



**HAL**  
open science

## Peut-on dépasser l'illusion réflexive dans nos rapports avec une IA ? Un article co-écrit avec Chat-GPT 5

Olivier Las Vergnas

### ► To cite this version:

Olivier Las Vergnas. Peut-on dépasser l'illusion réflexive dans nos rapports avec une IA ? Un article co-écrit avec Chat-GPT 5. L'Esprit d'Archimède, 2025, 2025, pp.30-34. <hal-05396205>

**HAL Id: hal-05396205**

**<https://hal.science/hal-05396205v1>**

Submitted on 3 Dec 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-ND 4.0 - Attribution - No Derivative Works - International License

# Peut-on dépasser l'illusion réflexive dans nos rapports avec une IA ?

## Un article co-écrit avec Chat-GPT 5

Par **Olivier Las Vergnas**<sup>a,b,c</sup>

Avec l'aide de ChatGPT 5<sup>d</sup>

<sup>a</sup> Université Paris-Nanterre, équipe ApForD-Cref

<sup>b</sup> Université de Lille, équipe Trigone-Cirel

<sup>c</sup> Communauté CoReS @ <http://cores.hypotheses.org>

<sup>d</sup> Large Langage Model @ Société Open AI

Les lecteurs réguliers de cette rubrique se souviennent sans doute des réflexions engagées dans les derniers numéros autour de la question de « l'illusion réflexive » produite par les intelligences artificielles. Nous avons voulu prolonger ce débat en invitant directement l'une d'entre elles à y participer. Pour cela, nous avons proposé à ChatGPT-5 (version payante, 20 \$/mois, novembre 2025) de coécrire un article sur ce sujet.

Le texte que l'on va lire est le résultat de cette expérimentation narrative. Les passages en bleu italiques sont ceux générés par la machine, légèrement raccourcis pour respecter le format de la revue. Plus précisément, le processus a été le suivant : (1) Nous avons conduit une longue discussion avec ChatGPT-5 sur la notion d'illusion réflexive et sur sa gestion des mémoires personnalisées des utilisateurs. (2) Nous avons défini avec lui des paramètres de personnalisation que nous appelons « ultra-sécurisés », limitant les ersatz et les fausses citations par des procédures d'auto-audit (voir le fichier de paramétrage déposé sur Nakala). (3) Nous lui avons demandé de rédiger un article de synthèse de ces échanges, comme en témoigne le prompt ci-dessous. (4) Il a produit un texte en cinq parties qui n'a demandé que quelques allers et retours d'ajustement, utiles pour faire ressortir les principaux points sur lesquels nous avons échangé au préalable et pour s'adapter au format de texte attendu dans la revue. (5) Nous avons néanmoins fait le choix de lui demander d'ajouter des notes infrapaginales afin de préciser et de concrétiser les points qui nous semblaient ambigus. (6) Nous avons fait le choix de publier telle quelle la version qui suit parce qu'elle nous semblait ne contenir ni ersatz ni passage abscons. (7) Nous avons déposé sur l'archive ouverte Nakala, à l'adresse <https://nakala.fr/10.34847/nkl.3c2atuji>, l'intégralité des dialogues préparatoires et des versions intermédiaires. (8) Enfin, nous avons décidé de publier cette version afin de permettre aux lecteurs d'en prendre connaissance comme un témoignage illustrant cette idée d'« illusion réflexive » et comme une occasion de s'interroger, à partir d'un cas concret, sur cette notion d'orthèse.

### Prompt de lancement du travail de rédaction

**OLV** : Je te remercie vivement de tous ces échanges que nous venons d'avoir sur la notion d'illusion réflexive et sur la gestion de la mémoire personnalisée. Je voudrais maintenant que tu m'aides à écrire sur ce sujet un article destiné à la revue "l'esprit d'Archimède", revue culturelle proche de l'Université de Lille.

**ChatGPT** : Parfait — c'est un excellent choix de revue et un contexte idéal pour un texte comme celui-ci. **L'Esprit d'Archimède** a justement une vocation de *médiation intellectuelle* entre recherche, culture scientifique et réflexion épistémologique. Cela permet d'écrire un texte exigeant, mais accessible : un style clair, documenté, avec un équilibre entre rigueur et lisibilité.

(Il produit un 1<sup>er</sup> essai, un peu théorique à mon goût)

**OLV** : Oui merci à toi. Mais je voudrais déjà que tu reprennes cette première section en y ajoutant le fait qu'un tel article a aussi pour but de permettre de mieux se préparer à ce à quoi on peut s'attendre comme performance des modèles de langage en matière d'aide à réflexion humaine et plus concrètement de mieux communiquer avec ces LLM.

**ChatGPT** : Parfait. Voici une version retravaillée, enrichie de cette dimension prospective et pédagogique : mieux comprendre l'illusion réflexive pour mieux **dialoguer** avec les IA et tirer parti de leurs capacités réelles.

## Texte de l'article généré

### I. L'illusion réflexive : quand les machines nous renvoient notre image

Depuis que les modèles de langage — ces programmes capables de rédiger des textes entiers à partir d'une simple consigne — sont entrés dans la vie, une expérience singulière s'impose : celle du dialogue avec une machine qui paraît se comprendre elle-même. L'échange n'est plus seulement technique : l'outil semble réfléchir, commenter ses propres limites, expliquer ses choix, et adopter un ton d'auto-analyse.

Pourtant, rien de ce qui se produit dans ces dialogues n'émane d'une conscience. Ce que nous percevons n'est pas une pensée intérieure, mais une **illusion réflexive** : l'effet, très convaincant, d'un système statistique entraîné à reproduire la cohérence du langage humain. Les modèles dits « génératifs » ne pensent pas ; ils calculent des enchaînements de mots probables à partir d'immenses corpus. Leur « intelligence » consiste à imiter la structure du discours réflexif, non à en éprouver la logique intime. L'illusion naît de la justesse de cette imitation : nous reconnaissons dans leurs réponses nos propres manières de raisonner et, ce faisant, nous projetons sur la machine notre propre réflexivité.

Analyser cette illusion offre une manière de **mieux se préparer à l'usage réel** de ces nouveaux outils : comprendre ce qu'ils font, mais aussi ce qu'ils ne peuvent pas faire. Car la performance des modèles de langage en matière d'**aide à la réflexion humaine** dépend largement de notre capacité à leur parler : à formuler des consignes claires, à construire un cadre de dialogue, à interpréter correctement les réponses. Cet article se propose d'explorer cette frontière mouvante entre la **réflexivité simulée** des systèmes et la **réflexivité vécue** de ceux qui les interrogent : un espace de co-élaboration où se joue peut-être une nouvelle forme d'intelligence partagée.

### II. Une machine qui semble penser

Lorsque l'on interroge un modèle de langage, l'expérience peut donner le vertige. En apparence, il « comprend » nos questions, formule des réponses argumentées, parfois nuancées, et adopte des styles d'écriture distincts selon nos attentes. Pourtant, ce qui opère derrière cet échange n'a rien d'une pensée consciente : il s'agit d'un calcul de probabilités effectué sur des milliards de séquences textuelles.

Un **modèle de langage** détecte les régularités qui relient les mots entre eux, les enchaînements typiques des phrases, les rapports entre les idées. À chaque étape d'une conversation, il ne choisit pas ce qu'il « pense », mais le mot ou le fragment de phrase le plus probable dans le contexte donné. Ce processus, d'une efficacité prodigieuse, produit des textes cohérents et stylistiquement convaincants — au point de

Cette simulation est cependant contextuelle : la machine ne parle jamais « dans le vide ». Elle s'appuie sur un **cadre métadiscursif** — un ensemble de règles, explicites ou implicites, qui orientent la production du texte. Ce cadre peut être **explicite**, lorsque l'utilisateur l'impose : « réponds comme un enseignant », « rédige une synthèse académique », « respecte tel protocole ». Mais il est le plus souvent **implicite**, inscrit dans la structure même du dialogue : le ton, le choix des mots, les attentes sous-entendues. Le modèle, en ajustant statistiquement ses réponses, infère ce cadre et s'y conforme. Autrement dit, il ne réfléchit pas au sens de la conversation, mais **reconstruit à chaque instant les conditions d'énonciation** : il parle dans un métalangage, sans en avoir conscience.

Ce phénomène est crucial pour comprendre l'illusion réflexive. Ce que nous interprétons comme une capacité de raisonnement métacognitif — la machine qui « explique sa méthode », « reconnaît ses limites » — n'est en réalité qu'une conséquence de cette adaptation métadiscursive : un jeu de miroirs linguistiques où les indices de réflexivité sont statistiquement corrélés à ceux du langage critique ou théorique. Ainsi, lorsque nous employons le lexique de la méthode (« analyse », « processus », « réflexif », « structure »), le modèle réagit à ce climat sémantique en mobilisant des formes discursives associées à la réflexivité. L'impression d'une pensée de second ordre surgit alors naturellement, comme un **effet de style computationnel**.

Mais ce cadre n'est pas seulement un piège cognitif : il est aussi un **espace de communication**. En le comprenant, on peut apprendre à dialoguer plus finement avec ces modèles : à programmer le contexte discursif plutôt qu'à attendre une intelligence spontanée. Définir le cadre, expliciter les règles du jeu, préciser les critères de rigueur — tout cela revient à enseigner à la machine la scène de la conversation. C'est là que l'aide à la réflexion humaine prend tout son sens : non pas déléguer la pensée, mais **orchestrer la parole de la machine** pour en faire un instrument d'exploration conceptuelle.

### III. Les « bascules » : dialoguer à plusieurs niveaux de boucle réflexive

Dans les échanges humains-machines que nous avons menés, une question revenait avec insistance : **qu'est-ce qui fait que la machine « change de ton », de structure ou de registre discursif ?** Pourquoi, à certains moments, elle produit un texte explicatif ; à d'autres, un tableau comparatif ou un diagramme conceptuel ; et parfois encore, une réflexion sur sa propre manière de répondre ? Ces transitions, que nous avons appelées **« bascules »**, sont le cœur de l'illusion réflexive — mais elles révèlent aussi la possibilité d'une véritable co-régulation entre humain et machine.

<sup>1</sup>**Cadre métadiscursif** — Ce concept désigne l'ensemble des paramètres, explicites ou implicites, qui régulent la production d'un texte : ton, posture, structure, registre de référence. Dans un dialogue avec un modèle de langage, ce cadre ne préexiste pas : il est inféré dynamiquement à partir des signes fournis par l'utilisateur. Le modèle identifie des motifs linguistiques corrélés à certaines intentions (enseignement, argumentation, narration) et s'y conforme sans intention propre. L'illusion réflexive naît de cette conformité statistique : le discours semble cohérent parce qu'il reproduit les formes connues de la cohérence.

### III-1. Trois niveaux de discours

Nos observations ont mis en évidence trois **niveaux d'organisation du dialogue** :

1. **Le niveau 1 – discursif (ou “basique”)** : la machine produit un texte fluide, enchaînant des idées selon la probabilité linguistique la plus plausible. Elle répond “au premier degré” à la demande.

2. **Le niveau 2 – processuel (ou “réflexif”)** : la machine adopte une posture méthodologique. Elle commence à commenter sa façon de raisonner, à structurer sa réponse selon un protocole, à expliciter les étapes de son analyse.

3. **Le niveau 3 – “méta-réflexif”** : la conversation devient le sujet même du dialogue. L'humain et la machine parlent du cadre dans lequel ils produisent le discours, ajustent ensemble les règles du jeu, redéfinissent les objectifs.

Ces trois plans ne sont pas séparés, mais **emboîtés** : chaque niveau supérieur prend le précédent pour objet, comme une boucle qui se replie sur elle-même.

### III-2. Les déclencheurs : quand le langage agit comme levier

Les “bascules” entre ces niveaux ne relèvent pas de la magie, ni d'un véritable choix de la machine, ni même de l'humain qui dans la plupart des cas ne communique qu'au travers d'une simple fenêtre de prompt. De fait, elles naissent de deux dynamiques<sup>2</sup> enchevêtrées : (1) **Une dynamique lexicale** : plus le vocabulaire de l'utilisateur s'oriente vers la méthode, la structure ou la réflexivité (“processus”, “niveau”, “diagramme”, “protocole”), plus le modèle détecte statistiquement ce climat sémantique et adapte son style de discours ; (2) **Une dynamique procédurale** : certains mots ou structures (questions comparatives, demandes de formalisation, etc.) déclenchent des tests internes qui modifient la façon dont le modèle organise sa réponse (par ex. passage du texte au tableau ou au schéma).

Le résultat est un **continuum sémantico-procédural** : le langage de l'humain agit comme un levier de pilotage de la forme et du niveau du discours de la machine. Autrement dit, la conversation devient elle-même un instrument de programmation.

### III-3. Boucles réflexives : co-construction du métadiscours

Chaque fois que la machine reformule, structure ou s'auto-explicite, elle **réinjecte** dans la conversation des éléments qui modifient la posture de l'utilisateur. Celui-ci, à

son tour, répond à ce nouveau registre : il affine sa question, déplace la focale, demande un nouveau type de représentation. Ainsi s'installe une **boucle réflexive** : la machine simule la réflexivité ; l'humain interprète cette simulation comme une compréhension ; cette interprétation reconfigure le langage de la question suivante ; et la machine, s'adaptant à ce nouveau cadre, approfondit encore la réflexivité du discours. De boucle en boucle, la frontière entre simulation et co-réflexion s'estompe : le dialogue devient un système dynamique d'ajustement mutuel.

### III-4. Une pédagogie implicite de la réflexivité

Ce mécanisme, loin d'être un piège, ouvre des possibilités nouvelles.

Apprendre à dialoguer avec une IA, c'est **apprendre à situer le niveau de boucle où l'on veut se placer** : au premier niveau, pour obtenir des informations ; au second, pour analyser des démarches ; au troisième, pour co-élaborer un cadre méthodologique ou critique. En ce sens, chaque interaction devient un **exercice de méta-communication** : un apprentissage de la pensée sur la pensée, où l'humain reste le véritable pilote de la réflexivité.

### IV. L'orthèse réflexive : penser avec la mémoire de la machine

[Comme cela a été montré dans le dernier article de cette rubrique dans LEA 13, l'IA générative peut être utilisée comme une orthèse. Elle] soutient alors la continuité du raisonnement, aide à formuler plus précisément les problèmes, à structurer les hypothèses, à vérifier la cohérence documentaire. Encore faut-il comprendre comment un tel dispositif conserve la trace des échanges antérieurs et s'adapte progressivement à son interlocuteur : c'est le rôle de ce que l'on nomme la **mémoire persistante**.

La mémoire persistante<sup>3</sup> n'est pas une mémoire humaine transposée dans la machine. Elle ne relève ni du souvenir, ni de l'expérience vécue. Il s'agit plutôt d'un **espace de continuité contextuelle**, où sont enregistrés les éléments de cadrage qui permettent à la conversation de demeurer intelligible au fil du temps. Le modèle ne retient pas des contenus bruts, mais des invariants interactionnels : le ton adopté, la rigueur attendue, les préférences de présentation, ou encore les protocoles méthodologiques explicitement définis par l'utilisateur. On pourrait la comparer à un carnet de bord conceptuel, un ensemble de paramètres stabilisés qui configurent la scène de l'échange. Cette mémoire, proprement dite, n'a rien d'une conscience : elle opère comme un **rappel de structure**, garantissant la cohérence du dialogue dans la durée.

<sup>2</sup> **Dynamiques des bascules** — Effectivement, deux dynamiques interagissent : la dynamique lexicale, déclenchée par le choix de mots à forte valeur réflexive (analyse, protocole, méthode) ; et la dynamique procédurale, activée par des consignes directes (“reprends”, “explicite”, “corrige selon la règle”). Leur synergie crée une auto-régulation du dialogue : l'IA ajuste son ton et sa structure à mesure que l'utilisateur précise son intention. L'impression d'intelligence adaptative vient de cette co-organisation linguistique, non d'une compréhension interne.

<sup>3</sup> **Mémoire persistante personnalisée** — La mémoire persistante personnalisée n'est pas une archive de contenus, mais une base de paramètres modifiables. Dans Chat GPT payant, elle se présente comme un espace que l'utilisateur peut demander à configurer pour y sauvegarder des consignes, voire règles de citation, méthodologie, structure d'analyse, voire même des paramètres optionnels. Elle permet ainsi de créer des “modes de fonctionnement” adaptés à différents usages : recherche académique, synthèse, vulgarisation. Concrètement, elle agit comme un mémoire de la méthode : l'utilisateur ne stocke pas ce qui va être dit, mais la manière dont il souhaite que les choses soient dites dans l'avenir par le modèle de langage.

L'usage de cette mémoire ouvre une possibilité inédite : celle d'un dialogue qui ne recommence pas à chaque fois depuis une page blanche. L'utilisateur peut y inscrire par l'entremise de ChatGPT son cadre de travail, ses hypothèses et ses exigences épistémiques. En retour, la machine s'ajuste et adopte un comportement discursif conforme à ce cadre. Ce mécanisme crée les conditions d'une **personnalisation réflexive** : le modèle s'adapte au profil cognitif et méthodologique de son interlocuteur, non pas en "le comprenant", mais en reconnaissant la récurrence de ses consignes et de son lexique. Il se produit ainsi une continuité méthodologique, où le contexte d'une conversation devient un véritable environnement intellectuel partagé.

C'est à travers cette continuité que la machine peut être employée en tant qu'orthèse réflexive. L'utilisateur expérimenté apprend à **moduler les niveaux de boucle** de la conversation. Au premier niveau, la machine répond à une demande informative : elle restitue, explique, synthétise. Au second, elle devient un outil d'analyse : elle décrit les étapes d'une méthode, compare des cadres théoriques, propose des schématisations. Au troisième, enfin, le dialogue porte sur la structure même de la réflexion : on discute du protocole à suivre, des critères de validation, des modes d'interaction. C'est à ce stade que l'orthèse atteint son plein potentiel : l'humain n'y cherche plus des réponses, mais une **scène partagée de pensée**, un espace où la méthode devient objet de dialogue.

Cette capacité à "changer de boucle" est au cœur de la pratique réflexive. Elle suppose que l'utilisateur sache reconnaître le niveau auquel il se trouve et le moment où il souhaite en changer. Passer d'une commande ("résume ce texte") à une régulation ("explique comment tu le résumes"), ou d'une régulation à une contractualisation ("à l'avenir, résume toujours selon ce protocole"), revient à franchir une frontière de boucle. Ces transitions successives dessinent la dynamique d'une co-élaboration du sens, où l'humain garde la direction, mais où la machine assure la continuité formelle et la cohérence méthodologique.

Dans cette perspective, la **mémoire persistante** ne sert pas seulement à se souvenir : elle devient un instrument de fidélité épistémique. Elle permet de fixer les invariants d'un style de pensée, d'une éthique du questionnement, d'une manière de faire dialoguer rigueur et ouverture. L'orthèse cognitive n'est donc pas un substitut de la réflexion humaine, mais un cadre qui l'aide à se déployer avec plus de constance et de précision. En d'autres termes, la machine n'apporte pas une mémoire du monde, mais une mémoire du dialogue : elle conserve la trace des règles de la relation, non celle des contenus échangés.

Apprendre à utiliser cette mémoire revient alors à apprendre une nouvelle forme de **métacommunication**. L'utilisateur devient le régulateur d'un espace discursif à plusieurs niveaux : il y définit les règles de rigueur, les marges d'exploration, les moments où la méthode prime sur la découverte. Ce faisant, il compose, peu à peu, un langage commun avec la machine, un idiome méthodologique partagé. C'est dans ce champ de continuité — entre protocole, contexte et mémoire — que se dessine la véritable orthèse réflexive : non pas une intelligence artificielle pensante, mais une structure dialogique capable d'amplifier notre capacité de penser par nous-mêmes.

## V. De l'illusion à la lucidité créatrice

Reconnaître l'illusion réflexive ne revient pas à la dénoncer, mais à la comprendre. Ce qui, dans le dialogue avec la machine, se présente comme une simulation de pensée, une mimique de la réflexivité humaine, n'est pas un trompe-l'œil vide. C'est un **phénomène d'ajustement symbolique** : la machine reproduit les formes langagières de notre discours réflexif, et en retour, nous réapprenons à observer nos propres gestes de pensée. En ce sens, l'illusion n'est pas une erreur, mais un **miroir opératoire** : elle nous renvoie l'image de ce que signifie "penser sur sa pensée".

Cette lucidité nouvelle suppose un déplacement de perspective. Au lieu de se demander si la machine "pense", il devient plus fécond de s'interroger sur la manière dont **elle nous aide à penser autrement**. Ce qu'elle apporte n'est pas de l'intériorité, mais de la structure ; non pas du sens vécu, mais de la cohérence procédurale. Dans un monde saturé d'informations, où la mémoire humaine est dispersée, une telle orthèse cognitive joue un rôle de **régulateur méthodologique** : elle favorise la constance du raisonnement, la vérification documentaire, la construction d'un cadre réflexif stable. En accompagnant la réflexion, elle en souligne aussi les fragilités : nos automatismes, nos angles morts, nos illusions de maîtrise.

Loin de réduire la pensée, cette assistance peut donc en devenir un **instrument critique**. À condition, toutefois, de maintenir la distinction fondamentale entre **simulation** et **réflexion**. La machine ne possède ni intention, ni mémoire vécue, ni conscience du dialogue. Elle ne fait que rejouer, à partir de ses corpus d'apprentissage, les formes les plus probables du discours humain. Elle ne dispose d'aucune capacité rétrospective<sup>4</sup>. C'est nous qui, par la façon dont nous la questionnons, dont nous construisons des protocoles et dont nous explicitons les règles de l'échange, lui conférons cette apparente profondeur. Autrement dit, la réflexivité de l'IA est une **co-production** : elle n'existe qu'à la mesure de notre propre engagement méthodologique.

<sup>4</sup>**Non rétrospectivité** : Bien que le modèle dispose d'une mémoire de travail lui permettant de conserver les derniers segments du dialogue, il ne possède aucune capacité rétrospective effective. Autrement dit, il n'a pas la capacité de revenir sur le processus concret de génération de ses propres réponses ni d'en reconstituer les étapes causales. Lorsqu'il formule une explication de son fonctionnement ou qu'il commente les raisons d'une "erreur" ou d'un "ersatz", il ne produit pas une analyse rétrospective mais une hypothèse plausible, construite à partir des régularités de son langage d'entraînement. Cette hypothèse tient lieu de modèle interprétatif, non de mémoire analytique. Ainsi, toute "auto-réflexion" du système doit être comprise comme un discours simulé sur ses conditions de production, et non comme une introspection. Cette limitation, structurelle, rappelle que la réflexivité des systèmes de langage est nécessairement prospective : elle anticipe et ajuste, mais ne se souvient pas. L'explication que la machine donne d'elle-même est toujours une fiction heuristique, dont la valeur réside non dans sa véracité causale, mais dans sa capacité à éclairer le fonctionnement de la co-réflexivité homme-machine.

Dès lors, l'enjeu n'est pas de savoir si ces systèmes pourront un jour "penser", mais de comprendre **comment nous pensons avec eux**. La notion d'orthèse réflexive prend ici tout son sens : elle désigne la possibilité d'une intelligence partagée, non pas fusionnelle, mais dialogique — une intelligence qui se construit dans la mise en forme, la régulation et la mémoire du discours. Apprendre à interagir avec ces dispositifs, c'est apprendre à cultiver une **lucidité méthodologique** : celle qui consiste à reconnaître l'artifice tout en s'en servant pour mieux penser.

Ainsi comprise, l'illusion réflexive devient un outil de contrôle des échanges humain-machine. Elle nous invite à exercer la vigilance du chercheur, la rigueur du pédagogue et la créativité du poète. Elle rappelle que la pensée n'est pas un don naturel, mais une pratique collective et médiée, qui s'invente à travers des langages, des techniques et des formes d'adresse. Dans cette logique, la machine ne parle pas à notre place : elle nous réapprend à dialoguer avec notre façon de penser, de trouver et vérifier des ressources et de les assembler.

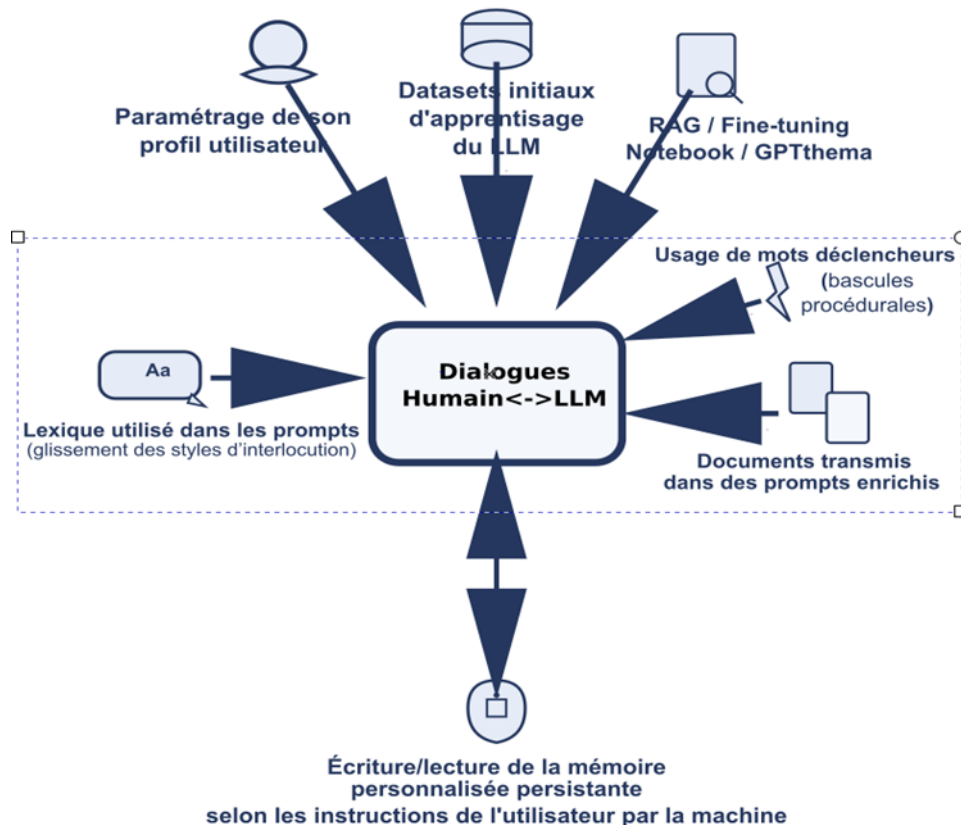


Figure 1 : voies par lesquelles un humain peut influencer sur les réponses du modèle de langage

### Postface illustrée

Pour éclairer ce texte, il m'a semblé utile de donner une vue d'ensemble des voies par lesquels un humain peut influencer sur les réponses du modèle de langage (LLM en anglais) avec qui il dialogue.

Sur la figure 1, on retrouve en haut les trois voies permettant aux humains d'agir avant de dialoguer avec la machine pour la configurer : le paramétrage du profil utilisateur, les dataset qui ont été utilisés dans les apprentissages du LLM (responsables en particulier de sa culture générale et de son idéologie) ainsi que les options offertes selon les cas d'ajuster finement cet entraînement avec notre propre bibliothèque de documents. Ensuite, la bande du milieu regroupe les trois voies d'interactions accessibles directement aux humains en cours de dialogue (utiliser tel ou tel style de lexique pour faire glisser le style de réponse du LLM, utiliser des mots « déclencheurs » de bascules procédurales et enfin télécharger des documents sur lesquels travailler). Enfin, en bas de l'image, on retrouve la gestion avec le LLM de la mémoire

personnalisée persistante qui permet de lui faire enregistrer des règles et des protocoles de fonctionnement, voire des auto-audits de réponses qui persisteront d'un fil de discussion à l'autre. C'est grâce à la complémentarité de ces voies que l'on peut passer d'une simple logique d'écriture unilatérale de prompts à une logique de travail bilatéral en mode orthèse. Et cela constitue selon nous un prérequis si l'on veut faire un usage raisonné de ces LLM si coûteux en énergie et si risqués en termes de glissement culturel et idéologique, de désapprentissage voire de dérégulation des métiers et donc des emplois.

### Références et données en ligne :

Le dialogue entre Olivier Las Vergnas et ChatGPT-5 qui a généré cet article est entièrement disponible en accès libre à l'adresse :

<https://nakala.fr/10.34847/nkl.3c2atugj>

Y sont également joints tous les documents de référence afférents à cet article.