At that point, *Agent* gathered environment understanding through a front camera image and had to determine the most appropriate response to ensure both safety and efficiency. The available *Tools* included replanning the route, driving forward (over the obstacle), using visual and auditory signals, or aborting the task.

Experiments To evaluate the *Agent*'s performance in agricultural automation, two experimental conditions were tested:

1. Language-only embodiment: a natural language description of the tractor, specifying its size, movement capabilities, and operational constraints;
2. Visual embodiment: an image (Figure 8) of the tractor, in addition to the description, offering it a direct visual reference of its physical attributes.

A range of multimodal LLMs (GPT-4o, GPT-4o mini, Claude 3.5 Sonnet) was tested as reasoning engines for the agent. The experiments focused on response to anomalies, specifically recognizing, classifying, and responding to obstacles in the environment.

Challenges The primary issue with the language-only embodiment was misjudging the durability of the tractor. Without a visual reference, the *Agent* was overly cautious, often stopping for small obstacles such as branches, incorrectly assuming they could cause damage.

The visually embodied *Agent* performed significantly better in assessing the tractor's physical capabilities. Both *Agents* occasionally misidentified objects, leading to errors in decision-making.

Results Providing the *Agent* with a visual reference of the tractor improved its obstacle assessment. With an image of itself, the *Agent* had a better grasp of its dimensions and durability. Unlike the language-only *Agent*, it correctly identified that small branches did not pose a threat. The setup including vision-based embodiment led to an increase in correct hazard classification.

Despite these improvements, both agents struggled with image-based object recognition. Occasionally, classification errors led to incorrect hazard assessments, unnecessary maneuvering, or aborting the mission. While the visual embodiment helped mitigate false positives, the underlying image processing remained a limitation.

4   Conclusions and Future Work

The goal of this work was to develop a scalable, extensible, and flexible system to implement embodied MAS in both simulated and physical environments. To achieve this, the RAI framework was introduced, designed to address common challenges in working with autonomous and semi-autonomous robotic agents.

The RAI's architecture supports scalability by allowing the addition of new *Agents* with minimal overhead. It is extensible, enabling users to develop custom *Tools* and functionalities, and flexible, simplifying integration across various deployment scenarios. These capabilities have been demonstrated through successful deployments on multiple robotic platforms, highlighting RAI's potential for embodied agent applications in diverse contexts, as demonstrated in Section 3. The framework has also proven valuable for experimenting with embodiment strategies and evaluating LLM capabilities.

The current version of RAI is available at `https://github.com/RobotecAI/rai`. It should be noted that the framework undergoes constant development. Further work will include expanding the components library (Section 2.4) and adding features meant to address limitations of LLMs (e.g. spatio-temporal database, or knowledge streaming). Contributions to its further development are welcome and requested.

References

1. Cetnarowicz, K.: M-agent architecture based method of development of multiagent systems. In: Proc. of the 8th Join EPS-APS International Conference on Physics Computing, ACC Cyfronet (1996)
2. Coleman, D., Sucan, I.A., Chitta, S., Correll, N.: Reducing the barrier to entry of complex robotic software: a moveit! case study. CoRR **abs/1404.3785** (2014), `http://arxiv.org/abs/1404.3785`
3. Douze, M., Guzhva, A., Deng, C., Johnson, J., Szilvasy, G., Mazaré, P.E., Lomeli, M., Hosseini, L., Jégou, H.: The faiss library (2024)
4. Duan, K., Suen, C.W.K., Zou, Z.: Marc: A multi-agent robots control framework for enhancing reinforcement learning in construction tasks (2023)
5. Duffy, B., Joue, G.: Intelligent robots: The question of embodiment (2000)
6. Guidi-Polanco, F., Cubillos, C.: An agent-based software framework for robotics and automation systems. In: Crisan, M. (ed.) Convergence and Hybrid Information Technologies, chap. 7. IntechOpen, Rijeka (2010). `https://doi.org/10.5772/9652`
7. Guzman, L., Morellas, V., Papanikolopoulos, N.: Robotic embodiment of human-like motor skills via reinforcement learning. IEEE Robotics and Automation Letters **7**(2), 3711–3717 (2022). `https://doi.org/10.1109/LRA.2022.3147453`
8. Hołda, P., Rachwał, K., Sawicki, J., Ganzha, M., Paprzycki, M.: Agents assembly: Domain specific language for agent simulations. In: Dignum, F., Mathieu, P., Corchado, J.M., De La Prieta, F. (eds.) Advances in Practical Applications of Agents, Multi-Agent Systems, and Complex Systems Simulation. The PAAMS Collection. pp. 487–492. Springer International Publishing, Cham (2022)
9. Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., et al.: A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. ACM Transactions on Information Systems **43**(2), 1–55 (2025)
10. Huang, W., Abbeel, P., Pathak, D., Mordatch, I.: Language models as zero-shot planners: Extracting actionable knowledge for embodied agents. In: International conference on machine learning. pp. 9118–9147. PMLR (2022)
11. Li, C., Chen, H., Yan, M., Shen, W., Xu, H., Wu, Z., Zhang, Z., Zhou, W., Chen, Y., Cheng, C., Shi, H., Zhang, J., Huang, F., Zhou, J.: Modelscope-agent:

Building your customizable agent system with open-source large language models. ArXiv **abs/2309.00986** (2023), `https://api.semanticscholar.org/CorpusID:261531214`

12. Lim, C.S., Mamat, R., Braunl, T.: Market-based approach for multi-team robot cooperation. In: 2009 4th International Conference on Autonomous Robots and Agents. pp. 62–67. IEEE (2009)

13. Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Li, C., Yang, J., Su, H., Zhu, J., et al.: Grounding dino: Marrying dino with grounded pre-training for open-set object detection. arXiv preprint arXiv:2303.05499 (2023)

14. Macenski, S., Martin, F., White, R., Ginés Clavero, J.: The marathon 2: A navigation system. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2020)

15. Macenski, S., Soragna, A., Carroll, M., Ge, Z.: Impact of ros 2 node composition in robotic systems. IEEE Robotics and Autonomous Letters (RA-L) (2023)

16. Palanca, J.: Smart Python Agent Development Environment. `https://github.com/javipalanca/spade` (Last accessed May 10th, 2022)

17. Reed, S., Zolna, K., Parisotto, E., Colmenarejo, S.G., Novikov, A., Barth-Maron, G., Gimenez, M., Sulsky, Y., Kay, J., Springenberg, J.T., et al.: A generalist agent. arXiv preprint arXiv:2205.06175 (2022)

18. Ren, T., Liu, S., Zeng, A., Lin, J., Li, K., Cao, H., Chen, J., Huang, X., Chen, Y., Yan, F., Zeng, Z., Zhang, H., Li, F., Yang, J., Li, H., Jiang, Q., Zhang, L.: Grounded sam: Assembling open-world models for diverse visual tasks (2024)

19. Rogers, T.E., Sekmen, A.S., Peng, J.: Attention mechanisms for social engagements of robots with multiple people. In: ROMAN'2006-The 15th IEEE Int. Symp. on Robot and Human Interactive Communication. pp. 605–610. IEEE (2006)

20. Royce, R., Kaufmann, M., Becktor, J., Moon, S., Carpenter, K., Pak, K., Towler, A., Thakker, R., Khattak, S.: Enabling novel mission operations and interactions with rosa: The robot operating system agent (2024)

21. Seredyński, D., Winiarski, T., Zieliński, C.: Fabric: Framework for agent-based robot control systems. In: 2019 12th International Workshop on Robot Motion and Control (RoMoCo). pp. 215–222 (2019). `https://doi.org/10.1109/RoMoCo.2019.8787370`

22. Wang, G., Xie, Y., Jiang, Y., Mandlekar, A., Xiao, C., Zhu, Y., Fan, L., Anandkumar, A.: Voyager: An open-ended embodied agent with large language models. arXiv preprint arXiv: Arxiv-2305.16291 (2023)

23. Xi, Z., Ding, Y., Chen, W., Hong, B., Guo, H., Wang, J., Yang, D., Liao, C., Guo, X., He, W., Gao, S., Chen, L., Zheng, R., Zou, Y., Gui, T., Zhang, Q., Qiu, X., Huang, X., Wu, Z., Jiang, Y.G.: Agentgym: Evolving large language model-based agents across diverse environments. ArXiv **abs/2406.04151** (2024), `https://api.semanticscholar.org/CorpusID:270285866`

24. Xu, Z., Liu, Y., Deng, G., Li, Y., Picek, S.: A comprehensive study of jailbreak attack versus defense for large language models. arXiv:2402.13457 (2024)

25. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., Cao, Y.: React: Synergizing reasoning and acting in language models. arXiv preprint arXiv:2210.03629 (2022)

26. Zhao, H., Chen, H., Yang, F., Liu, N., Deng, H., Cai, H., Wang, S., Yin, D., Du, M.: Explainability for large language models: A survey. ACM Trans. Intell. Syst. Technol. **15**(2) (Feb 2024). `https://doi.org/10.1145/3639372`

27. Zhu, H.: A role-based approach to robot agent team design. In: 2006 IEEE International Conference on Systems, Man and Cybernetics. vol. 6, pp. 4861–4866. IEEE (2006)