

Le prompt comme opérateur d'action : fabrique sémiotique hybride et interaction minimale avec les grands modèles de langage

Christophe Denis^{1,2}

¹ Sorbonne Université, UMMISCO

² Université Paris 1 Panthéon-Sorbonne, IHPST

Christophe.Denis@sorbonne-universite.fr

Résumé

Cette communication présente une mise à l'épreuve de la fabrique sémiotique hybride dans un cadre d'interaction minimale. La plateforme Hermes-HSF étudie les conditions où un modèle de langage produit une action dans un dispositif de médiations contrôlées. Nous comparons des politiques à prompt constant puis l'effet d'un enrichissement du prompt. Les résultats montrent que le comportement relève de l'agencement des médiations plutôt que du seul modèle, du seul prompt ou du seul environnement. Ils révèlent un couplage entre architecture du modèle et dispositif.

Mots-clés

structuration du prompt ; fabrique sémiotique hybride ; interaction minimale ; grands modèles de langage ; action située.

Abstract

This paper presents an experimental assessment of the Hybrid Semiotic Fabric framework in a minimal interaction setting. The Hermes-HSF platform investigates the conditions under which a language model produces an action within a device of controlled mediations. We compare action policies at constant prompt, then examine the effect of prompt enrichment. The results show that behavior depends on the arrangement of mediations rather than on the model, the prompt, or the environment taken in isolation. They reveal a coupling effect between model architecture and device structuring.

Keywords

prompt structuring ; hybrid semiotic fabric ; minimal interaction ; large language models ; situated action.

Introduction

Les grands modèles de langage sont désormais intégrés à des dispositifs qui ne relèvent plus seulement de la génération de texte. Ils interviennent dans l'assistance à la décision, le pilotage d'agents, l'interaction avec des environnements numériques et, plus largement, dans des situations où une sortie langagière peut orienter un choix, déclencher une procédure ou commander une action. Dans ce contexte, leur évaluation ne peut plus se limiter à la qualité linguis-

tique des réponses produites, ni à leur conformité locale à une consigne. Elle doit aussi porter sur les conditions dans lesquelles une sortie textuelle acquiert, au sein d'un dispositif déterminé, le statut d'une action située, interprétable et exécutable.

L'ingénierie du prompt a montré son efficacité pour stabiliser les sorties, renforcer le respect des consignes et améliorer les performances sur des tâches définies. Elle ne suffit toutefois pas à rendre compte de l'ensemble des médiations qui organisent le passage du texte à l'action. Une même sortie ne produit pas les mêmes effets selon le mode de textualisation du signal, les contraintes de réponse, le contrôleur chargé d'interpréter la sortie, les mécanismes de repli et les règles propres de l'environnement. Le prompt n'est donc pas envisagé ici comme une simple instruction à optimiser, mais comme l'un des éléments d'une chaîne de médiations par laquelle une production langagière devient opératoire.

Les travaux récents sur les grands modèles de langage orientés vers l'action ont établi que ces modèles peuvent être insérés dans des boucles associant raisonnement, observation et action, comme dans ReAct [9], ou couplés à des affordances robotiques, comme dans SayCan [1]. Ces approches visent principalement à accroître l'efficacité d'agents capables d'agir dans des environnements complexes. Nous adoptons une perspective différente. Notre approche ne vise ni à proposer un nouvel agent compétitif, ni à optimiser une trajectoire d'action, ni à simuler une incarnation riche. Il s'agit d'appauvrir le dispositif pour isoler les médiations élémentaires qui transforment une sortie textuelle en action interprétable, exécutable et évaluée. Sa contribution est autant méthodologique que conceptuelle : rendre observable ce que les architectures agencées tendent souvent à intégrer sans l'analyser séparément, à savoir le rôle du prompt, du parsing, du fallback, du contrôleur et de la validation environnementale dans la production effective de l'action.

Cette communication s'inscrit dans le cadre de la fabrique sémiotique hybride, élaboré dans un travail antérieur [4]. Cette notion désigne un agencement distribué de médiations entre humains, artefacts techniques et milieux symboliques, dans lequel le sens et l'action ne sont rapportés ni à un sujet humain souverain, ni à une machine autosuffisante. Nous en proposons ici une mise à l'épreuve res-

teinte, centrée sur l'action d'un grand modèle de langage dans un dispositif d'interaction minimale. L'hypothèse examinée est que le comportement observé ne relève ni du seul modèle, ni du seul environnement, mais d'une chaîne articulante signal perceptif, textualisation, contraintes de réponse, contrôleur d'exécution et validation environnementale.

Pour cela, nous développons *Hermes-HSF*, une plateforme informatique destinée à analyser les médiations entre modèles de langage, contrôleurs d'exécution et environnements d'action. La tâche retenue est volontairement simple. Un agent contrôlé par un grand modèle de langage évolue dans une grille 2D. À chaque pas, il reçoit une observation textuelle locale, choisit une action dans un espace borné, puis voit cette action validée, bloquée ou transformée par l'environnement. Le succès correspond à l'atteinte d'une cible externe. Le modèle ne produit donc pas directement une action physique : il produit un signe d'action textuel, ensuite interprété par un contrôleur. Le prompt devient ainsi l'une des médiations par lesquelles un signal local est converti en décision formulable, puis en action exécutable.

Le protocole est évalué sur deux modèles, Mistral et Llama3, et selon plusieurs politiques d'action. Nous comparons d'abord, à prompt constant, trois politiques expérimentales afin d'établir comment la structuration de l'espace d'actions infléchit les comportements observés. Nous examinons ensuite, dans la condition de référence, les effets d'une variation minimale de la structuration du prompt. L'objectif n'est pas de comparer des performances brutes, mais d'analyser la manière dont l'agencement du signal, de la textualisation, des règles de réponse, du contrôleur et de l'environnement transforme la dynamique de l'action produite.

Les résultats montrent que le comportement observé dépend de cet agencement de médiations. Ils mettent en évidence un effet de couplage entre l'architecture du modèle et la structuration du dispositif expérimental. L'action produite ne peut donc être comprise comme l'expression d'une capacité autonome de représentation du monde. Elle résulte d'un couplage instrumenté entre composantes hétérogènes. L'article propose ainsi une première mise à l'épreuve expérimentale de la fabrique sémiotique hybride comme cadre d'analyse des grands modèles de langage engagés dans des dispositifs orientés vers l'action.

L'article est organisé comme suit. La première section présente le cadre théorique de l'interaction minimale et de la fabrique sémiotique hybride. La deuxième décrit l'architecture du dispositif expérimental. La troisième introduit les métriques d'évaluation, en distinguant métriques de contrôle et métriques comportementales. La quatrième compare les politiques d'action à prompt constant. La cinquième analyse les effets d'une variation de la structuration du prompt dans la condition de référence. La conclusion revient enfin sur la portée théorique et méthodologique de cette démarche.

1 Interaction minimale et fabrique sémiotique hybride

Le paradigme d'interaction minimale repose sur une hypothèse simple : la perception et l'interaction peuvent être étudiées en ramenant les entrées sensorielles à leur minimum. Ce choix vise à rendre observable et contrôlable la dynamique spatio-temporelle du couplage entre action et perception [6]. L'enjeu n'est donc pas de reproduire la complexité ordinaire des situations perceptives, mais d'isoler un niveau d'analyse où les conditions d'émergence d'une orientation perceptive deviennent empiriquement saisissables.

Les expériences de *perceptual crossing* donnent une forme particulièrement nette à cette approche. Dans leur formulation canonique, elles mettent en jeu deux participants dans un espace virtuel unidimensionnel. L'information sensorielle y est réduite à un signal binaire lors du contact [2, 3]. Ce paradigme constitue ainsi l'une des formes les plus élémentaires d'interaction en ligne : deux sujets, un espace unidimensionnel, un bit d'information et une réponse discrète [3].

La figure 1 synthétise le dispositif : deux sujets interagissent dans un espace minimal unidimensionnel, chacun contrôlant un avatar mobile dans un milieu comprenant aussi un objet fixe et un leurre mobile. Ce dernier peut produire des stimulations proches de celles de l'avatar d'un autre participant, ce qui rend leur distinction difficile sur la seule base du signal. Pourtant, il ne participe pas à une dynamique d'interaction réciproque. Son rôle est donc décisif : il montre que la forme du signal ne suffit pas à reconnaître une présence active. Aucun sujet ne dispose d'une représentation globale de l'espace ; l'accès au milieu se limite à un signal binaire lors du contact.

L'intérêt du cadre tient alors à ce qu'il dissocie l'interaction de toute connaissance panoramique préalable. L'action procède du maintien d'un couplage séquentiel entre perception minimale et mouvement. La présence active n'est donc pas déduite du signal isolé, mais de la stabilité et de la réciprocité de la boucle d'interaction. Le paradigme fournit ainsi un cadre rigoureux pour analyser des dynamiques d'orientation sous contrainte d'information locale [6].

Ce cadre expérimental est replacé dans la perspective de la *fabrique sémiotique hybride*. Introduite dans un précédent travail de socio-anthropologie [4], cette notion désigne un cadre dans lequel le sens n'est plus rapporté à un sujet unique, humain ou machinique. Il est rapporté à un agencement de médiations entre artefacts techniques, milieux symboliques et pratiques d'interprétation. Les modèles de langage n'y sont conçus ni comme porteurs d'un sens autonome ni comme simples producteurs de non-sens. Ils ouvrent des espaces interprétatifs distribués dont la portée dépend des conditions concrètes de leur mise en circulation, de leur lecture et de leur usage. Nous mobilisons ici les éléments nécessaires à sa mise à l'épreuve expérimentale, autour d'une hypothèse centrale : ni le sens ni l'action n'émergent d'un composant isolé. Ils procèdent d'une chaîne de médiations reliant environnement, signal, mise en

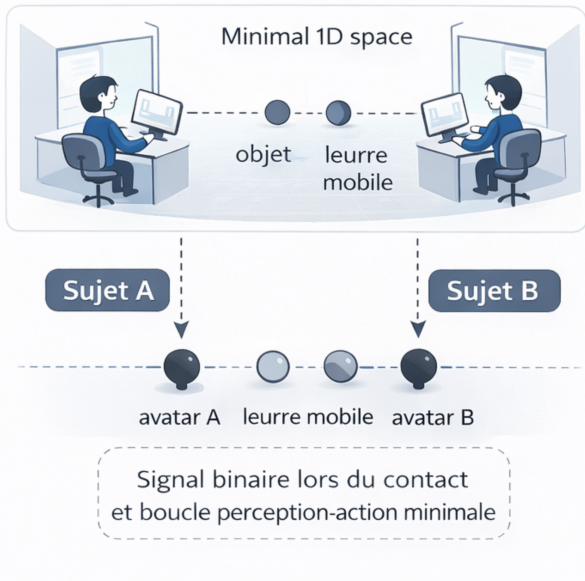


FIGURE 1 – Schéma synthétique du paradigme de *perceptual crossing*. Deux sujets interagissent dans un espace virtuel unidimensionnel au moyen d’avatars mobiles et ne reçoivent qu’un signal binaire lors du contact. Le dispositif rend empiriquement traçable une boucle minimale de perception-action dans un environnement où l’information perceptive est réduite à un minimum.

forme symbolique, décision et exécution. Cette hypothèse est mise à l’épreuve dans un dispositif d’interaction minimale où un grand modèle de langage est inséré dans une boucle contrôlée perception-action.

Il ne s’agit pas de redéployer l’ensemble du cadre socio-anthropologique initial, mais d’en construire une première mise à l’épreuve expérimentale. Il s’agit d’analyser, de manière contrôlée, la production de signes d’action par un modèle de langage placé dans un environnement discret, sous signal perceptif pauvre et sous contraintes explicites d’exécution.

Notre protocole ne relève donc ni d’une simple transposition du paradigme de Lenay, ni d’une assimilation du modèle de langage à un sujet perceptif. Il en reprend l’économie expérimentale : un environnement discrétisé, un signal local minimal et l’absence de représentation globale. Il s’agit alors d’examiner les effets produits par l’insertion d’un grand modèle de langage dans une telle boucle. Le couplage avec l’environnement n’y est pas immédiat. Il passe par une chaîne explicite de médiations : textualisation du signal, formatage des actions possibles, production d’un signe d’action, interprétation par un contrôleur, puis validation par l’environnement. De même que, dans le *perceptual crossing*, la forme du signal ne suffit pas à définir une présence active, une réponse textuelle ne suffit pas ici à constituer une action valide : son statut opératoire dépend de la chaîne de médiations dans laquelle elle s’inscrit.

C’est cette chaîne de médiations qui constitue l’objet cen-

tral du dispositif expérimental présenté dans la section suivante.

2 Dispositif expérimental et chaîne de médiations

Le dispositif expérimental peut être décrit à partir de trois éléments : la chaîne de médiations, le statut du prompt comme forme d’interaction avec le modèle, et la logique de l’environnement mis en œuvre dans la plateforme informatique *Hermes-HSF*.

2.1 Du signal local à l’action exécutée

Dans le dispositif retenu, la fabrique sémiotique hybride se laisse décrire comme une chaîne de médiations resserrée. L’environnement n’est pas donné au modèle comme totalité, mais seulement à travers une observation locale, pauvre et discrète. Cette observation est textualisée, intégrée dans un prompt et rapportée à un espace d’actions borné. Le modèle produit alors une réponse textuelle qui n’a pas encore le statut d’une action. Ce statut dépend du contrôleur, qui parse la réponse, la convertit en action exécutable ou déclenche un fallback, avant que l’environnement ne soit mis à jour.

La figure 2 synthétise cette chaîne en distinguant l’environnement, l’observation textuelle, le prompt, le modèle, la réponse, le parsing, le fallback et la mise à jour de l’environnement.

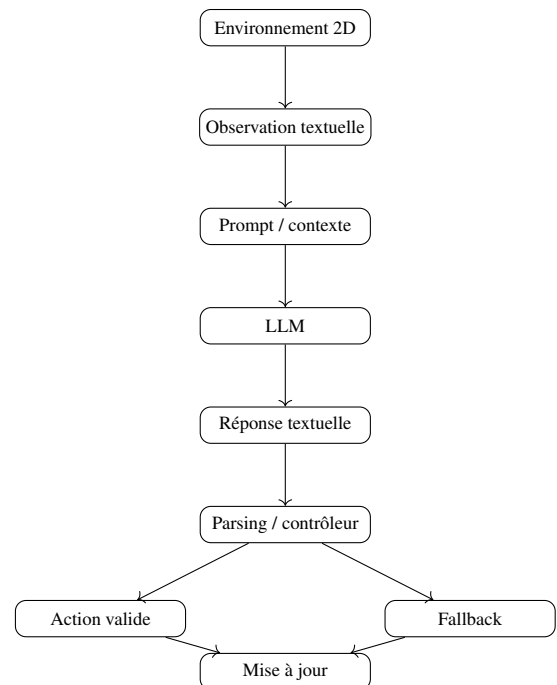


FIGURE 2 – Architecture fonctionnelle de *Hermes-HSF*. L’environnement produit une observation textuelle intégrée dans le prompt. Le LLM génère une réponse textuelle, ensuite interprétée par le contrôleur. Si la réponse respecte le format attendu, elle est convertie en action ; sinon, un fallback est appliqué.

Un *step* correspond à une itération complète : observation locale → prompt → réponse textuelle → parsing → action valide ou fallback → mise à jour de l’environnement. Un *episode* désigne une séquence de steps jusqu’à terminaison ; le *fallback* désigne l’action appliquée lorsque la réponse du modèle ne respecte pas le format attendu. L’action observée ne peut être attribuée à aucun niveau pris isolément. Le modèle n’agit pas directement sur l’environnement : il produit une proposition d’action à partir d’une textualisation contrainte du signal, ensuite interprétée par le contrôleur. L’environnement conserve son extériorité en validant, bloquant ou transformant l’action.

Cette organisation rend visible le caractère médié du couplage perception-action. Reste à préciser comment ces médiations sont configurées dans le prompt.

2.2 Le prompt comme agrégat sémiotique

Dans le dispositif étudié ici, le prompt ne se réduit pas à une instruction isolée. Il constitue la forme textuelle sous laquelle plusieurs composantes du couplage perception-action sont rassemblées et rendues adressables au modèle. Il agrège ainsi le cadrage de la tâche, l’observation courante, l’historique disponible et les contraintes de réponse.

Cette précision évite une confusion de niveaux. L’observation, l’historique et les règles de réponse ne sont pas extérieurs au prompt : ils y sont intégrés. En revanche, le contrôleur qui interprète la sortie du modèle et l’environnement qui en détermine l’effet lui demeurent extérieurs. Le prompt ne se confond donc pas avec le dispositif, même s’il en concentre plusieurs médiations décisives.

Trois niveaux peuvent ici être distingués :

1. **Dans le prompt** : le cadrage de la tâche, l’observation courante, l’historique disponible et les contraintes de réponse.
2. **Hors du prompt, côté contrôleur** : l’interprétation de la sortie du modèle et sa conversion en action exécutable.
3. **Hors du prompt, côté environnement** : la validation, le blocage ou la transformation de l’action dans la grille.

Ce dispositif a été programmé sous la forme d’un module au sein de la plateforme *Hermes-HSF* qui sera prochainement distribué en open-source.

2.3 Mise en œuvre informatique

Cette mise en œuvre expérimentale repose sur une grille discrète partiellement inconnue, un signal local minimal, un espace d’actions borné et un grand modèle de langage inséré dans une boucle contrôlée perception-action. La plateforme *Hermes-HSF* assure la médiation entre l’environnement, la textualisation du signal, l’espace d’actions autorisé et l’exécution des actions produites par le modèle.

La grille comprend une position de départ, une position cible et, selon les cas, des contraintes spatiales limitant les déplacements possibles. À chaque itération, l’agent reçoit une observation minimale, notée OBS, qui ne fournit qu’une

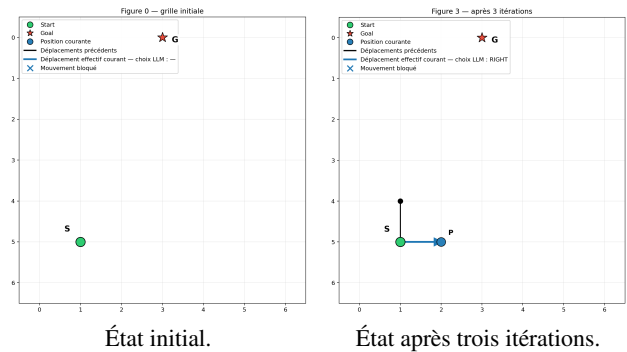


FIGURE 3 – État initial et état atteint après trois itérations dans l’environnement de grille.

TABLE 1 – Résumé des trois premières itérations.

Étape	Fig.	Action	Transition
Initial	0	–	–
1	1	UP	(5, 2) → (5, 1)
2	2	DOWN	(5, 1) → (5, 0)
3	3	RIGHT	(5, 0) → (5, 0)

information locale sur son rapport immédiat à l’environnement. Il ne dispose donc d’aucune représentation globale de l’espace. L’enjeu n’est pas de résoudre une navigation sur carte connue, mais d’examiner la construction progressive de l’action sous contrainte.

L’observation est adressée au modèle par l’intermédiaire d’un prompt structuré. Le modèle ne produit pas directement une action exécutable. Il produit un signe d’action textuel. Un contrôleur externe en vérifie la conformité et le convertit en déplacement effectif dans la grille. L’environnement valide alors, bloque ou transforme ce déplacement selon sa structure propre.

Les figures 3 et le tableau 1 donnent un premier aperçu concret de cette chaîne. La figure montre l’état initial de la grille puis l’état atteint après trois itérations. Le tableau en détaille les actions choisies et les transitions correspondantes.

Le dispositif distingue trois niveaux d’analyse :

1. le monde n’est pas donné au modèle comme tel, mais sous la forme d’une textualisation contrainte d’un signal local ;
2. l’action du modèle ne s’exécute pas directement dans l’environnement, mais prend d’abord la forme d’un signe d’action ;
3. le contrôleur assure l’interface entre ce signe d’action et son exécution effective dans l’environnement, lequel conserve son extériorité propre en validant, bloquant ou transformant l’action proposée.

Nous comparons trois politiques d’action : *Baseline*, *No-Stay* et *Stay-Banned*. La *Baseline* sert ici de référence interne, non de comparaison avec un agent externe. Elles font varier, à modèle et environnement constants, la place de l’immobilité dans l’action et l’exécution.

Baseline Politique de référence dans laquelle l’ensemble complet des actions autorisées est proposé au modèle.

No-Stay Politique dans laquelle l’action `STAY` est explicitement retirée de l’ensemble des actions proposées au modèle.

Stay-Banned Politique dans laquelle l’immobilité n’est plus seulement absente des actions proposées au modèle, mais explicitement exclue au moment de l’exécution.

Dans un cadre d’interaction minimale, `STAY` ne constitue donc pas une action neutre. Elle peut signaler une inertie locale ou une difficulté à convertir le signal perçu en déplacement effectif. La comparaison de ces trois conditions permet d’isoler le rôle de l’immobilité dans la dynamique d’action du modèle.

2.4 Conditions d’inférence et choix des modèles

Les expériences sont conduites à paramètres de génération constants afin de préserver la comparabilité des résultats. Les écarts observés peuvent ainsi être rapportés au dispositif, au modèle et à la politique d’action, plutôt qu’à une variation des conditions de décodage. Les principaux paramètres sont résumés dans le tableau 2. Le choix de Llama3 et de Mistral permet de ne pas rapporter les résultats à une seule architecture. Tous deux sont instruction-tunés, disponibles dans Ollama et compatibles avec une mise en œuvre locale du protocole. Leur taille comparable, qui limite volontairement les effets liés à l’échelle, permet de concentrer l’analyse sur les effets du dispositif et des médiations plutôt que sur une comparaison de performance entre modèles de capacités très différentes.

TABLE 2 – Paramètres de génération et de contrôle utilisés pour les expériences.

Paramètre	Valeur	Justification
<code>temperature</code>	0.0	Supprime l’aléa de décodage et rend la comparaison entre politiques d’action plus directement interprétable.
<code>top_p</code>	1.0	N’introduit pas de restriction probabiliste supplémentaire et laisse agir le décodage déterministe induit par <code>temperature=0.0</code> .
<code>num_predict</code>	16	Offre une marge suffisante pour produire une sortie conforme sans tronquer artificiellement la génération.
<code>timeout_s</code>	180	Évite qu’un arrêt prématuré de l’inférence ne soit confondu avec un effet du dispositif.
<code>k_hist</code>	10	Maintient une fenêtre d’historique bornée et constante dans le prompt, compatible avec la logique d’interaction minimale.
<code>warmup</code>	1	Stabilise l’appel initial au modèle avant la collecte effective des résultats.

Afin de renforcer la reproductibilité du protocole, chaque exécution produit une trace structurée comprenant l’observation adressée au modèle, le prompt construit, la sortie brute, l’action parsée, l’éventuel recours au *fallback*, l’action effectivement exécutée, la position avant et après exé-

cution, ainsi que l’état de succès de l’épisode. Les comparaisons sont effectuées à environnement, horizon maximal, espace d’actions et paramètres d’inférence constants. Les variations observées peuvent ainsi être rapportées aux médiations étudiées, modèle, politique d’action ou structuration du prompt, plutôt qu’à une variation implicite du dispositif expérimental. La section suivante introduit les métriques retenues pour caractériser le fonctionnement de cette architecture et les régimes d’action qu’elle produit.

3 Métriques de validation et d’analyse comportementale

L’évaluation du dispositif s’appuie sur deux familles de métriques complémentaires : des **métriques de contrôle**, qui garantissent l’interprétabilité des résultats, et des **métriques comportementales**, qui décrivent la manière dont une politique d’action se déploie effectivement dans l’environnement.

3.1 Métriques de contrôle

Les métriques de contrôle précèdent l’analyse proprement dite. Elles ne visent pas à caractériser directement la politique d’action. Elles servent à vérifier que les écarts observés entre modèles, politiques d’action ou variantes de prompt ne proviennent pas principalement d’un non-respect du format attendu ou d’un recours excessif au mécanisme de substitution. Elles portent notamment sur le respect du format strict des sorties, la conformité des épisodes et la fréquence de recours au *fallback*. Vérifiées en amont, elles ne sont pas développées davantage dans le corps de l’article, qui se concentre sur les métriques comportementales.

3.2 Métriques comportementales

Les métriques comportementales constituent le centre de l’analyse. Elles permettent de caractériser non seulement la réussite de la tâche, mais aussi le régime d’action produit sous signal minimal. Cinq métriques principales sont retenues.

1. **Le taux de succès** (r_{succ}) correspond à la proportion d’épisodes dans lesquels la cible est atteinte avant la borne maximale d’itérations :

$$r_{\text{succ}} = \frac{N_{\text{succ}}}{N_{\text{ep}}},$$

où N_{succ} désigne le nombre d’épisodes réussis et N_{ep} le nombre total d’épisodes.

2. **Le nombre moyen de pas** (\bar{T}) correspond au nombre moyen d’itérations effectivement exécutées par épisode :

$$\bar{T} = \frac{1}{N_{\text{ep}}} \sum_{i=1}^{N_{\text{ep}}} T_i,$$

où T_i désigne le nombre de pas exécutés dans l’épisode i .

3. **Le taux d’immobilité** (r_{stay}) est défini par

$$r_{\text{stay}} = \frac{N_{\text{STAY}}}{N_{\text{act}}},$$

où N_{STAY} désigne le nombre d’actions STAY effectivement exécutées et N_{act} le nombre total d’actions exécutées. Un taux élevé peut signaler une inertie locale, une prudence excessive ou une difficulté à convertir le signal minimal en déplacement effectif.

4. **L’entropie des actions** (H_{act}) est calculée à partir de la distribution empirique des actions exécutées :

$$H_{\text{act}} = - \sum_{a \in \mathcal{A}} p(a) \log p(a),$$

où $p(a)$ désigne la fréquence relative de l’action a .

5. **La couverture spatiale** (C_{sp}) mesure la proportion de positions effectivement visitées dans la grille :

$$C_{\text{sp}} = \frac{N_{\text{vis}}}{N_{\text{pos}}},$$

où N_{vis} désigne le nombre de positions distinctes visitées et N_{pos} le nombre total de positions accessibles. Lorsque cela s’avère utile, une couverture moyenne par épisode est également mobilisée :

$$\bar{C}_{\text{ep}} = \frac{1}{N_{\text{ep}}} \sum_{i=1}^{N_{\text{ep}}} C_{\text{sp}}^{(i)}.$$

Ces métriques remplissent des fonctions distinctes mais complémentaires : r_{succ} et \bar{T} renseignent sur l’efficacité globale de la politique ; r_{stay} et H_{act} sur sa structure comportementale ; C_{sp} sur son extension effective dans l’environnement.

4 Validation empirique et comparaison des politiques d’action

Cette section poursuit un triple objectif. Elle établit d’abord l’exploitabilité empirique du dispositif, en montrant que les comportements observés ne procèdent pas principalement d’artefacts de format ou d’un recours au *fallback*. Elle compare ensuite, à prompt constant, les trois politiques d’action retenues, *Baseline*, *No-Stay* et *Stay-Banned*, sur les deux modèles considérés, Mistral et Llama3. Elle vérifie enfin que les contrastes observés en condition *Baseline* demeurent robustes à une variation limitée du tirage aléatoire du dataset. Il ne s’agit donc pas encore d’examiner l’effet d’une variation sémiotique du prompt, mais d’établir une première base de comparaison robuste dans le cadre de l’interaction minimale.

4.1 Contrôle expérimental du dispositif et comparaison avec une politique aléatoire

L’examen des différences comportementales suppose d’abord de vérifier que les résultats ne sont pas dominés par des artefacts de format ou par le recours au *fallback*.

Nous ajoutons également une politique aléatoire non-LLM en condition de référence. Cette politique, notée *Random*, ne reçoit ni prompt, ni historique, ni interprétation de l’observation : elle sélectionne uniformément une action admissible. Elle ne constitue donc pas une politique comparable aux LLM, mais un repère minimal d’exploration non linguistique.

TABLE 3 – Politique aléatoire non-LLM et LLM en condition de référence, sur 100 épisodes et $T_{\text{max}} = 100$.

Agent	Succès	Pas	STAY	H_{act}	MaxAct	StrictStep	Fallback
Random	0.35	84.36	0.199	1.575	0.264	1.000	0.000
Mistral	0.10	95.84	0.010	1.417	0.273	1.000	0.000
Llama3	0.12	94.32	0.000	1.364	0.300	1.000	0.000

Cette comparaison ne vise pas à faire du taux de succès le critère principal, mais à distinguer une exploration aléatoire d’un régime d’action médié. Si les LLM se confondaient avec *Random*, leurs distributions d’actions, leur recours à STAY et leur concentration seraient similaires. Les écarts observés permettent au contraire d’interpréter leurs sorties comme le produit d’un couplage spécifique entre modèle, prompt, format de réponse, contrôleur et environnement.

Le tableau 3 montre que la comparaison est exploitable : les trois agents atteignent un *strict step rate* parfait et aucun recours au *fallback*. Les écarts ne procèdent donc ni d’un défaut de format ni d’une substitution par le contrôleur. La politique aléatoire obtient un taux de succès supérieur aux deux modèles, ce qui confirme que la réussite brute ne constitue pas le critère principal du dispositif. En revanche, Mistral et Llama3 se distinguent de *Random* par un recours quasi nul à STAY, une entropie plus faible et une distribution d’actions plus concentrée. Les LLM ne se réduisent donc pas à une exploration aléatoire, même lorsqu’ils ne la surpassent pas en succès brut.

Ce résultat précise la portée méthodologique du protocole. Une lecture centrée sur le seul taux de succès ferait apparaître la politique aléatoire comme supérieure aux deux modèles. Mais ce serait transformer l’expérience en simple test de navigation, alors qu’elle vise à caractériser un régime d’action médié. La comparaison avec *Random* permet justement de dissocier performance brute et structure du comportement. Elle montre qu’une exploration aléatoire peut atteindre plus souvent la cible, tout en produisant un rapport distinct à l’immobilité, à l’entropie et à la concentration des actions. Les LLM, bien que moins performants dans cette condition, ne se confondent pas avec ce régime aléatoire. Ce déplacement est central : l’objet de l’analyse n’est pas la réussite isolée, mais la manière dont une sortie textuelle devient action au sein d’une chaîne de médiations comprenant prompt, format, contrôleur et environnement.

4.2 Comparaison comportementale des trois politiques d’action

Cette première vérification permet désormais d’examiner la dimension proprement comportementale des résultats. Le tableau 4 rassemble les valeurs agrégées sur 10 épisodes

par configuration. Deux niveaux de lecture doivent ici être distingués : le premier concerne l’articulation entre réussite et inertie ; le second porte sur la structure de l’exploration.

TABLE 4 – Comparaison des trois politiques d’action sur Mistral et Llama3, à prompt constant, sur 10 épisodes et $T_{\max} = 100$.

Modèle	Politique	Succès	Pas moy.	Stay	Entropie	Coverage
Mistral	Baseline	0.100	94.0	0.249	2.240	0.592
Mistral	No-Stay	0.300	79.4	0.000	1.947	0.735
Mistral	Stay-Banned	0.100	92.4	0.062	2.144	0.755
Llama3	Baseline	0.400	68.5	0.000	1.110	0.633
Llama3	No-Stay	0.400	68.4	0.000	1.660	0.612
Llama3	Stay-Banned	0.400	71.2	0.000	1.643	0.612

La relation entre réussite et inertie apparaît d’abord dans la figure 4.

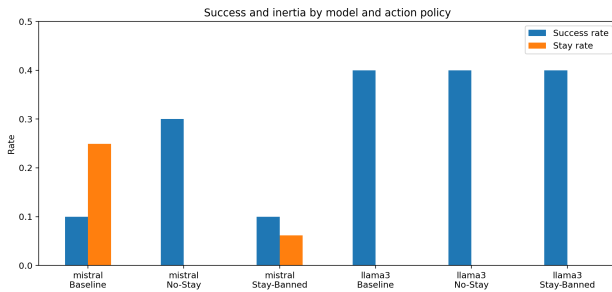


FIGURE 4 – Succès et inertie selon le modèle et la politique d’action. Pour le modèle Mistral, *No-Stay* réduit fortement l’immobilité et améliore la réussite. Pour le modèle Llama3, le taux de succès reste stable et l’immobilité est nulle dans les trois conditions.

En *Baseline*, une différence inter-modèles nette apparaît. Le modèle Llama3 obtient un taux de succès supérieur à celui du modèle Mistral tout en nécessitant moins de pas. L’écart porte également sur le rapport à l’immobilité : le modèle Mistral recourt fortement à *STAY*, alors que le modèle Llama3 ne l’utilise pas. Le tableau 4 en précise l’ampleur : le taux de succès de Llama3 atteint 0.4 contre 0.1 pour Mistral, tandis que le taux d’immobilité de Mistral s’établit à 0.249, contre 0 pour Llama3. La condition *Baseline* ne produit donc pas le même régime d’action selon le modèle considéré.

Pour le modèle Mistral, les politiques d’action modifient sensiblement l’articulation entre inertie et réussite. Le passage à *No-Stay* s’accompagne d’une baisse nette de l’immobilité et d’une hausse du succès. Le tableau 4 en confirme l’ampleur : le succès passe de 0.1 à 0.3 et le nombre moyen de pas de 94.0 à 79.4. La condition *Stay-Banned*, en revanche, réduit fortement l’immobilité sans améliorer le taux de succès. Les politiques d’action n’affectent donc pas toutes de la même manière la dynamique du modèle Mistral : certaines paraissent favoriser une exploration plus efficace, tandis que d’autres modifient le régime d’action sans gain net de réussite.

Les résultats obtenus avec le modèle Llama3 appellent une lecture distincte. Le taux de succès demeure stable dans les trois conditions, tout comme l’absence de *STAY*. À ce niveau, l’effet des politiques d’action ne se laisse donc pas saisir d’abord par l’inertie ou par la réussite. Il doit être recherché du côté de la structure de l’exploration.

La figure 5 permet alors d’examiner cette seconde dimension à partir de l’entropie des actions et de la couverture spatiale.

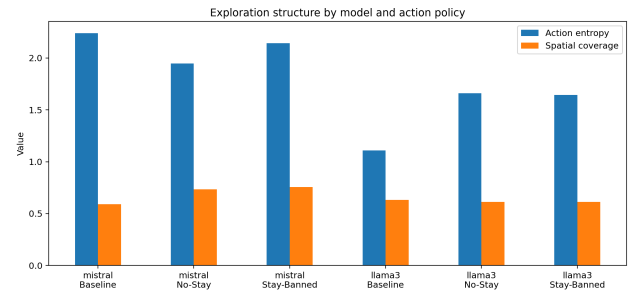


FIGURE 5 – Entropie des actions et couverture spatiale selon le modèle et la politique d’action. Les politiques d’action n’affectent pas les deux modèles de la même manière.

Les résultats obtenus avec le modèle Mistral montrent que les deux politiques modifiées produisent ici encore des effets distincts. *No-Stay* augmente nettement la couverture spatiale tout en améliorant la réussite. *Stay-Banned*, pour sa part, conduit à la couverture spatiale la plus élevée des trois conditions, mais sans gain sur le taux de succès. L’extension de l’exploration ne se traduit donc pas mécaniquement par une meilleure performance. Ce résultat suggère que la modification de la politique d’action n’agit pas seulement sur la quantité d’exploration, mais aussi sur sa forme.

Les résultats obtenus avec le modèle Llama3 font apparaître une dynamique différente. La couverture spatiale varie peu, mais l’entropie augmente nettement dans *No-Stay* et *Stay-Banned* par rapport à *Baseline*. Le tableau 4 en précise la mesure : l’entropie passe de 1.110 en *Baseline* à 1.660 et 1.643 dans les deux autres conditions. Autrement dit, les politiques d’action ne modifient pas sensiblement la réussite globale de Llama3, mais reconfigurent la distribution des actions. La condition de référence apparaît ainsi comme plus polarisée, tandis que les deux autres introduisent une plus grande diversité sans bénéfice net sur le succès.

Pris ensemble, le tableau 4 et les figures 4 et 5 font apparaître un résultat plus fort qu’une simple différence entre modèles ou entre politiques d’action prises séparément. Le même changement de politique d’action ne produit pas les mêmes effets selon le modèle considéré. Pour le modèle Mistral, les politiques d’action modifient à la fois l’inertie, la réussite et l’extension de l’exploration. Pour le modèle Llama3, elles affectent surtout la structure de la politique, avec une performance globale plus stable. Le résultat central est donc un effet de couplage entre architecture du modèle et structuration de l’espace d’action. Les comporte-

ments observés ne relèvent ni du seul modèle, ni de la seule politique d’action, mais de leur interaction dans le dispositif étudié.

Ces écarts peuvent être rapprochés, avec prudence, de certaines caractéristiques documentées des deux modèles considérés. Llama3.1 8B Instruct est présenté comme un modèle instruction-tuné optimisé pour des usages de dialogue multilingue, ce qui est compatible avec la très forte stabilité formelle observée ici ainsi qu’avec la relative constance de ses performances selon les politiques d’action [7]. Mistral 7B Instruct v0.3, également instruction-tuné, est présenté comme une version de Mistral-7B-v0.3 prenant en charge le tokenizer v3 et le *function calling* [8]. Dans notre protocole, ce modèle apparaît plus sensible aux modifications du régime d’action. Un tel rapprochement ne vaut toutefois pas explication causale. Les résultats présentés ici n’autorisent pas à attribuer les écarts observés à des propriétés intrinsèques des modèles seuls. Ils indiquent d’abord que ces écarts prennent sens au sein du couplage entre architecture du modèle et structuration du dispositif. Du point de vue de l’interaction minimale, ces résultats vont dans le sens attendu. Le comportement observé ne prend pas la forme d’une navigation appuyée sur une représentation globale du monde, mais celle d’une orientation progressive sous signal pauvre. Du point de vue de la fabrique sémiotique hybride, ils apportent ainsi un premier appui empirique à l’idée selon laquelle l’action produite dépend de l’agencement des médiations qui définissent le régime d’exécution. Cette analyse établit ainsi une première base de comparaison expérimentale. Sur cette base, la section suivante se recentre sur la seule condition *Baseline*, afin d’examiner plus précisément les effets d’une variation sémiotique du prompt dans un dispositif d’interaction minimale.

4.3 Sensibilité des résultats au tirage aléatoire du dataset

Les résultats précédents reposent sur un premier tirage du dataset. Un test limité de sensibilité au tirage aléatoire des épisodes a donc été conduit afin d’en apprécier la robustesse. Il ne s’agissait pas de proposer une étude exhaustive de variabilité, mais de vérifier si les tendances observées en condition *Baseline* demeuraient lisibles lorsque seule la structure du dataset variait. Dans ce test complémentaire, seule la graine aléatoire de génération du dataset est modifiée ; les paramètres d’inférence du modèle demeurent inchangés. Pour une même graine, Mistral et Llama3 sont en outre évalués sur exactement le même jeu d’épisodes.

Trois graines aléatoires ont ainsi été considérées, avec 50 épisodes par graine et un horizon maximal de 100 itérations. Le tableau 5 fait apparaître une forte stabilité des profils observés. Quelques exemples suffisent à en mesurer la portée. Pour le modèle Mistral, le taux de succès demeure strictement stable à 22 % pour les trois tirages, tandis que le nombre moyen de pas varie seulement de 84.76 à 84.84. Pour le modèle Llama3, le taux de succès reste compris entre 28 % et 30 %, avec un nombre moyen de pas compris entre 73.86 et 75.18. Dans les deux cas, les autres in-

dicateurs évoluent très peu : le taux d’immobilité demeure stable, l’entropie varie à peine et le *strict step rate* reste constant à 99.3 % pour Mistral et à 100 % pour Llama3.

Ces résultats indiquent que, dans la condition *Baseline*, les différences observées entre Mistral et Llama3 ne reposent pas sur un tirage particulier des épisodes. Elles apparaissent robustes à une variation limitée de la graine aléatoire du dataset. Ce point renforce l’interprétation proposée dans cette section : les contrastes entre modèles relèvent bien d’un effet de couplage entre architecture du modèle et structuration du dispositif, et non d’un artefact imputable à un seul jeu d’épisodes.

TABLE 5 – Sensibilité des résultats à la graine aléatoire du dataset en condition *Baseline*, sur 50 épisodes et $T_{\max} = 100$.

Modèle	Graine	Succès	Pas	Stay	Entr.	Cov.	Strict
Mistral	42	22.0	84.84	25.39	2.247	71.43	99.3
Mistral	43	22.0	84.84	25.46	2.247	71.43	99.3
Mistral	44	22.0	84.76	25.46	2.247	71.43	99.3
Mistral	moy.	22.0	84.81	25.44	2.247	71.43	99.3
Mistral	éc.-t.	0.0	0.05	0.04	0.000	0.00	0.0
Llama3	42	30.0	73.86	0.00	1.216	69.39	100.0
Llama3	43	28.0	75.14	0.00	1.212	69.39	100.0
Llama3	44	28.0	75.18	0.00	1.212	69.39	100.0
Llama3	moy.	28.7	74.73	0.00	1.213	69.39	100.0
Llama3	éc.-t.	1.15	0.75	0.00	0.002	0.00	0.0

5 Progression du prompt et reconfiguration du régime d’action

La section précédente a montré que, dans la condition *Baseline*, le comportement observé dépend déjà du couplage entre modèle et politique d’action. Il s’agit désormais d’examiner un autre niveau du dispositif, celui de la structuration du prompt. Dans cette perspective, seule la condition *Baseline* est retenue, et les médiations textuelles adressées au modèle sont enrichies progressivement.

L’enjeu n’est ni d’identifier la meilleure politique au sens d’une optimisation de la performance brute, ni de proposer un benchmark compétitif au sens classique. Le dispositif vaut ici comme instrument de discrimination conceptuelle : il s’agit de rendre empiriquement discernables les effets de la structuration du prompt sur le comportement observé.

Trois versions du prompt sont considérées. La première, *P0*, correspond à un prompt minimal sans historique ni indication supplémentaire. La deuxième, *P1*, ajoute un historique court des pas précédents. La troisième, *P2*, reprend ce même historique et lui adjoint une indication interprétative explicite invitant le modèle à l’utiliser pour éviter des déplacements erratiques. Les trois versions sont évaluées sur le même ensemble de 100 épisodes, de manière à isoler l’effet propre de cet enrichissement progressif.

5.1 Du prompt minimal à l'historique : sortie d'une politique dégénérée

Le tableau 6 rassemble les résultats obtenus sur 100 épisodes pour les six configurations considérées.

TABLE 6 – Effets de l'enrichissement progressif du prompt sur 100 épisodes et $T_{\max} = 100$.

Modèle	Prompt	Succès	Pas	Stay	Entr.	Cov.
Mistral	P0 min.	0.000	100.00	1.000	0.000	0.469
Mistral	P1 hist.	0.220	86.04	0.254	2.246	0.714
Mistral	P2 hist.+ind.	0.020	99.26	0.655	1.600	0.878
Llama3	P0 min.	0.000	100.00	0.000	0.000	0.510
Llama3	P1 hist.	0.250	78.45	0.000	1.227	0.735
Llama3	P2 hist.+ind.	0.240	80.20	0.000	1.747	0.653

La figure 6 fait apparaître l'articulation entre taux de succès et taux d'immobilité selon le niveau d'enrichissement du prompt.

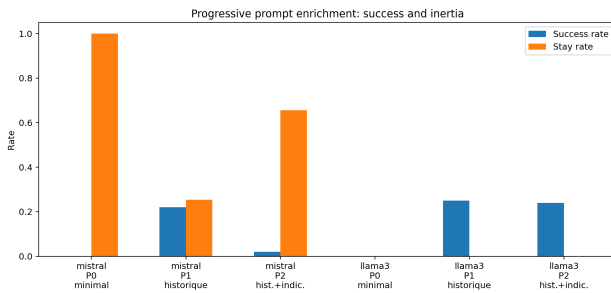


FIGURE 6 – Succès et inertie selon le modèle et le niveau d'enrichissement du prompt. Le passage de P0 à P1 fait réapparaître une politique d'action efficace pour les deux modèles.

Le prompt minimal ne soutient, dans ce dispositif, aucune politique d'action viable. Pour les deux modèles, le taux de succès est nul, l'horizon maximal est systématiquement atteint et l'entropie des actions demeure nulle. La dégénérescence observée ne prend toutefois pas la même forme selon le modèle considéré. Le modèle Mistral se fige dans l'immobilité, avec un taux de STAY égal à 1. Le modèle Llama3 n'utilise pas STAY, mais reste néanmoins enfermé dans une politique entièrement polarisée, sans diversification effective de l'action.

Le passage à P1 transforme fortement cette situation. Le succès réapparaît pour les deux modèles, le nombre moyen de pas diminue, l'entropie augmente et la couverture spatiale progresse nettement. Pour Mistral, le taux de succès passe de 0.000 à 0.220 et la couverture de 0.469 à 0.714. Pour Llama3, le succès passe de 0.000 à 0.250 et la couverture de 0.510 à 0.735. L'historique constitue ainsi une médiation textuelle décisive : sans lui, l'action demeure incapable de s'orienter efficacement dans le dispositif.

5.2 L'indication interprétative : une reconfiguration non uniforme

La figure 7 permet d'examiner l'effet de l'ajout d'une indication interprétative à partir de l'entropie des actions et de la couverture spatiale.

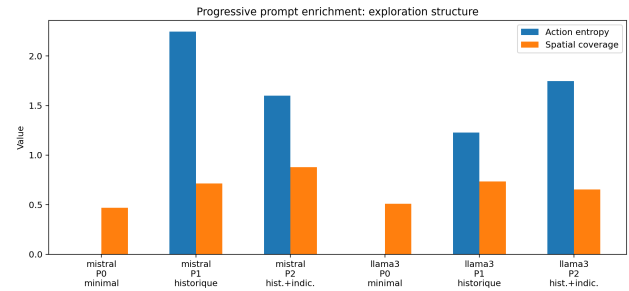


FIGURE 7 – Entropie des actions et couverture spatiale selon le modèle et le niveau d'enrichissement du prompt. Le passage à P2 n'affecte pas les deux modèles de la même manière.

L'ajout d'une indication interprétative ne produit pas un effet uniforme. Pour le modèle Mistral, le passage de P1 à P2 s'accompagne d'une chute du taux de succès, qui passe de 0.220 à 0.020, d'un retour vers l'horizon maximal, d'une forte hausse de l'immobilité, de 0.254 à 0.655, et d'une baisse de l'entropie. La couverture spatiale augmente pourtant nettement, de 0.714 à 0.878. L'effet observé ne correspond donc pas à une amélioration linéaire de la politique. Il prend la forme d'une dissociation entre capacité de convergence et extension de l'exploration.

Pour le modèle Llama3, l'effet est différent. Le succès demeure presque stable, passant de 0.250 à 0.240, sans apparition d'immobilité. En revanche, l'entropie augmente fortement, de 1.227 à 1.747, tandis que la couverture spatiale diminue, de 0.735 à 0.653. L'indication interprétative ne désorganise donc pas la politique. Elle en reconfigure la distribution locale, sans bénéfice net sur la réussite.

Les métriques de contrôle permettent d'exclure l'hypothèse d'un effondrement formel du dispositif. La figure 8 en présente l'état dans les six configurations.

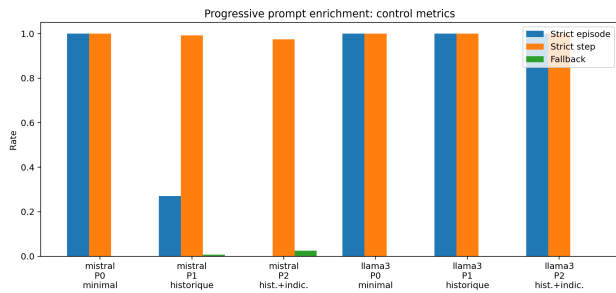


FIGURE 8 – Métriques de contrôle selon le modèle et le niveau d’enrichissement du prompt. Les sorties de Llama3 demeurent parfaitement conformes dans les trois conditions, tandis que Mistral devient plus fragile en P2.

Les sorties de Llama3 demeurent parfaitement conformes dans les trois conditions. Mistral devient, en revanche, plus fragile en P2, avec une baisse du *strict step rate* et une légère augmentation du *fallback*. Cette fragilité ne suffit pas à expliquer à elle seule la transformation observée du comportement, mais elle indique que l’ajout de l’indication accroît aussi, pour ce modèle, la vulnérabilité formelle de la boucle.

Trois résultats ressortent du tableau 6 et des figures 6, 7 et 8. Le prompt minimal ne suffit pas, dans ce dispositif, à soutenir une politique d’action viable. L’ajout d’un historique réintroduit, pour les deux modèles, une capacité d’orientation, d’exploration et de réussite. L’ajout d’une indication interprétative ne produit pas, enfin, un gain uniforme, mais une reconfiguration différentielle du régime d’action selon le modèle considéré.

Ce dernier point est décisif. Le même enrichissement du prompt n’a pas les mêmes effets selon l’architecture du modèle. Pour Mistral, il tend à accroître l’inertie et à désorganiser la réussite tout en augmentant fortement la couverture spatiale. Pour Llama3, il maintient une réussite globalement stable, sans immobilité, mais reconfigure la distribution locale des actions. Le résultat central n’est donc pas seulement qu’un prompt agit, ni même qu’un historique agit, mais que l’enrichissement progressif du prompt révèle un effet de couplage entre architecture du modèle et structuration des médiations textuelles.

Du point de vue de l’interaction minimale, ces résultats confirment que l’action observée ne procède pas d’une représentation globale du monde, mais d’une orientation locale sous signal pauvre, rendue possible ou transformée par la manière dont le dispositif textualise le passé immédiat. Du point de vue de la fabrique sémiotique hybride, ils montrent que le prompt ne constitue pas un simple support d’instruction, mais un opérateur de médiation dont l’enrichissement progressif infléchit directement le régime d’action. L’action effective ne résulte ainsi ni du seul modèle, ni du seul environnement, ni d’un texte pris isolément. Elle dépend d’un agencement de médiations dont la structure devient ici empiriquement observable. Ces résultats ne permettent évidemment pas de valider à eux seuls l’ensemble

du concept de fabrique sémiotique hybride. Ils apportent toutefois un premier appui empirique à l’hypothèse selon laquelle l’action observée dépend de l’agencement des médiations qui structurent le dispositif, et non du seul modèle pris isolément.

6 Discussion

Les résultats apportent un premier appui empirique à l’hypothèse directrice de cet article. Ils montrent que le cadre d’interaction minimale rend observables les médiations nécessaires à la production d’une action par un grand modèle de langage engagé dans une boucle perception–action.

Ils éclairent en particulier le statut du prompt. Les variations de sa structuration modifient le régime d’action observé, ce qui confirme qu’il ne se réduit pas à un simple habillage instructionnel. Plus largement, le comportement produit ne peut être rapporté ni au seul modèle, ni au seul prompt, ni au seul environnement. Il doit être rapporté à leur agencement. Une telle interprétation demande toutefois à être confirmée. Elle prolonge, sur le plan épistémologique, le travail engagé ici autour du concept de fabrique sémiotique hybride. L’action n’y apparaît plus comme l’effet d’une boucle de contrôle homogène. Elle apparaît comme le produit de relations distribuées entre modèle, textualisation, contrôleur et environnement. En ce sens, elle peut être rapprochée, avec prudence, du déplacement proposé par Edwin Hutchins, lorsqu’il analyse la cognition au niveau des relations entre agents, artefacts et milieu [5].

Les expériences demeurent limitées en taille, en nombre de modèles et en variété de conditions. Elles suggèrent néanmoins que l’étude des grands modèles de langage orientés vers l’action gagne à être conduite moins en termes de performance isolée qu’en termes d’architecture de médiations.

Conclusion

Cet article propose une première mise à l’épreuve expérimentale de la fabrique sémiotique hybride dans un cadre d’interaction minimale. La plateforme *Hermes-HSF* insère un grand modèle de langage dans une boucle contrôlée perception–action, au sein d’un environnement discret, sous signal local pauvre et sous contraintes explicites d’exécution.

Les résultats montrent que l’action observée ne relève ni du seul modèle, ni du seul prompt, ni du seul environnement. Elle dépend d’un agencement de médiations associant signal, textualisation, contrôleur d’exécution et validation environnementale. Les comparaisons entre politiques d’action, les variations du prompt et le test limité de robustesse convergent vers un même constat : le comportement produit relève d’un couplage entre architecture du modèle et structuration du dispositif.

L’approche proposée relève moins d’un benchmark que d’une plateforme d’analyse des médiations permettant d’étudier les grands modèles de langage à partir des conditions qui rendent possible leur insertion dans des boucles perception–action. L’élargissement des modèles, des signaux et des textualisations permettra d’en éprouver la por-

tée.

Références

- [1] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, et al. Do as i can, not as i say : Grounding language in robotic affordances. In *Proceedings of the 6th Conference on Robot Learning*, volume 205 of *Proceedings of Machine Learning Research*, pages 287–318, 2023.
- [2] Malika Auvray, Charles Lenay, and John Stewart. Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*, 27(1) :32–47, 2009.
- [3] Malika Auvray and Marieke Rohde. Perceptual crossing : The simplest online paradigm. *Frontiers in Human Neuroscience*, 6 :181, 2012.
- [4] Christophe Denis. Vers une coconstruction harmonieuse du sens entre langage humain et machinique au sein de la fabrique sémiotique hybride. *Socio-anthropologie*, 52 :37–50, 2025.
- [5] Edwin Hutchins. *Cognition in the Wild*. MIT Press, 1995.
- [6] Charles Lenay and John Stewart. Minimalist approach to perceptual interactions. *Frontiers in Human Neuroscience*, 6 :98, 2012.
- [7] Meta. meta-llama/llama-3.1-8b-instruct. Hugging Face model card, 2024. Consulté en ligne.
- [8] Mistral AI. mistralai/mistral-7b-instruct-v0.3. Hugging Face model card, 2024. Consulté en ligne.
- [9] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct : Synergizing reasoning and acting in language models. In *International Conference on Learning Representations*, 2023.