Intelligence humaine et artificielle

Par Claudine Leleux
Source : https://www.claudineleleux.be/Articles.html

On ne peut ouvrir un seul journal, voire une seule revue, sans voir pointer la peur que nous apporte l'intelligence artificielle (IA). Va-t-elle nous détruire, comme elle détruira une série de métiers : du scénariste au photographe, du radiologue au producteur de disque, pour ne pas parler de l'analyste d'entreprise et du policier d'investigation ? Permet-elle au contraire des avancées inédites dans la connaissance et la compréhension du monde ? Et surtout, va-t-elle nous supplanter ?

Le présent article veut surtout faire la différence entre l'intelligence humaine (IH) et l'intelligence artificielle (IA) pour montrer que cette dernière ne peut supplanter la première même si elle peut l'augmenter et même si, tel le pharmakon (φάρμακον) que dénonçait Platon qui serait la meilleure et la pire des choses, elle peut être remède ou poison, selon l'usage qu'on en fait.

Avant de montrer en quoi l'intelligence humaine sera toujours nécessaire pour valider les résultats de l'intelligence artificielle, précisons d'abord le concept d'IA.

Qu'est-ce que l'intelligence artificielle ?

Pour s'assurer que l'intelligence artificielle ne peut supplanter l'intelligence humaine, essayons d'abord de définir l'intelligence dite artificielle.

Le concept est déjà galvaudé à la base puisque les Anglo-saxons ont recouru à la notion d'intelligence au sens d'*Intelligence Service* c'est-à-dire au sens d'un service de renseignement gouvernemental[1] qui collecte, analyse et diffuse des informations pour la sécurité nationale, la politique étrangère et la défense d'un pays. Les mots clefs ici sont *collecte*, *informations* ou *renseignements*.

Nous verrons que l'IA repose en effet sur la collecte d'informations qui ont été appelées « données » ou « data ». Par extension, on entend par intelligence artificielle la capacité d'une machine, d'un ordinateur, munis d'un programme développé par un ou plusieurs humains et appelé algorithme[2] ou ensemble d'algorithmes :

- à parler en langage familier dans des « chat » donnant l'illusion qu'on parle avec une personne, un être humain ;
- à résumer tous les arguments en faveur et en défaveur de la réponse à un problème, dans un langage familier ;
- à comparer des milliers de données, y compris des radios, des images ou des voix, en y détectant le moindre indice que l'ouïe ou l'œil humains ne suffisent pas à percevoir ;
- à organiser des parties de textes, d'images ou de voix de façon à créer de nouveaux ensembles, inconnus jusqu'ici
- ...

La référence aux faits

Un peu à la manière dont la police confectionne les portraits-robots, l'intelligence artificielle puise dans de grands ensembles de données les éléments qui vont composer la réponse à une question (qu'on appelle aussi « prompt »). Dans l'exemple du portrait-robot, les données

[1] Comme le britannique *MI6* (Secret Intelligence Service) pour l'étranger ou la *CIA* aux États-Unis.

[2] Un algorithme est une suite finie et non ambiguë d'instructions permettant de résoudre un problème ou d'accomplir une tâche à partir d'un ensemble de données initiales. Il est comparable à une recette de cuisine qui donne des étapes précises pour obtenir un plat à partir d'ingrédients. Le mot vient du nom du mathématicien persan Al-Khwârizmî (Gemini).

correspondent à quelque chose d'existant ou pas mais reposent en fin de compte sur un témoignage humain qui peut s'avérer sujet à caution. De même lorsque l'IA procède avec un ensemble de données, celles-ci peuvent certes faire référence aux faits au sens où ils existent, mais peuvent aussi faire référence à des données qui sont fausses.

Une première difficulté avec l'IA comme avec l'IH sera de départager le vrai du faux.

Si les critères sont des faits ou des existants, cela ne pose pas trop de problèmes. En revanche, si certaines données sont fausses, la question se pose de savoir quelle critique épistémologique opérer.

Au début, l'IA générative se limitait à calculer la probabilité du mot ou du token[3] suivant pour compléter une phrase, en s'appuyant sur un modèle de langage à grande échelle (*Large Language Model*, ou LLM). Cette approche consistait simplement à générer, à partir d'un vaste ensemble de données d'entraînement – souvent issues de sources variées, y compris celles détenues par les grandes entreprises technologiques comme les GAFAM –, une séquence de mots probables en réponse à un prompt donné.

Récemment, ces IA génératives ont été enrichies par des systèmes de récupération de données en temps réel[4] ou mises à jour fréquemment, afin de fournir des réponses plus contextualisées. C'est le cas notamment de la technique RAG[5] (Retrieval-Augmented Generation, ou Génération Augmentée par la Récupération). Cela accroît les chances que l'IA réponde au prompt en intégrant des informations les plus récentes et les plus pertinentes, mais cela ne garantit en aucun cas leur validité ou leur exactitude.

Les critères épistémologiques permettant d'évaluer la validité (vérité, fiabilité, exactitude) et la pertinence (adéquation au besoin informationnel) d'une source d'information s'avère particulièrement critique dans le contexte des IA génératives augmentées par récupération (RAG). Ces critères aident à distinguer les jugements épistémiques (liés à la valeur de vérité : crédibilité, autorité, exactitude) des jugements pragmatiques (liés à l'utilité : pertinence thématique, accessibilité).

Un cadre largement adopté pour structurer cette évaluation est le test **CRAAP** (Currency, Relevance, Authority, Accuracy, Purpose) que nous devons adapter au contexte des sources récupérées par une IA : Actualité (Currency) ; Pertinence (Relevance) ; Auteur (Authority) ; Exactitude (Accuracy) ; Objectivité / But (Purpose)

D'autres critères complémentaires, issus notamment des approches francophones en sciences de l'information, viennent enrichir cette grille :

- Fiabilité de la source : identification claire de l'origine et traçabilité.
- Qualité du contenu : logique, cohérence et absence d'erreurs manifestes.
- Vérification croisée : recoupement avec d'autres sources indépendantes pour confirmation.

En résumé, aucun critère n'est absolu ; l'évaluation reste toujours contextuelle et repose sur une combinaison de plusieurs dimensions. Dans le cadre des IA équipées de RAG, la traçabilité des sources citées renforce sensiblement la validité, mais une vigilance humaine demeure indispensable pour détecter les limites (biais de récupération, obsolescence).

Fausses informations dans les revues scientifiques

À ces difficultés, il faut ajouter que ces dernières années, nous avons vu augmenter le nombre de fausses informations dans des revues scientifiques prestigieuses. Au point d'attirer l'attention des savants sur la nécessité de vérifier la méthodologie de recherche utilisée et la fiabilité des mesures entreprises. Et ce, d'autant qu'un nouveau marché existe pour offrir,

---

[3] Selon Renaud Larsen, dans les modèles de langage à grande échelle (*Large Language Models* ou LLM), comme GPT, Grok ou Llama, le texte n'est pas traité mot par mot, mais sous forme de **tokens**. Un token est l'unité de base que le modèle utilise pour représenter et traiter le langage.

[4] Anne Catherine Simon, professeur à l'UCL dans *Le Soir* du Mercredi 11 juin 2025, page 24.

[5] Merci à Renaud Larsen pour m'avoir permis de publier les paragraphes suivants à partir de son cours sur l'IA.

notamment en Chine, aux Indes ou en Iran, moyennant paiement, des contributions aux revues scientifiques, parfois même des contributions que ces revues auraient déjà acceptées sous couvert d'anonymat. Marché qui, contre paiement, prend aussi en charge la multiplication des citations de ces articles frauduleux de façon que leur auteur apparaisse comme un expert en la matière[6].

Heureusement, les chercheurs commencent à mettre au point des outils pour détecter de tels articles frauduleux. Par exemple, comme ceux-ci, pour éviter d'être repérés pour plagiat, ont tendance à insérer des synonymes pour modifier l'apparence[7], des chercheurs ont eu l'idée de rechercher dans ces articles référencés les « phrases torturées »[8], c'est-à-dire ces indices de synonymes qui déforment le sens de la phrase, voire sa conformité à l'usage.

Ces chercheurs ont pu aussi mettre au point un outil qui détecte dans les bibliographies les articles qui ont été retirés pour incorrects ou faux par les comités éditoriaux des revues scientifiques et tentent d'interroger les auteurs qui ont référencé ces articles frauduleux de façon à vérifier l'impact de la fraude sur leur recherche.

Toujours au niveau des sources et de la difficulté de repérer les sources pertinentes, il faut ajouter la fabrication ces dernières années d'articles ou de photos par l'intelligence artificielle générative.

Discover pour Androïd et Google Chrome renvoient sans cesse à des sites soi-disant d'information mais qui sont générés par l'IA en fonction du profil de l'utilisateur[9].

Heureusement, l'erreur étant humaine, certains recopient le texte généré par IA en recopiant par mégarde le « bouton » que l'IA met à disposition de son utilisateur : « regenerate response ». Ce bouton sert alors de marqueur pour détecter par intelligence humaine une contribution générée par IA[10].

La véracité des données doit ainsi être contrôlée par des humains qui doivent mettre au point des outils de vérification et l'IA est peu de chose sans l'IH.

On doit s'attendre en plus à un foisonnement de données en expansion et l'inclusion de plus en plus importante d'informations elles-mêmes générées par IA. Au point de connaître un véritable « effondrement du modèle » et la suppression de toute fiabilité à moins de continuer à renouveler les données par des données humaines[11].

Seule l'intelligence humaine pourra nous prémunir contre ces erreurs et fraudes et élaborer les outils pour former l'IA à les débusquer. L'IA n'étant pas intelligente par elle-même, elle peut certes nous aider mais elle ne peut rien sans l'intelligence humaine. Celle-ci, en effet, a la capacité de réfléchir au tout produit par l'IA alors que l'IA ne peut pas réfléchir sur elle-même. Ou, comme le formule Jean-Marc Ferry dans les *Grammaires de l'intelligence*[12], l'IA peut certes s'adresser à quelqu'un (2e personne) en se référant à quelque chose (3e personne) mais sans un engagement de sa (première) personne.

Référence aux normes et aux valeurs

Jusqu'à présent, nous avons abordé la question de l'intelligence artificielle relativement aux faits, avérés ou non, mais quels rapports peut-elle entretenir avec les normes et les valeurs ? En effet, d'un point de vue épistémologique, il importe d'être capable de discerner les uns et les autres[13].

---

[6] Guillaume Cabanac, « Comment dépolluer la science ? », *Rencontres* sur le site de Cairn.info du 15 juin 2025.

[7] Par exemple « clameur » plutôt que « bruit », *ibid*.

[8] *Ibid*.

[9] *Le Monde* du 26 septembre 2025, p. 15.

[10] *Ibid*.

[11] Renaud Larsen, Jacques Steiwer [2025], *De la souveraineté*, éd. Samsa.

[12] Paris, 2004, Éd. du Cerf, Collection "Passages".

[13] Claudine Leleux, « Faits, normes et valeurs : une confusion dommageable pour l'enseignement moral et citoyen », Bruxelles, 2019, 26 pp., Open edition.

Hilary Putnam déclarait en 2002 : « Aujourd'hui, ce n'est pas se placer dans une tour d'ivoire que de se demander quelles différences il y a entre les jugements "factuels" et les jugements de "valeurs". Il se pourrait bien que l'enjeu en soit un problème – littéralement – de vie et de mort.[14] »

Je ne pense pas qu'il y aille de la vie ou de la mort et qu'il faille à ce point durcir le trait. Mais, en revanche, la question que Putnam mentionne, ajoutée au défi actuel de la place grandissante qu'occupe l'intelligence artificielle, mérite qu'on s'y attarde.

Tout d'abord, il nous faut bien distinguer jugement de fait, jugement normatif et jugement de valeur. Pour nous aider à bien délimiter les champs de ceux-ci, nous pouvons partir de l'exemple cité par Nathalie Heinich[15] :

> La température est de 25° (énoncé descriptif qui prétend à l'exactitude)
>
> Il fait trop chaud (énoncé expressif qui prétend à l'authenticité)
>
> Baisse le chauffage (énoncé normatif qui prétend à la justesse)

Ou de l'exemple cité par Ruwen Ogien[16] :

> Les spaghettis sont cuits *al dente* (énoncé descriptif qui prétend à l'exactitude)
>
> Les spaghettis cuits *al dente* sont bien meilleurs (énoncé expressif qui prétend à l'authenticité)
>
> Pour être *al dente*, les spaghettis ne doivent pas rester trop longtemps dans l'eau bouillante (énoncé normatif qui prétend à la justesse).

L'enjeu épistémologique de cette différence « linguistique » a trait à la validité de l'énoncé, en l'occurrence ici les énoncés des données apprises (l'*e-learning*) par l'IA générative. S'il l'on peut montrer que les énoncés descriptifs et les énoncés normatifs peuvent faire l'objet d'une certaine validité – et donc prétendre à une certaine objectivité *vs* subjectivité – selon des conditions à définir, les jugements de valeur, eux, peuvent tout au plus valoir pour ceux qui les énoncent.

Jugements de goût ou de valeur

Dans ce dernier cas, par conséquent, il y a lieu de penser que l'IA générative ne repèrera que les jugements de valeurs majoritaires et éliminera de son champ d'*e-learning* les jugements très singuliers. Ou que les programmateurs d'algorithmes privilégieront certains jugements de valeurs au profit d'autres. Par exemple, à la question « Préfères-tu la liberté à la sécurité ? », Chat GPT répond :

> « Donc si tu me poses la question personnellement, je dirais que je n'ai pas de préférence en tant qu'IA. Mais pour un humain, le choix dépend souvent du contexte :
>
> - en temps de **crise ou de danger**, on préfère la sécurité ;
> - en temps de **paix et de stabilité**, on redemande plus de liberté. »

---

[14] Hilary Putnam [2002], *Fait / Valeur : la fin d'un dogme et autres essais*, trad. M. Caveribère et J.-P. Cometti, éd. de l'éclat, coll. "tiré à part", p. 12.

[15] Nathalie Heinich [2017], *Des valeurs. Une approche sociologique*, Gallimard, coll. "Bibliothèque des sciences humaines", p. 125.

[16] Ogien Ruwen, « Repenser les relations entre les faits, les normes et les valeurs », *Les Sciences de l'éducation – Pour l'Ère nouvelle*, 2012/1 (Vol. 45), pp. 17-31. DOI : 10.3917/lsdle.451.0015. URL : https://www.cairn.info/revue-les-sciences-de-l-education-pour-l-ere-nouvelle-2012-1-page-17.htm : « "Ces spaghettis sont cuits al dente", semblent avoir pour caractéristique de décrire des états de choses. D'autres comme : "Les spaghettis cuits al dente sont bien meilleurs" ne semblent pas décrire des états de choses mais indiquer quels états de choses sont désirables ou préférables. D'autres enfin, comme "Pour être al dente, les spaghettis ne doivent pas rester trop longtemps dans l'eau bouillante" disent quels états de choses sont obligatoires, permis ou interdits. Les énoncés qui décrivent des états de choses sont appelés descriptifs ou factuels. Ceux qui posent ce qui est désirable, préférable, sont appelés évaluatifs. C'est aux énoncés qui contiennent les expressions dites "déontiques" (du grec *deon* : devoir), "Il est obligatoire", "Il est interdit", "Il est permis", que l'appellation normatifs convient le mieux.[16] »

Et de poursuivre en disant ce que tel ou tel auteur en pense :

> « Certains (comme Benjamin Franklin) disaient déjà que « ceux qui abandonnent une liberté essentielle pour un peu de sécurité temporaire ne méritent ni l'une ni l'autre ».
>
> D'autres philosophes (comme Hobbes) soutenaient au contraire que les humains acceptent de limiter leur liberté pour que l'État leur garantisse la sécurité, sans quoi règne la peur permanente. »

Il en va de même à propos d'un jugement de goût ou d'un jugement esthétique. Par exemple, à la question « Aimes-tu Rotko ? », Chat GPT répond :

> « Je n'ai pas de goûts personnels comme un être humain, mais je peux te dire que Rothko est un peintre qui suscite énormément de passions. Beaucoup de gens disent qu'ils « aiment » ou qu'ils se sentent profondément bouleversés par ses grands aplats de couleurs, parce qu'ils créent une sorte d'espace méditatif, silencieux, presque spirituel.
>
> Certains, au contraire, trouvent son travail trop minimaliste ou ne comprennent pas l'émotion qu'il est censé provoquer. »

En résumé, le programmateur de l'e-learning a prévu que lorsque l'on demande à ChatGPT de se prononcer à propos d'un goût, d'une préférence ou d'une valeur, il affirmera ne pas avoir de goût, de préférence ou de valeur personnels mais fera état de ce qu'éprouvent certains et, au mieux, les uns et les autres en fonction des données auxquelles il aura accès.

À cet égard, il peut y avoir **un biais cognitif par l'effet de contexte**. En effet, du point de vue quantitatif des données, une majorité d'entre elles étant réunies par des moteurs de recherche majoritaires voire monopolistiques, elles peuvent être issues d'un même contexte culturel et/ou géopolitique induisant par là même un profil réduit et gommant de nombreuses différences.

Jugements normatifs

En matière de normes, il faut différencier les normes juridiques et les normes morales[17] mais avant même cette distinction, dans quelle mesure l'IA peut-elle simplement repérer qu'il s'agit d'une norme, voire d'un idéal de la raison et non d'un fait. Ainsi dans l'exemple illustré par le premier article de la Déclaration Universelle des Droits de l'homme de 1948 : « Tous les êtres humains naissent libres et égaux en dignité et en droits. », même si formellement l'énoncé se présente comme étant descriptif d'un fait, il indique au contraire un devoir, une norme d'action, fût-elle idéale et non appliquée. C'est le sens de l'énoncé qui nous dit qu'il s'agit d'une norme idéale et ce sens n'est pas réductible à un fait ou à une donnée quantifiable. À la question : « Que signifie "Les hommes naissent libres et égaux" ? », ChatGPT repère parfaitement qu'il s'agit d'un droit (juridique) qui entraîne des obligations légales :

D'abord le « fait » :

> « La formule *"***Les hommes naissent libres et égaux"** est extraite de la **Déclaration des droits de l'homme et du citoyen de 1789** (article 1 : *" Les hommes naissent et demeurent libres et égaux en droits*"). »

Ensuite, le commentaire de l'article (l'exposé des motifs) :

> « Elle condense plusieurs idées :
>
> **Naissance** :
> → On affirme qu'à la base, tout être humain possède une dignité et une valeur intrinsèques qui ne dépendent ni de sa naissance, ni de son rang social, ni de sa religion, ni de sa richesse.
>
> **Liberté** :
> → Chaque personne a, dès sa venue au monde, le droit d'agir librement, dans le respect de la liberté des autres et des lois. La liberté est vue comme un état naturel de l'homme, et non comme un privilège accordé par un pouvoir.
>
> **Égalité** :
> → L'égalité n'est pas une identité (les hommes sont différents par leurs talents, leurs situations…), mais une **égalité en droits** : personne ne devrait avoir plus ou moins de droits fondamentaux qu'un autre du fait de sa naissance ou de son statut.

[17] Claudine Leleux, « Faits, normes et valeurs : une confusion dommageable pour l'enseignement moral et citoyen », *ibid*.

En somme, c'est une affirmation **philosophique et politique** :

- Philosophique, parce qu'elle pose un principe universel de dignité humaine.
- Politique, parce qu'elle s'oppose aux sociétés d'Ancien Régime où les droits et libertés dépendaient de la classe sociale (noblesse, clergé, tiers état). »

Et ChatGPT repère parfaitement, par comparaison, qu'il s'agit d'un idéal et non d'un fait :

« Autrement dit, ces deux affirmations sont des **idéaux** : elles disent moins ce qui *est* que ce qui *doit être*. Elles servent de repères pour corriger les inégalités et les oppressions qui existent toujours. »

Si l'IA a été entraînée à repérer le fait d'une norme et une norme d'un idéal, qu'en est-il maintenant de la validité des normes énoncées par l'IA ?

Pour répondre à cette question, il nous faut faire la distinction entre la validité d'un énoncé et celle d'une énonciation. En effet, depuis le tournant linguistique, nous avons dû différencier la validité d'un énoncé (logique) et celle d'une énonciation (pragmatique)[18].

Énoncé et énonciation

Jusqu'au tournant linguistique, les logiciens abordaient la question de l'énoncé comme s'il était une proposition susceptible d'être vraie ou fausse. Ainsi, par exemple, la proposition *p* « Il pleut ou il ne pleut pas » est logiquement toujours vraie tandis que la proposition *q* « Il pleut et il ne pleut pas » est toujours logiquement fausse. Comme on le sait, toute proposition en tant que telle n'est pas susceptible de vérité ou de fausseté logique. Par exemple, si je dis : « Il pleut », la proposition est indécidable en tant que telle. Tout au plus, est-elle une proposition qui prétend qu'il pleut et c'est, comme le montrera Jürgen Habermas[19], la discussion des personnes concernées, en l'occurrence celles à qui je m'adresse, qui devront en reconnaître la validité ou au contraire la fausseté en référence à la pluie qui tombe ou non. D'où l'intérêt d'aborder les propositions non pas sous l'angle sémantique (l'énoncé) mais dans leur dimension pragmatique ou performative (l'énonciation) : « Qu'est-ce que je fais lorsque je dis qu'il pleut ? » veut dire que je prétends trouver un accord avec mon interlocuteur sur ce que je dis. L'accord ou la validité n'est pas intrinsèque à l'énoncé mais le résultat d'un accord entre interlocuteurs.

De même, lorsque je dis : « Il est vrai que X paie ses impôts », une enquête, une discussion, une vérification, permettront de conclure à l'exactitude ou la fausseté de l'énoncé. S'il est exact, sans doute le philosophe comme l'homme de la rue seront d'accord pour dire qu'il s'agit d'un *fait*. Le fait n'est donc pas quelque chose qui existe mais qui résulte d'une discussion à l'issue de laquelle un état de choses sera établi.[20]

Du coup, la phrase de David Hume que cite Putnam[21] comme étant problématique : « où est ce fait que nous appelons ici crime[22] » permet d'être relativement vite élucidée. Le crime est un fait et jugé en tant que tel lors du « jugement ». La « vérité judiciaire », le fait, est le résultat d'une discussion. Pour qualifier un événement de fait, on peut certes recourir à l'expérience

---

[18] John Langshaw Austin, *Quand dire, c'est faire*.

[19] Jürgen Habermas [1972], « Théories relatives à la vérité », trad. R. Rochlitz, dans *Logique des sciences sociales et autres essais*, PUF, 1987.

[20] Comme le dit Habermas dans deux autres exemples :

« Le fait que le feu soit passé à l'orange ou qu'une pomme soit déjà jaune, c'est dans le contexte pratique de la circulation ou de la vente de fruits mûrs, une information (la communication d'une expérience relative à l'action) ; on peut dire également que ce sont des faits, mais on ne le dit – autrement dit on ne parle de faits – que lorsqu'il est nécessaire, après un accident de la circulation, par exemple, d'éclaircir l'état de choses (*Sachverhalt*) et de savoir si, à un moment donné, ce feu est passé à l'orange, ou dans le cas où il faut établir, dans le cadre d'une expérience horticole, si cette pomme, à un moment donné, était déjà jaune. » (Jürgen Habermas, *op. cit.*, pp. 282-283).

[21] Hilary Putnam [2002], o*p. cit.*, p. 16.

[22] David Hume [1751], *Enquête sur les principes de la morale*, (1991) trad. P. Baranger et P. Saltel, GF, n° 654, p. 208.

mais, comme le souligne Habermas, celle-ci devra faire l'objet d'une interprétation qui devra elle-même être discutée :

> « […] pour prendre méthodiquement appui sur l'expérience, par exemple dans l'expérimentation, on est obligé de recourir à des interprétations dont la validité ne peut s'avérer que dans la discussion.[23] »

C'est aussi dans ce sens qu'il est permis de comprendre l'opposition de Nietzsche aux positivistes lorsqu'il affirme :

> « non, justement il n'y a pas de faits, seulement des interprétations.[24] »

même s'il le dit dans une optique perspectiviste plutôt que discursive à l'instar d'Habermas.

Première conclusion sur la différence entre l'intelligence artificielle et humaine

Si nous suivons ici Habermas, nous voyons combien l'intelligence artificielle se démarque de l'intelligence humaine : la première procède par probabilités tandis que la seconde établit discursivement les faits. La discussion fait place à l'altérité et à la justification (les raisons) d'un accord ou non avec une énonciation alors que l'IA pourra tout au plus faire l'inventaire de ces raisons sans pouvoir se mettre d'accord avec elles.

Prétention à la validité et validité

Outre la distinction entre fait et jugement de fait et entre énoncé et énonciation, l'approche habermassienne – de présupposer pragmatiquement que lorsque je dis ou j'argumente, je prétends m'entendre avec un interlocuteur – nous conduit aussi à distinguer d'une part la prétention à la validité et, d'autre part, la validité. Lorsque j'affirme une proposition *p*, j'entends m'entendre avec mon interlocuteur sur *p*. *p* peut être un énoncé descriptif / constatif qui énonce ce qu'il en *est* de quelqu'un ou de quelque chose mais elle suppose un interlocuteur qui confirme l'exactitude de ce que je dis ou non.

Comment passer pour Habermas de la *prétention* à la validité (à l'exactitude) à la validité (l'exactitude) de l'énoncé ?

Soit l'énoncé est tout simplement logiquement vrai ou faux (contradiction logique). Soit l'énonciation est fausse. Par exemple, l'énonciation de Nietzsche selon laquelle « il n'y a pas de vérité[25] » est fausse dans la mesure où prétendre à l'absence de vérité revient à en énoncer une, à savoir qu'il n'y en a pas. C'est la « contradiction performative » qui vient compléter la boîte à outils d'invalidation des énoncés du logicien. C'est-à-dire la contradiction entre l'énoncé et l'énonciation. Par exemple, lorsque je dis : « J'étais sur un bateau qui a fait naufrage et il n'y a pas eu de survivant » est une contradiction entre ce que je dis et ce que je fais en le disant. L'adresse à un interlocuteur à propos de ce naufrage et de ceux qui n'en auraient pas survécu *contredit* ce qui est dit.

Lorsqu'il s'agira de problèmes plus complexes, dans le cas de l'énoncé d'hypothèses scientifiques ou techniques, de modèles mathématiques ou de théories scientifiques, de faits historiques et même d'expertises esthétiques ou philosophiques, la discussion ne peut se confondre avec une simple interlocution. Pour que les énoncés soient exacts, il faut que la discussion remplisse les critères d'une « situation idéale de parole » et se produise entre toutes les personnes concernées, en l'occurrence au sein de la communauté des savants

[23] Jürgen Habermas [1972], *op. cit.*, p. 283.

[24] Friedrich Nietzsche [1886], *Fragments posthumes* dans *Œuvres Philosophiques Complètes*, vol. XII, 7[60], pp. 304-305 (texte allemand dans *Nachlassene Fragmente. Herbst 1885 bis Herbst 1887*, Walter de Gruyter, Berlin-New York, 1974, p. 323 : « Gegen den Positivismus, welcher bei dem Phänomen stehen bleibt „es giebt nur Thatsachen“, würde ich sagen: nein, gerade Thatsachen giebt es nicht, nur Interpretationen. »)

[25] Friedrich Nietzsche [1888], *Le crépuscule des idoles ou Comment philosopher à coups de marteau*, trad. J.C. Hémery, Folio essais, 1974, n° 88, p. 31. « Le “monde vrai”, une idée qui ne sert plus à rien, qui n'engage même plus à rien – une idée inutile, superflue, *par conséquent* une idée réfutée : abolissons-la. »

concernés. Ce devoir, cette contrainte, se justifie pour Apel et Habermas[26], par la présupposition pragmatique de toute personne qui argumente auprès d'autres.

Une situation *idéale* de parole doit remplir au moins quatre conditions selon Habermas :

> « Personne ne peut sérieusement entamer une argumentation s'il ne suppose pas une situation de parole qui garantisse, en principe, la publicité de l'accès, l'égalité de participation, la sincérité des participants, des prises de position sans contrainte, etc.[27] »

Publicité. Cela signifie que tous les savants ou les experts concernés par la prétention à la validité de l'énoncé doivent être au courant du déroulement de la discussion (publicité) pour pouvoir y participer et, toujours idéalement, aucun d'entre eux ne peut en être exclu. Les contre-exemples sont légion et invalident ainsi de fait certains énoncés scientifiques. Outre les climato-sceptiques, certains savants reprochent au GIEC[28] par exemple de valider leurs conclusions par les représentants gouvernementaux plutôt que par la communauté scientifique.

Égalité de participation. Il va de soi que la discussion doit faire place à une égalité de participation entre les savants. Aucun argument d'autorité ne peut, par exemple, prévaloir au nom de la notoriété de certains experts plutôt que d'autres.

Liberté, sans contrainte. Certains savants qui dépendent de subventions, pour leurs recherches par exemple, doivent pouvoir réellement discuter sans risque de perdre ces moyens alloués, a fortiori sans risque de perdre leur emploi.

Sincérité. Que ce soit à cause du risque de représailles ou à cause de ce que Bachelard appelait les « obstacles épistémologiques », aucun participant ne devrait être de mauvaise foi mais devrait sérieusement se laisser convaincre par l'argument le meilleur.

Les situations empiriques devront dans la mesure du possible se rapprocher au plus près de la situation idéale de parole pour que la validité d'un énoncé descriptif / constatif soit valide. Cette validité, pour les tenants de la pragmatique universelle, Apel et Habermas, est toujours provisoire, tant qu'une nouvelle discussion, à la suite par exemple d'une nouvelle expérience, ne vienne la contester ou la problématiser.

Pour ceux qui, comme Hilary Putnam[29], reprocheraient à Habermas de recourir aux normes/valeurs de publicité, égalité, liberté et sincérité, il s'agit, selon ce dernier, de bien discerner deux types de normativité : d'une part, une normativité morale, voire juridique, et d'autre part, une normativité qui, à l'instar de toutes les règles de logique qui favorisent l'intelligibilité du discours, n'a pas de sens pratique :

> « Pour répondre à l'objection de circularité, j'aimerais indiquer que le contenu des présuppositions universelles de l'argumentation n'est encore nullement "normatif" au sens moral. […] La répartition égalitaire des libertés de communiquer *à l'intérieur* de la discussion et l'exigence de sincérité *pour* la discussion sont des obligations et des droits *relatifs à l'argumentation* et nullement *d'ordre moral*. […] Les règles constitutives du jeu d'argumentation définissent l'échange des arguments et des prises de position par oui ou par non ; elles ont un sens épistémique qui consiste à permettre la justification de certains énoncés, et non le sens pratique immédiat de motiver des actions.[30] »

Autrement dit, les publicité, sincérité, liberté et égalité, au sens de normes épistémiques, permettent de vérifier si les conditions d'une situation idéale de parole sont remplies.

---

[26] Jürgen Habermas [2000], « Valeurs et normes. À propos du pragmatisme kantien de Hilary Putnam », trad. R. Rochlitz dans Rainer Rochlitz (dir) [2002], *Habermas. L'usage public de la raison*, PUF, coll. "Débats philosophiques", p. 235.

[27] Jürgen Habermas [1991] *De l'éthique de la discussion*, trad. M. Hunyadi, Paris, éd. du Cerf, 1992, p. 122.

[28] Groupe d'Experts Intergouvernemental sur l'évolution du Climat.

[29] Hilary Putnam [2002], *op. cit.*, p. 15 : « l'objectivité postulée par Habermas *présuppose* en réalité l'objectivité d'au moins quelques valeurs ».

[30] Jürgen Habermas [1996], « Le contenu cognitif de la morale », trad. R. Rochlitz dans Jürgen Habermas [1996], *L'intégration républicaine. Essai de théorie politique*, Fayard, 1998, p. 61.

Il nous paraît évident de rappeler à ce propos toute la distance qui sépare l'intelligence artificielle et l'intelligence humaine. La première, à supposer qu'elle rencontre le critère de « publicité », échoue aux trois autres. La machine ne peut exprimer de sentiments réels, cela n'a aucun sens de dire qu'elle est sincère ou pas puisqu'elle procède comme on l'a programmée. Pour les mêmes raisons, elle n'est pas libre mais dépend des programmateurs de ses algorithmes qui ne sont forcément pas à égalité de participation avec les autres savants.

L'IA, à la différence de l'intelligence humaine, ne peut répondre aux critères d'une situation idéale de parole et ne peut donc juger de la validité discursive des faits pas plus que de la validité des normes. En outre, les critères de validité des normes sont appris par l'IA par l'intermédiaire d'humains chargés de valoriser ou de disqualifier des normes d'action et des comportements inappropriés. C'est ainsi que, par exemple, comme le rappelle Didier van Cauwelaert, de nombreux :

> « […] esclaves modernes, délocalisés en grande partie dans le "Sud global", qui passent dix heures par jour derrière leur écran à visionner ce qu'il y a de pire – meurtres, massacres, viols, zoophilie… – de manière à faire des annotations afin que l'IA prenne en compte que ce sont des "comportements inappropriés" »[31].

Deuxième conclusion sur la différence entre l'intelligence artificielle et humaine

Autrement dit, l'IA ne peut recourir à la discussion pour valider ses énoncés et se présentera toujours comme la somme de prétentions à la validité qui devront, pour être validées, recourir à l'intelligence humaine.

Notons enfin que la prétention à la validité peut être ou non corroborée, ce qui signifie que la validité est toujours provisoire et ainsi en va-t-il de la validité des théories scientifiques. Ces révisions peuvent même donner lieu à des changements de paradigmes qui sont à l'origine des révolutions scientifiques, comme l'a montré Thomas Kuhn[32]. Marius Bertolucci dans un livre récent, *L'homme diminué*, insiste sur cette particularité de l'intelligence humaine, qu'elle révise ses croyances ce dont n'est pas capable l'IA en tant que telle :

> « La science est une discipline révisionniste en cela qu'elle révise constamment ses croyances en fonction des nouvelles découvertes. L'ère des données massives propose la fin de l'histoire des sciences.[33] »

Claudine Leleux
13 décembre 2025

[31] Entretien avec Didier van Cauwelaert dans Le Soir du 14 novembre 2025, p. 3.

[32] Thomas Kuhn, *La structure des révolutions scientifiques* [1962/1970], trad. L. Meyer, Champs Flammarion, 1983, pp. 23-24 : « […] dans les révolutions scientifiques. Chacune d'elles a exigé que le groupe rejette une théorie scientifique consacrée par le temps en faveur d'une autre qui était incompatible. Chacune a par conséquent amené un déplacement des problèmes offerts à la recherche scientifique et des critères selon lesquels les spécialistes décident de ce qui doit compter comme un problème admissible ou comme une solution légitime. »

[33] Marius Bertolucci [2023], *L'homme diminué par l'IA*, Hermann, Éd. Kindle, p. 146.