

Repenser le rôle des LLMs : de système de raisonnement central à source de connaissances pour les systèmes d'IA symboliques

Guillaume Sarthou*, Bastien Dussard*

*LAAS-CNRS, Université de Toulouse, CNRS
prenom.nom@laas.fr

Résumé. Au regard des développements actuels autour des grands modèles de langage (LLMs), ce résumé étendu vise à repenser le lien entre ces LLMs et les graphes de connaissances. Tandis que la majorité de la recherche actuelle considère les LLMs comme des composants centraux d'un système décisionnel, notre travail explore plutôt l'utilisation des LLMs comme sources de connaissances à structurer pour alimenter des modèles basés sur la connaissance, tel que des ontologies. Sous le prisme d'applications pour des agents artificiels tel que des robots, nous examinons les avantages et défis de cette approche.

Les grands modèles de langage (LLMs) possèdent naturellement un vaste répertoire de connaissances générales sur notre monde, grâce à la diversité de leurs jeux de données d'apprentissage. Cependant, il a été largement démontré que ces modèles présentent des limites en matière de fiabilité, d'explicabilité et d'efficacité lors de la résolution de problèmes complexes (Han et al., 2024; Jiang et al., 2024), et même dans le cas de problèmes simples de raisonnement logique (Ariyani et al., 2025). À l'opposé, les algorithmes d'intelligence artificielle (IA) symboliques offrent des solutions vérifiables, génèrent des explications transparentes de leur raisonnement (qu'il s'agisse d'un succès ou d'un échec) et atteignent un niveau d'efficacité remarquable. De plus, ces algorithmes sont souvent peu gourmands en ressources (comparés à leurs homologues basés réseaux de neurones), et sont donc une option de choix pour des applications embarquées nécessitant un niveau de sûreté accrue, tel que la robotique. Toutefois, ces systèmes nécessitent des connaissances structurées et cohérentes, ce qui représente un défi coûteux et chronophage à mettre en œuvre, notamment en raison de la nécessité de faire intervenir des experts pour modéliser ces connaissances spécifiques au problème considéré.

Les limitations intrinsèques des modèles de langage ne pouvant être résolus uniquement par l'augmentation illimitée du volume de données ou de la complexité des modèles, des systèmes d'IA hybrides (neuro-symbolique) ont été proposés comme une solution prometteuse. En combinant les connaissances générales et l'adaptabilité des LLMs avec la précision, l'explicabilité et l'efficacité de l'IA basée sur des modèles symboliques, ces systèmes hybrides ont déjà démontré leur capacité à améliorer la prise de décision, renforcer les capacités de résolution de problèmes et réduire la dépendance à l'ingénierie manuelle des connaissances (Sharma et al., 2025).

Une utilisation naturelle des LLMs est l'interfaçage avec l'utilisateur humain. Dans ce type d'utilisation, les raisonnements logiques sont effectués par un système basé sur la connaissance

Repenser le rôle des modèles de langage

tel que des ontologies ou des graphes de connaissance, puis le résultat de ces raisonnements est converti en langage naturel par un LLM. Grâce à cela, les raisonnements deviennent prouvables et des explications en langage formel peuvent être générées à faible coût. De plus, les raisonnements peuvent être effectués sur des graphes de grande envergure tout en assurant une validité logique au cours du temps. Le LLM vient alors combler le manque de lisibilité de ces raisonnements en apportant du sens commun. De plus, le LLM n'a pas besoin d'être informé de l'ensemble de la connaissance manipulée par le système et peut donc être nourri par la seule connaissance ayant été utilisée par le raisonnement logique. Cependant, bien que le raisonnement initial ait été validé en termes de logique formelle, il a été démontré que cette traduction de phrase sans raisonnement profond et sur un ensemble restreint de la connaissance peut, dans des cas non négligeables, être déformée et mal interprétée par le LLM, fournissant finalement des résultats erronés qui ne reflètent pas le raisonnement réalisé (Dussard et al., 2025).

Pour les raisonnements nécessitant du sens commun et ne pouvant pas être réalisés par des systèmes basés sur des modèles, le besoin de guider le LLM par rapport à une connaissance experte liée à une situation courante reste primordial. Des approches consistent alors à alimenter les LLMs avec des connaissances expertes, soit spécifiques à un problème donné, soit identifiées automatiquement comme pouvant être liées au problème à résoudre. De telles approches (comme le *Retrieval Augmented Generation*, RAG (Lewis et al., 2020)) aident à guider le traitement des LLMs en focalisant l'attention sur des connaissances utiles et pertinentes. Toutefois, ces méthodes restent confrontées à plusieurs limites. Premièrement, la qualité des connaissances récupérées dépend fortement de la pertinence et de l'exhaustivité des documents ou des données structurées fournies. Si le mécanisme de récupération échoue à extraire les informations les plus pertinentes (par exemple, on peut se demander si la similarité cosinus est toujours exacte), le LLM peut encore produire des résultats incorrects ou trompeurs. De plus, bien que le RAG améliore la cohérence factuelle en ancrant les réponses dans des connaissances externes, il ne traite pas naturellement l'efficacité du raisonnement ni la capacité à décomposer des tâches complexes en étapes logiques. De plus, ces approches basées sur la récupération peinent face aux domaines de connaissances dynamiques et évolutives, ce qui constitue un défi particulièrement critique en robotique, où l'adaptation en temps réel à un environnement changeant est essentielle. En somme, peu importe le niveau de logique fourni en entrée, l'utilisation de LLM en boucle ouverte demeure risquée et pourrait nécessiter de fermer la boucle avec une validation formelle par rapport à l'entrée initiale.

Face à ces problématiques, les recherches se sont orientées vers des paradigmes plus efficaces, notamment ceux qui déplacent les calculs de l'entraînement vers l'inférence, permettant ainsi une meilleure mise à l'échelle et une réduction des ressources nécessaires (Wu et al., 2025). Ces nouvelles approches se concentrent sur l'amélioration de la capacité des LLMs à raisonner de manière structurée, en introduisant des cadres de pensée étape par étape, ce qui transforme la manière dont les modèles traitent les informations et génèrent des réponses. Dans cette nouvelle génération de LLM, les techniques de raisonnement permettent de décomposer les problèmes complexes en sous-étapes intermédiaires, offrant ainsi une meilleure traçabilité des décisions prises. Ces méthodes peuvent être classées en deux grandes catégories (Snell et al., 2024) : la vérification des recherches, qui consiste à générer plusieurs échantillons de réponses et à sélectionner celle jugée la plus pertinente par un vérificateur (souvent un modèle spécialisé via *fine-tuning*), et la modification de la distribution, où les modèles sont entraînés à produire des étapes de raisonnement intermédiaires avant de fournir leur réponse finale. Ces

deux approches renforcent la robustesse des résultats obtenus en introduisant des mécanismes de validation ou de correction intra-processus. Cependant, malgré ces avantages, ces techniques restent probabilistes, car elles ne reposent pas sur une source de connaissance externe vérifiable. En l'absence de vérification formelle ou d'interface avec des bases de connaissances structurées, les erreurs peuvent se propager à travers les étapes de raisonnement. Ainsi, si un modèle commence par un processus de pensée incorrect, il peut s'attacher à cette erreur et la perpétuer, ce qui souligne la nécessité de combiner ces méthodes avec des stratégies de validation ou d'intégration de connaissances externes pour atteindre une meilleure fiabilité.

Dans cette nouvelle génération d'architectures, les agents LLM représentent une évolution majeure par rapport aux approches de raisonnement simples, en permettant aux systèmes d'utiliser des outils externes et de coordonner leurs actions pour résoudre des tâches complexes. Cette transition s'inscrit naturellement dans l'évolution des techniques de raisonnement, où les méthodes qui permettaient aux LLMs de décomposer les problèmes en étapes logiques sont maintenant intégrées dans des systèmes plus ambitieux capables d'interagir avec l'environnement et des sources de connaissances externes. Les agents intègrent des capacités d'auto-réflexion, de planification et d'orchestration de sous-tâches (Plaat et al., 2025), ce qui leur permet de structurer leur approche de manière plus systématique. L'usage d'outils spécifiques, tels que des bases de connaissances, des moteurs de raisonnement formel ou des outils de recherches sur le Web, permet de générer des résultats intermédiaires vérifiables et de garantir une certaine traçabilité du processus décisionnel. Cependant, malgré ces avancées, le cœur du raisonnement et de la synthèse des informations demeure probabiliste : les agents peuvent planifier des actions et orchestrer des outils, mais le processus de décision fondamental reste basé sur des distributions de probabilité et des générations de réponses non garanties. Les approches d'IA agentique (souvent qualifiées de "LLM augmenté") améliorent ainsi les chances de succès en structurant les processus de traitement, mais elles ne garantissent toujours pas une vérification complète des résultats. De plus, l'implémentation de ces systèmes nécessite souvent l'utilisation de multiples LLMs spécialisés, chacun chargé d'une tâche précise, ce qui pose des défis importants en termes de scalabilité et de complexité computationnelle. Cette limitation devient particulièrement critique pour les applications embarquées telles qu'en robotique, où les contraintes de ressources et de puissance de traitement rendent difficile l'implémentation de solutions basées sur de nombreux modèles spécialisés. Ainsi, bien que les agents LLM offrent une voie prometteuse pour améliorer les capacités des systèmes d'IA, ils restent soumis aux limitations fondamentales de la nature probabiliste de leur raisonnement central.

Dans ces approches actuelles, les grands modèles de langage sont intégrés comme des composants centraux des systèmes d'IA, gérant directement la chaîne de décision et la génération des réponses (ou des actions dans un cadre robotique). Cependant, malgré les résultats impressionnants, un changement de perspective devrait être étudié, en considérant les LLMs non plus comme le cœur décisionnel, mais comme des ressources cognitives supportant des systèmes fondés sur la connaissance structurée, suivant donc la voie ouverte par la communauté neuro-symbolique (Hammond et Leake, 2023). Cette perspective propose d'exploiter la robustesse, l'explicabilité et l'efficacité d'exécution des systèmes symboliques et des graphes de connaissances établis, tout en les complétant avec l'intuition de sens commun et de la richesse contextuelle encodée dans les LLM/MLM. Plutôt que de fonctionner comme un système de raisonnement central, le modèle de langage fournirait des heuristiques non-structurées et des inférences contextuelles qui sont difficilement modélisables par des règles formelles seules.

Repenser le rôle des modèles de langage

Cette synergie permettrait d'atteindre une meilleure généralisation et une adaptation plus rapide aux variations environnementales sans sacrifier la rigueur du processus décisionnel.

Néanmoins, l'efficacité de telles architectures repose sur la qualité et l'exhaustivité de la base de connaissances initiale, dont l'élaboration et la maintenance à grande échelle représentent une difficulté à adresser. Une direction de recherche complémentaire et stratégique consiste à exploiter les capacités des LLMs pour automatiser l'acquisition, la structuration et la validation de nouvelles connaissances structurées au sein du système. Contrairement aux efforts visant à corriger les lacunes des LLMs par les graphes de connaissances, cette approche vise à résoudre les lacunes des systèmes basés sur les graphes de connaissances par les LLMs. Les principaux défis seraient alors : (1) L'identification proactive des lacunes de connaissance par le moteur d'inférence du système. (2) L'interrogation augmentée (RAG interne) du LLM/MLM, où la requête est enrichie par les connaissances et règles existantes pour garantir la cohérence contextuelle. (3) La transformation des réponses non-structurées en schémas formels (par ex., triplets) et leur insertion dans le graphe de connaissance. (4) La vérification et l'identification de conflits ainsi que la résolution itérative de ces incohérences, potentiellement via des itérations avec le LLM. Au lieu de questionner régulièrement le LLM pour une même connaissance, ces systèmes pourraient donc apprendre au fil de son utilisation en réduisant progressivement leur recours au modèle de langage. Outre les bénéfices en termes de traçabilité et d'explicabilité, ce paradigme pourrait répondre aux impératifs d'opérabilité des systèmes décisionnels embarqués. En confiant le raisonnement critique à des structures optimisées, il favoriserait l'efficacité d'exécution et réduirait considérablement l'empreinte énergétique et computationnelle par inférence. Enfin, en limitant la dépendance à des inférences externalisées coûteuses et volumineuses, cette approche pourrait assurer la frugalité des systèmes décisionnels tout en privilégiant l'autonomie, réduisant donc la dépendance aux fournisseurs tiers. Cette autonomie opérationnelle renforcerait la capacité d'utilisation à long terme et faciliterait l'intégration d'une réflexion sur le respect la vie privée et la souveraineté des données, éléments cruciaux pour le déploiement généralisé et éthique des systèmes robotiques.

Références

- Ariyani, N. F., Z. Bouraoui, R. Booth, et S. Schockaert (2025). There's no such thing as simple reasoning for LLMs. In *Findings of the Association for Computational Linguistics : ACL 2025*.
- Dussard, B., A. Clodic, et G. Sarthou (2025). Evaluating embeddable language models in verbalizing rule-based inferences through justifications. In *International Conference on Robot and Human Interactive Communication (34th Edition-IEEE RO-MAN)*.
- Hammond, K. J. et D. B. Leake (2023). Large language models need symbolic ai. In *NeSy*, pp. 204–209.
- Han, S., H. Schoelkopf, Y. Zhao, Z. Qi, M. Riddell, W. Zhou, J. Coady, D. Peng, Y. Qiao, L. Benson, et al. (2024). Folio : Natural language reasoning with first-order logic. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 22017–22031.

- Jiang, B., Y. Xie, Z. Hao, X. Wang, T. Mallick, W. J. Su, C. J. Taylor, et D. Roth (2024). A peek into token bias : Large language models are not yet genuine reasoners. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 4722–4756.
- Lewis, P., E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, et al. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems* 33, 9459–9474.
- Plaat, A., M. van Duijn, N. van Stein, M. Preuss, P. van der Putten, et K. J. Batenburg (2025). Agentic large language models, a survey. *arXiv preprint arXiv :2503.23037*.
- Sharma, K., P. Kumar, et Y. Li (2025). Og-rag : Ontology-grounded retrieval-augmented generation for large language models. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 32950–32969.
- Snell, C., J. Lee, K. Xu, et A. Kumar (2024). Scaling llm test-time compute optimally can be more effective than scaling model parameters. *arXiv preprint arXiv :2408.03314*.
- Wu, Y., Z. Sun, S. Li, S. Welleck, et Y. Yang (2025). Inference scaling laws : An empirical analysis of compute-optimal inference for llm problem-solving. In *The Thirteenth International Conference on Learning Representations*.

Summary

In light of current developments surrounding Large Language Models (LLMs), this extended summary aims to rethink the relationship between LLMs and knowledge graphs. While most current research considers LLMs as central components of a decision-making system, our work explores their use as knowledge sources to be structured and used to feed knowledge-based models, such as ontologies. Through the lens of applications for artificial agents, such as robots, we examine the advantages and challenges of this approach.