RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

ONERA
THE FRENCH AEROSPACE LAB

www.onera.fr

PROPOSITION DE SUJET DE THESE

Intitulé : Embodied Chain of Thought for Enhanced and Safe Robotic Tasks

Référence : **TIS-DTIS-2026-**22
(à rappeler dans toute correspondance)

Début de la thèse : Dernier trimestre 2026	**Date limite de candidature :**

Thématiques scientifiques : Intelligence Artificielle et Décision, Robotique & Autonomie

Mots clés : Foundation Models for Robotics, Chain of Thoughts, Trustworthy AI

Profil et compétences recherchées :
Master 2 Recherche or Engineering Diploma, Specialization in Artificial Intelligence or Robotics

Compétences : Artificial Intelligence, Robotics

Outils : TensorFlow, PyTorch

Présentation du projet doctoral, contexte et objectif :

Context & Motivation:

Foundational models (i.e., large, pretrained neural networks) are now commonly used to solve robotic tasks [1, 2, 3, 4, 5], such as visual scene understanding, human interaction [6], action planning [4, 5, 10, 11, 12], manipulation [8], and navigation [9]. Many of these models use language to specify the task at hand, introducing contextual information, samples, references, or knowledge through "prompting" [13, 14].

Foundation models' language modality allows for the implementation of a Chain-of-Thought (CoT), which refers to the use of explicit, accessible, step-by-step reasoning to improve decision-making and problem-solving processes. The CoT is expected to produce relevant outputs and explain and justify them transparently. In recent chatbots, this is visible as the "thinking" phase that precedes the final answer generation phase. Most developments have been applied to large language models (LLMs), which are purely symbolic, as well as to large multimodal models (LMMs) [17, 18] and, more recently, to vision-language-action (VLA) models and presented as an embodied chain-of-thought (ECoT) [19, 20, 21], where reasoning is grounded in environmental observations.

Foundation models are also complex systems that may exhibit unexpected behaviors, including hallucinations [22], overconfidence, sensitivity to adversarial attacks [23,24], and domain shift. These behaviors pose critical safety concerns when the models are deployed in robots that interact physically with their environments [25,26,27]. One possible use of a CoT is to detect unintended behavior or enhance trust by documenting a robot's reasoning. However, the CoT paradigm has recently been questioned regarding its conditioning capacity [28, 29, 30, 31] and faithfulness [32, 33], raising concerns about its reliability as a source of information.

Objectives & Research Questions:

This PhD research is expected to make significant contributions to the field of robotics and embodied AI, addressing two types of issues:

- The development of an embodied CoT framework that combines symbolic reasoning with real-time robot actions and perception, emphasizing effective representation of embodied knowledge, applied to robotic tasks such as object search in an unknown environment or object manipulation.
- The exploitation of the embodied CoT for enhancing safety, either for better human-robot interaction or for monitoring robot behavior.

The work is expected to answer some of the following research questions:
How can the CoT framework integrate and process multimodal sensory data (visual, auditory, and tactile) to improve the robot's understanding of, interaction with, and learning in its environment? What constitutes good content and syntax for a CoT in robotics? What does it represent? How faithful is the embodied CoT? What type of information is encoded by the CoT, and how is it used in inference? What does it explain?

References

[1] Firoozi, R., Tucker, J., Tian, S., Majumdar, A., Sun, J., Liu, W., Zhu, Y., Song, S., Kapoor, A., Hausman, K., Ichter, B., Driess, D., Wu, J., Lu, C., Schwager, M., 2023. Foundation Models in Robotics: Applications, Challenges, and the Future. https://doi.org/10.48550/arXiv.2312.07843
[2] Hu, Y., Xie, Q., Jain, V., Francis, J., Patrikar, J., Keetha, N., Kim, S., Xie, Y., Zhang, T., Zhao, S., Chong, Y.Q., Wang, C., Sycara, K., Johnson-Roberson, M., Batra, D., Wang, X., Scherer, S., Kira, Z., Xia, F., Bisk, Y., 2023. Toward General-Purpose Robots via Foundation Models: A Survey and Meta-Analysis. https://doi.org/10.48550/arXiv.2312.08782
[3] Jang, D.-S., Cho, D.-H., Lee, W.-C., Ryu, S.-K., Jeong, B., Hong, M., Jung, M., Kim, M., Lee, M., Lee, S., Choi, H.-L., 2024. Unlocking Robotic Autonomy: A Survey on the Applications of Foundation Models. Int. J. Control Autom. Syst. 22, 2341–2384. https://doi.org/10.1007/s12555-024-0438-7
[4] Ma, Y., Song, Z., Zhuang, Y., Hao, J., King, I., 2024. A Survey on Vision-Language-Action Models for Embodied AI. https://doi.org/10.48550/arXiv.2405.14093
[5] Kawaharazuka, K., Matsushima, T., Gambardella, A., Guo, J., Paxton, C., Zeng, A., 2024. Real-World Robot Applications of Foundation Models: A Review. https://doi.org/10.48550/arXiv.2402.05741
[6] Hunt, W., Ramchurn, S.D., Soorati, M.D., 2024. A Survey of Language-Based Communication in Robotics. https://doi.org/10.48550/arXiv.2406.04086
[7] Intelligence, P., Black, K., Brown, N., Darpinian, J., Dhabalia, K., Driess, D., Esmail, A., Equi, M., Finn, C., Fusai, N., Galliker, M.Y., Ghosh, D., Groom, L., Hausman, K., Ichter, B., Jakubczak, S., Jones, T., Ke, L., LeBlanc, D., Levine, S., Li-Bell, A., Mothukuri, M., Nair, S., Pertsch, K., Ren, A.Z., Shi, L.X., Smith, L., Springenberg, J.T., Stachowicz, K., Tanner, J., Vuong, Q., Walke, H., Walling, A., Wang, H., Yu, L., Zhilinsky, U., 2025. $π_{0.5}$: a Vision-Language-Action Model with Open-World Generalization. https://doi.org/10.48550/arXiv.2504.16054
[8] Wang, Z., Zhou, Z., Song, J., Huang, Y., Shu, Z., Ma, L., 2024. Towards Testing and Evaluating Vision-Language-Action Models for Robotic Manipulation: An Empirical Study. https://doi.org/10.48550/arXiv.2409.12894
[9] Zhang, Y., Ma, Z., Li, J., Qiao, Y., Wang, Z., Chai, J., Wu, Q., Bansal, M., Kordjamshidi, P., 2024. Vision-and-Language Navigation Today and Tomorrow: A Survey in the Era of Foundation Models. https://doi.org/10.48550/arXiv.2407.07035Sapkota, R., Cao, Y., Roumeliotis, K.I., Karkee, M., 2025. Vision-Language-Action Models: Concepts, Progress, Applications and Challenges. https://doi.org/10.48550/arXiv.2505.04769
[10] Pertsch, K., Stachowicz, K., Ichter, B., Driess, D., Nair, S., Vuong, Q., Mees, O., Finn, C., Levine, S., 2025. FAST: Efficient Action Tokenization for Vision-Language-Action Models. https://doi.org/10.48550/arXiv.2501.09747
[11] Shi, L.X., Ichter, B., Equi, M., Ke, L., Pertsch, K., Vuong, Q., Tanner, J., Walling, A., Wang, H., Fusai, N., Li-Bell, A., Driess, D., Groom, L., Levine, S., Finn, C., 2025. Hi Robot: Open-Ended Instruction Following with Hierarchical Vision-Language-Action Models. https://doi.org/10.48550/arXiv.2502.19417
[12] Wang, B., Zhang, J., Dong, S., Fang, I., Feng, C., 2025. VLM See, Robot Do: Human Demo Video to Robot Action Plan via Vision Language Model. https://doi.org/10.48550/arXiv.2410.08792
[13] Arenas, M.G., Xiao, T., Singh, S., Jain, V., Ren, A., Vuong, Q., Varley, J., Herzog, A., Leal, I., Kirmani, S., Prats, M., Sadigh, D., Sindhwani, V., Rao, K., Liang, J., Zeng, A., 2024. How to Prompt Your Robot: A PromptBook for Manipulation Skills with Code as Policies, in: 2024 IEEE International Conference on Robotics and Automation (ICRA). Presented at the 2024 IEEE International Conference on Robotics and Automation (ICRA), pp. 4340–4348. https://doi.org/10.1109/ICRA57147.2024.10610784
[14] Liu, F., Fang, K., Abbeel, P., Levine, S., 2024. MOKA: Open-Vocabulary Robotic Manipulation through Mark-Based Visual Prompting. Presented at the First Workshop on Vision-Language Models for Navigation and Manipulation at ICRA 2024.
[15] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., Zhou, D., 2023. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. https://doi.org/10.48550/arXiv.2201.11903
[16] Chu, Z., Chen, J., Chen, Q., Yu, W., He, T., Wang, H., Peng, W., Liu, M., Qin, B., Liu, T., 2023. A Survey of Chain of Thought Reasoning: Advances, Frontiers and Future. https://doi.org/10.48550/arXiv.2309.15402
[17] Zhang, R., Zhang, B., Li, Y., Zhang, H., Sun, Z., Gan, Z., Yang, Yinfei, Pang, R., Yang, Yiming, 2024. Improve Vision Language Model Chain-of-thought Reasoning.
[18] Chen, Y., Sikka, K., Cogswell, M., Ji, H., Divakaran, A., 2024. Measuring and Improving Chain-of-Thought Reasoning in Vision-Language Models. https://doi.org/10.48550/arXiv.2309.04461
[19] Mu, Y., Zhang, Q., Hu, M., Wang, W., Ding, M., Jin, J., Wang, B., Dai, J., Qiao, Y., Luo, P., 2023a. EmbodiedGPT: Vision-Language Pre-Training via Embodied Chain of Thought. https://doi.org/10.48550/arXiv.2305.15021
[20] Zhao, Q., Lu, Y., Kim, M.J., Fu, Z., Zhang, Z., Wu, Y., Li, Z., Ma, Q., Han, S., Finn, C., Handa, A., Lin, T.-Y., Wetzstein, G., Liu, M.-Y., Xiang, D., 2025. CoT-VLA: Visual Chain-of-Thought Reasoning for Vision-Language-Action Models. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1702–1713.
[21] Zawalski, M., Chen, W., Pertsch, K., Mees, O., Finn, C., Levine, S., 2024. Robotic Control via Embodied Chain-of-Thought Reasoning. https://doi.org/10.48550/arXiv.2407.08693
[22] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A., Fung, P., 2023. Survey of Hallucination in Natural Language Generation. ACM Comput. Surv. 55, 248:1-248:38.
[23] Wang, T., Liu, D., Liang, J.C., Yang, W., Wang, Q., Han, C., Luo, J., Tang, R., 2024. Exploring the Adversarial Vulnerabilities of Vision-Language-Action Models in Robotics. https://doi.org/10.48550/arXiv.2411.13587
[24] Liu, A., Zhou, Y., Liu, X., Zhang, T., Liang, S., Wang, J., Pu, Y., Li, T., Zhang, J., Zhou, W., Guo, Q., Tao, D., 2024. Compromising Embodied Agents with Contextual Backdoor Attacks. https://doi.org/10.48550/arXiv.2408.02882
[25] Wu, X., Chakraborty, S., Xian, R., Liang, J., Guan, T., Liu, F., Sadler, B.M., Manocha, D., Bedi, A.S., 2024. Highlighting the Safety Concerns of Deploying LLMs/VLMs in Robotics. https://doi.org/10.48550/arXiv.2402.10340
[26] Zhang, W., Kong, X., Braunl, T., Hong, J.B., 2024a. SafeEmbodAI: a Safety Framework for Mobile Robots in Embodied AI Systems. https://doi.org/10.48550/arXiv.2409.01630
[27] Qi, Y., Kyebambo, G., Xie, S., Shen, W., Wang, S., Xie, B., He, B., Wang, Z., Jiang, S., 2024. Safety Control of Service Robots with LLMs and Embodied Knowledge Graphs. https://doi.org/10.48550/arXiv.2405.17846
[28] Goyal, S., Ji, Z., Rawat, A.S., Menon, A.K., Kumar, S., Nagarajan, V., 2024. Think before you speak: Training Language Models With Pause Tokens. https://doi.org/10.48550/arXiv.2310.02226

[29] Pfau, J., Merrill, W., Bowman, S.R., 2024. Let's Think Dot by Dot: Hidden Computation in Transformer Language Models. https://doi.org/10.48550/arXiv.2404.15758
[30] Shojaee, P., Mirzadeh, I., Alizadeh, K., Horton, M., Bengio, S., Farajtabar, M., 2025. The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity. https://doi.org/10.48550/arXiv.2506.06941
[31] Sui, Y., Chuang, Y.-N., Wang, G., Zhang, J., Zhang, T., Yuan, J., Liu, H., Wen, A., Zhong, S., Zou, N., Chen, H., Hu, X., 2025. Stop Overthinking: A Survey on Efficient Reasoning for Large Language Models. https://doi.org/10.48550/arXiv.2503.16419
[32] Arcuschin, I., Janiak, J., Krzyzanowski, R., Rajamanoharan, S., Nanda, N., Conmy, A., 2025. Chain-of-Thought Reasoning In The Wild Is Not Always Faithful. https://doi.org/10.48550/arXiv.2503.08679
[33] Zhao, C., Tan, Z., Ma, P., Li, D., Jiang, B., Wang, Y., Yang, Y., Liu, H., 2025. Is Chain-of-Thought Reasoning of LLMs a Mirage? A Data Distribution Lens. https://doi.org/10.48550/arXiv.2508.01191

Co-encadrement : Liming Chen, LIRIS, Ecole Centrale de Lyon (liming.chen@ec-lyon.fr)

Laboratoire d'accueil à l'ONERA	**Directeur de thèse**
Département : Traitement de l'information et Systèmes	Nom : Stéphane Herbin
Lieu (centre ONERA) : Palaiseau	Laboratoire : ONERA DTIS
Contact : Stéphane Herbin	Tél. : 01 80 38 65 59
Tél. : 01 80 38 65 73 Email : stephane.herbin@onera.fr	Email : stephane.herbin@onera.fr

Pour plus d'informations : https://www.onera.fr/rejoindre-onera/la-formation-par-la-recherche