Using Large Language Models to Guide Purpose-Driven Grounded Learning in Autonomous Robots

Sergio Martínez-Alonso, Emanuel Fallas-Hernández, Alejandro Romero, Jose A. Becerra, and Richard J. Duro
GII, CITIC research center, Universidade da Coruña, A Coruña, Spain
{sergio.martinez3, emanuel.fallas, alejandro.romero.montero, jose.antonio.becerra.permuy, richard}@udc.es

***Abstract*—Large Language Models (LLMs) are increasingly used in autonomous robotics, often in end-to-end frameworks that map natural language instructions directly to actions. While effective in constrained settings, such approaches suffer from limited grounding, poor explainability, and weak knowledge reuse. We propose an alternative integration strategy in which LLMs act as a high-level cultural knowledge repository that informs, but does not control, a robotic cognitive architecture. All perception, learning, and action selection remain grounded in embodied interaction. We show that leveraging LLM-provided priors during exploration in previously unknown domains significantly accelerates the acquisition of grounded goals compared to intrinsically motivated exploration. Experiments in simulation and on a real robot demonstrate improved early learning efficiency, goal discovery, and interpretability. Our results suggest that embedding LLMs within hybrid cognitive architectures enables scalable, explainable, and open-ended robotic learning.**

***Index Terms*—large language models, human-robot interaction, grounding, explainability, robotics**

I. Introduction

Large Language Models (LLMs) have recently attracted considerable attention in robotics and Human–Robot Interaction (HRI) due to their ability to interpret and generate natural language, enabling intuitive communication of instructions, goals, and preferences [1]–[3]. This has led to a growing number of robotic systems that integrate LLMs as core components of their control pipelines, often adopting end-to-end paradigms in which language inputs are mapped directly to low-level actions [4], [5]. While effective in constrained scenarios, these approaches raise fundamental challenges related to grounding, explainability, and long-term knowledge reuse.

A central limitation of end-to-end LLM-based robotics is insufficient grounding in sensorimotor experience, as decisions are largely driven by linguistic priors rather than knowledge acquired through embodied interaction [6]. Moreover, the internal reasoning processes of LLMs remain opaque, hindering explainability and making it difficult to inspect, verify, or correct robot behavior in safety-critical or collaborative contexts [7]. Finally, most existing approaches lack explicit mechanisms for online learning and structured knowledge retention, limiting their ability to transfer and reuse knowledge across tasks and environments [8]. These issues are particularly problematic in open-ended and lifelong learning scenarios, where robots must continuously adapt to previously unknown tasks and conditions without relying on predefined curricula or static task specifications [9], [10].

Related work has explored several directions to mitigate these limitations. LLM-based systems such as SayCan [11], ProgPrompt [12], and Code-as-Policies [13] improve high-level reasoning or transparency but largely retain end-to-end control assumptions. Multimodal approaches such as PaLM-E [5] and RT-2 [14] extend perception and action capabilities, yet still struggle with grounding and knowledge transfer across contexts. Other efforts integrate reinforcement learning to improve grounding through interaction [15], but these methods are often sample-inefficient and difficult to scale to real-world, open-ended learning and they lack structured mechanisms for long-term knowledge retention [16].

Prior to LLM-based robotics, many different cognitive architectures were proposed to support structured reasoning, grounding, and lifelong learning [17]. In parallel, open-ended learning approaches based on intrinsic motivation and curriculum learning have been investigated to enable continual knowledge acquisition [18]–[20]. While principled, these methods often suffer from slow learning and strong dependence on carefully designed curricula, limiting their applicability in real-world autonomous robotics. Recent lifelong learning methods address knowledge retention and transfer [21], [22], but integrating human guidance in a grounded and explainable manner remains an open challenge.

In this work, we argue that many of the limitations of current LLM-based robotic systems arise from assigning LLMs a direct control role. Instead, we propose an integration paradigm in which the LLM is embedded within a cognitive architecture as a high-level cultural and social knowledge repository, rather than as an end-to-end action generator. The cognitive architecture remains responsible for perception, learning, and action selection through embodied interaction, ensuring proper grounding The LLM complements this process by providing structured priors that guide exploration, and suggest candidate goals or strategies, thus accelerating learning without supplant-

This work was funded by the European Union's Horizon 2020, research and innovation programme under GA 101070381 ('PILLAR-Robots - Purposeful Intrinsically-motivated Lifelong Learning Autonomous Robots'), by Xunta de Galicia (EDC431C-2021/39), by the Spanish Science and Education Ministry (PID2021-126220OB-I00), and the Ministry for Digital Transformation and Civil Service and Next-Generation EU/RRF (TSI-100925-2023-1), and CITIC (Xunta de Galicia - Convenio para o desenvolvemento de accións estratéxicas de I+D+i 2025-2026).

ing the robot's autonomous learning mechanisms.

All knowledge acquired through interaction is stored and refined within the cognitive architecture using bottom-up contextual associative learning mechanisms [23]. This enables explicit, interpretable representations that can be read as pseudo-rules, supporting explainability and facilitating knowledge reuse over the robot's lifetime. Thus, the main contribution of this paper is a procedure for leveraging LLM-extracted knowledge within a cognitive architecture to reduce the exploration space faced by a robot in new situations, thereby enabling more efficient open-ended and lifelong learning.

We validate the proposed approach in both simulated and real-world robotic scenarios. The results show that embedding LLMs as a cultural knowledge repository that provides priors to constrain exploration within a structured cognitive architecture significantly accelerates early learning, improves adaptability, and supports the reuse of grounded knowledge across contexts. Consequently, it provides a viable path toward autonomous, explainable, and lifelong robotic learning.

The code used to obtain the results of this paper is available in a public repository.[1] Furthermore, replication instruction can be found in the documentation website.[2]

II. Proposed Approach

A. Overview of the system

In this work, we extend the e-MDB cognitive architecture for autonomous robots (see [24], [25] for a more detailed explanation on its operation) with an LLM based knowledge repository. e-MDB is geared towards providing robots with lifelong open-ended learning capabilities, meaning that they can continuously acquire new knowledge and adapt to changing environments that were not predicted at design-time.

Functionally speaking, the e-MDB is built around an associative long-term memory (LTM) that stores all the knowledge and processes that make up the robot's cognitive capabilities. This includes skills (in the form of policies), utility and world models to deliberate, as well as purposes, drives and goals to guide the decision processes through a motivational engine that manages the priorities of the system. This motivational engine introduces a hierarchical structure of robot purposes, as internal representations of domain independent objectives acquired from humans in the form of language strings, and goals, as representations of domain-specific areas in state space that increase purpose satisfaction and which are discovered and acquired autonomously by the robot as it interacts within different domains. Robot purposes that are innately designed into the cognitive architecture are called needs, and those that are derived from human instructions when the robot is deployed are called missions. A simplified diagram of the knowledge structure of the e-MDB is presented in Fig. 1.

A unique aspect of the e-MDB is how contextual information is stored. Whenever the robot experiences some reward (a drive value is reduced) in a new situation, it creates a context unit (C-node) linking the last executed skill, the active world model (representing its domain of operation), and the active goal. This unit is associated with a new or existing perceptual class (P-node) that encodes an area of the robot perceptual space that includes the current perceptions of the robot. Later, if the robot encounters a similar situation, the P-node is activated and can trigger the same skill, provided the same goal and world model are active. Depending on the success of the execution of this skill, the system will reinforce the P-node activation or exclude that perception from its activation space [26], [27].

With regards to procedural knowledge, e-MDB encapsulates its actions within policies (in the RL sense). Procedures can be both operational, i.e. leading to actions by the robot actuators, or cognitive, i.e. leading to actions over the knowledge in the LTM of the cognitive architecture itself. For instance, learning algorithms or intrinsic motivations, such as novelty [28] or curiosity [29], are instantiated as cognitive policies. One per learning algorithm in the case of learning and one per intrinsic motivation executor in the case of intrinsic motivations. These policies are triggered by context and executed in the same way as physical actions. However, it is important to take into account that in addition to perceptions related to external sensing, cognitive robots must consider perceptions coming from internal metacognitive variables that provide information on the operation of the cognitive structures.

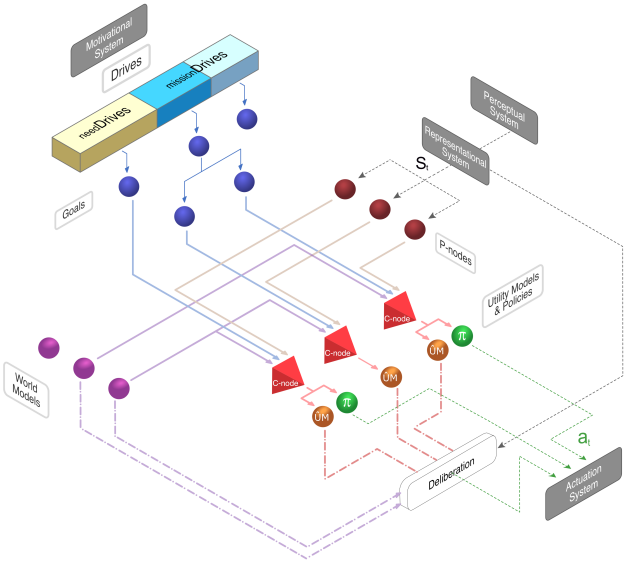

Fig. 1. Simplified diagram of the structure of the e-MDB LTM. The robot's decision-making is driven by the most active C-node, defined by the triplet {World Model, Goal, P-node}, which represents the current context. This C-node determines the next action.

B. LLMs as a knowledge injection mechanism

Bearing in mind the way the e-MDB cognitive architecture works, we propose the implementation of an LLM-guided

[1] https://github.com/pillar-robots/wp5_gii

[2] https://docs.pillar-robots.eu/projects/emdb_experiments_gii/en/latest/experiments/fruit_shop_experiment.html

cognitive policy that will enable the robot to populate the long-term memory (LTM) with initial hypotheses about how to accomplish its purposes.

From a motivational perspective, this cognitive policy is linked to an exploration need. It will be triggered when the architecture lacks sufficient knowledge to determine the next policy to execute. In such cases, exploration is required until a goal or a known situation is found. The LLM will guide this exploration towards the robot's purpose, by providing exploratory actions towards possible related goals that make sense for the domain the robot is in. In other words, instead of exploring randomly or using classical intrinsic motivations to guide the discovery of possible goals in the current domain that are appropriate to fulfill the purpose, the cognitive architecture, through a cognitive policy activated by the exploration need, will query the LLM for an initial set of possible goals to test and possible procedures to reach them. Based on the general knowledge contained in the LLM (which acts as a cultural repository, like grandparent or parent guidance for a child), it will provide a set of hypotheses on goals and actions towards them that will generally reduce the search space very significantly. As the robot is actually using its own sensorimotor system to explore whether these goals are valid by executing policies to reach them in the real domain, when these are successful, it will be able to create grounded representations for them linked to the domain and to its perceptions. Additionally, as this knowledge will be internally represented as e-MDB pseudo-rules (C-node type associations) in the LTM, explainability will be facilitated.

In short, a cognitive policy for this LLM guided exploration (see Fig. 2) takes as input a string corresponding to the robot purpose. This string is used to build a query to the LLM (see the code repository for an example) that includes: (1) the robot purpose (the aforementioned input); (2) a description of the perceptions received by the robot; (3) the set of policies available in LTM at that point in time, with an explanation of their behavior; (4) the most recent episodes (an episode is composed by: perceptions at time t, action taken, reward obtained, and perceptions at time $t+1$); and (5) the desired format for the answer, a suggested next policy. When the cognitive policy receives the answer from the LLM, it activates the suggested policy for execution. Finally, if the execution succeeds and the robot receives reward, a new goal and context (C-node and P-node) are created in LTM, thereby confirming the utility of the answer provided by the LLM and grounding that new knowledge (see [26] for details about how these nodes are created and managed). The pseudo-code of the behavior of this policy is shown in Algorithm 1.

C. Guided Exploration, explainability and Lifelong Learning

The proposed mechanism enables purpose-driven exploration by activating an LLM-guided cognitive policy whenever the robot encounters a domain in which no goals are associated with the current purpose or the means to reach them are unknown. In such situations, the policy initiates the procedure described in Section II-B, using the LLM to suggest candidate goals and action sequences that are meaningful for the current domain. If this process fails, the cognitive architecture can fall back on intrinsic motivation–based strategies, such as novelty [28] or competence improvement [30]. In typical scenarios, however, the LLM-guided policy substantially accelerates the discovery of goals and action sequences that satisfy the robot's purpose, effectively replacing low efficiency exploration strategies with more efficient exploration guided by the cultural knowledge encoded in the LLM.

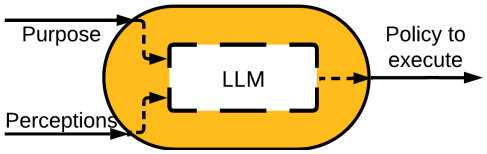

Fig. 2. Simplified diagram of the LLM-guided policy. The LLM receives as input a natural-language description of the purpose and the robot's current perceptions, and outputs an activation signal to the policy to execute.

Algorithm 1 LLM-Guided Cognitive Policy Execution

Input the robot purpose H
Input a description of the robot's perceptions P
Specify available policies Π
Define the episode format: $(P_t, A_t, r_{t+1}, P_{t+1})$
Set maximum number of episodes M
Initialize episode list E
while true **do**
    Add the last episode to E
    if $|E| > M$ **then**
        Remove the oldest episode from E
    end if
    Send (H, P, Π, E) to the LLM
    Receive response R from the LLM
    Execute policy R
end while

Importantly, the LLM does not replace the learning mechanisms of the cognitive architecture, but rather biases and guides exploration while preserving grounding through embodied sensorimotor interaction. Knowledge is acquired only through execution and validation in the real environment and is stored in the robot's Long-Term Memory (LTM) as pseudo-rules encoding contextual associative relationships. LLM-generated suggestions are consolidated only when they lead to reward, ensuring grounding. Each stored context can be interpreted as an explicit rule: *"if perception P_i is observed, the active goal is G_j, and the robot is in domain k (the current world model is W_k), then execute skill S_l"*, supporting explainability. When created under LLM guidance, these rules can also be linguistically annotated, further enhancing transparency.

By selectively reusing relevant policies stored in LTM when new purposes arise and disregarding irrelevant ones, the LLM-guided cognitive policy supports lifelong learning. This mechanism enables continual adaptation across tasks and domains while preserving grounded, interpretable, and reusable knowledge throughout the robot's lifetime.

III. Robotic experiment

The proposed approach was empirically validated in several different setups. Given the limitation in paper extension here, we will only present the results of one very representative robotic experiment. It is conducted using a TIAGo++ mobile manipulator controlled by the e-MDB cognitive architecture. As shown in Fig. 3, the robot is positioned in front of a table containing fruits of different types and a set of objects typically found in a fruit stand. The experimental setup includes a red collection area for fruits, a blue weighing area equipped with a scale, two baskets used to sort fruits based on weight, and a button that is unrelated to the classification task and serves as a distractor.

The experiment is structured into four sequential stages corresponding to distinct purposes in the robot's operational lifetime. At each stage, the robot is presented with a new purpose and is required to acquire or refine different components of its knowledge while potentially leveraging previously learned knowledge to achieve the purpose. The four stages are defined as follows:

1) *Interaction with objects*: The robot learns basic interaction primitives with objects in the environment, including grasping a piece of fruit and pressing a button. During this stage, the scale is deactivated to isolate object manipulation skills.
2) *Placing a piece of fruit*: The robot is asked to place a piece of fruit at a specified location, namely the center of the table, building upon the manipulation skills acquired in the previous stage.
3) *Exploration with an active scale*: The robot is asked to interact with the scale by placing a piece of fruit on it, thus acquiring knowledge about the weighing process.
4) *Fruit classification*: The robot learns to sort fruits according to their weight, explicitly reusing and combining knowledge acquired in the earlier stages to fulfill this more complex purpose.

This setup enables a systematic evaluation of knowledge reuse and transfer across different learning contexts. The durations of each stage are 1000, 2000, 1000 and 8000 iterations respectively. The environment is reset after each correct classification or 20 failed attempts.

Since the focus of this experiment is on evaluating the effectiveness of the proposed approach for efficient goal exploration using LLMs as a cultural knowledge repository, we assume that lower-level components, such as action policies, have been previously learned. This assumption allows us to isolate and analyze the contribution of LLM-guided exploration to goal discovery and learning. Accordingly, the robot has access to the eight pre-learned policies listed in Table I, which it can invoke to interact with the environment during the experiment.

For comparative evaluation, we also run the experiment using a simple novelty-based exploration strategy over the policy space as a baseline. Additionally, since the reward signal in this task is highly delayed, we also consider an auxiliary baseline in which the robot receives intermediate rewards for each correct sub-step toward task completion. Although such intermediate feedback is not available in real-world open-ended learning scenarios, this setting provides an upper-bound reference for learning efficiency and serves as a proxy for near-optimal performance. All experimental conditions are executed using the same cognitive architecture, policies, and environmental setup, ensuring that observed performance differences can be attributed solely to the exploration and learning strategies under comparison.

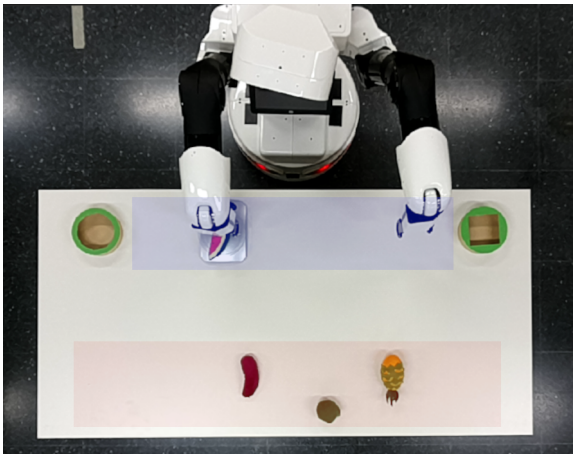

Fig. 3. Experimental setup with a TIAGo++ robot.

The robot's perceptual state is derived from multiple sensors. An RGB-D camera mounted on the ceiling detects the positions of the scale and fruits in polar coordinates, with angular values from -1.4 to 1.4 radians and radial distances from 0.2 to 1.9 meters. The camera also estimates fruit dimensions, with sizes from 0.03 to 0.1 meters, and detects the activation state of a button light (binary values for on or off). In addition, the robot maintains a binary signal indicating whether its gripper is currently holding a piece of fruit.

The scale provides two additional sensory signals: one indicating whether the scale is actively weighing a fruit (binary), and another encoding the validity of the weighing outcome, with discrete values corresponding to *not weighed*, *valid*, and *invalid*. All perceptual signals are normalized to the interval $[0, 1]$ and concatenated into a ten-dimensional perceptual state vector, which serves as input to the cognitive architecture during learning and decision-making.

Finally, in terms of LLM, the experiment employs the Phi-4 14B language model developed by Microsoft. The complete prompt employed during each phase of the experiment is provided in the code repository.

IV. Results and discussion

This section presents the experimental results obtained. Our analysis focuses on the impact of LLM-guided exploration on the behavior of the cognitive architecture and on its learning efficiency.We examine how guidance derived from the LLM shapes the exploration process, accelerates the acquisition of task-relevant knowledge, and influences policy selection.

TABLE I
POLICIES AVAILABLE IN THE REAL ROBOT EXPERIMENT.

Policy	Description
Pick fruit	Use a gripper to grasp a fruit within the robot's reach.
Test fruit	Leave the fruit in the scale to weigh it (it must be grasped from the same side as the scale).
Accept fruit	Place a valid fruit in the accepted basket.
Discard fruit	Place an invalid fruit in the rejected basket.
Change hands	Move the fruit from one gripper to the other (its maximum dimension must be below 0.085).
Place fruit	Leave the fruit in the center of the table.
Ask nicely	Ask experimenter to provide more fruits, if none.
Press button	Press the button.

These results were obtained by executing the real-robot experiment 20 times for each exploration strategy, including the baseline condition, which only addresses the final stage of the task (stage 4). To accelerate data collection while preserving the structure of the interaction dynamics, part of the runs were carried out in a discrete-event simulator calibrated using the real robot. The evaluation focuses on assessing the effectiveness of purpose-driven learning supported by an LLM acting as a cultural knowledge repository. Specifically, we analyze (i) learning rates across approaches, (ii) the usage patterns of exploration policies, (iii) the grounded contextual structures created within the cognitive architecture, and (iv) the reuse of acquired knowledge as evidence of lifelong learning.

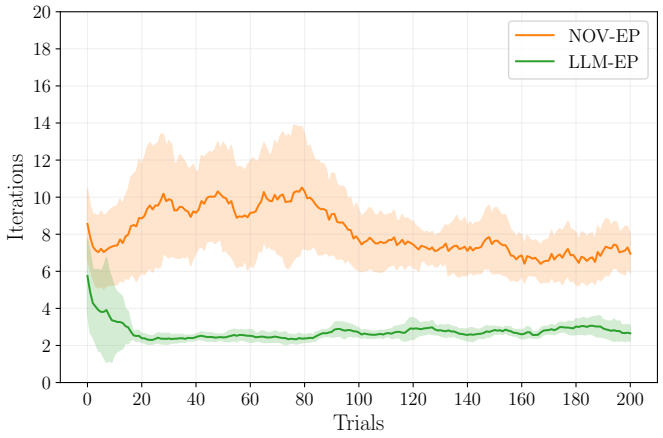

Fig. 4. Iterations needed to achieve the fruit placement purpose for the two exploration approaches during stage 2 of the experiment (95 % confidence interval).

We will start by comparing the results of the performance of the LLM based exploration policy (LLM-EP) and a traditional novelty based exploration policy (NOV-EP) in the second learning stage, in which the purpose of the robot was to place fruits in the center of the table. In this learning stage, not all interaction effects have been learned yet, and therefore, exploratory behavior is of paramount importance. As Fig. 4 shows, the purpose is achieved much faster and more consistently by the LLM-EP than by NOV-EP. Note that NOV-EP can take between 3 and 5 times more iterations to achieve

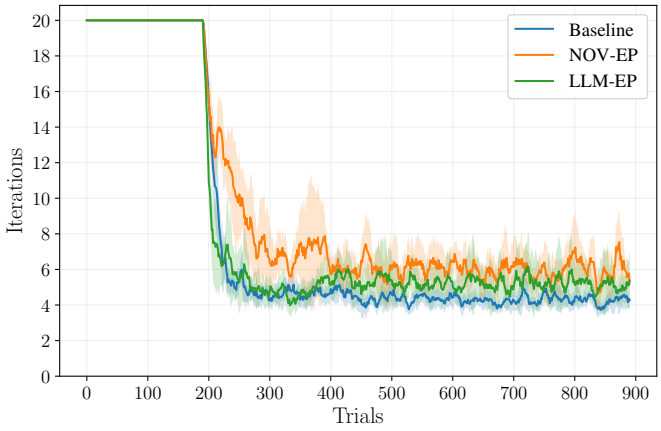

Fig. 5. Iterations needed to reach the fruit classification goal for the three approaches (95 % confidence interval).

the purpose than the LLM-EP in every run. It uses between 6 and 10 iterations, with quite a bit of variability between runs. On the other hand, LLM-EP consistently achieves the purpose in around 2 iterations with very little variability between runs after only 20 learning trials. Summarizing, the focused exploration generated using LLM-EP results in better and much more consistent performance in achieving the purpose.

With regards to the final purpose of performing the whole fruit classification task, which should incorporate all that was learned in previous stages, Fig. 5 compares the learning rate of the baseline approach, NOV-EP and LLM-EP. The first 200 trials correspond to the initial three stages of the experiment, which are not considered in the figure. Once the fourth stage begins, the number of iterations required to reach the final goal decreases substantially faster for LLM-EP than for NOV-EP, with a learning rate closely matching that of the baseline optimal condition. The LLM-guided approach stabilizes after approximately 300 trials, whereas the novelty based exploration policy requires roughly 400 trials to reach comparable stability. Furthermore, in LLM-EP, the average number of iterations needed to achieve the purpose remains consistently closer to that of the baseline.

It is worth noting that in this fourth stage all approaches exhibit considerable variance between runs during the stabilized phase. This variability arises from the stochastic nature of the environment, particularly the random placement of fruits in different runs. In some episodes, the fruit may be initially located close to the scale or already within the gripper, enabling rapid task completion. In other cases, the fruit may be placed further away or require external intervention, resulting in longer interaction sequences.

The effect of LLM guidance on exploration behavior is further illustrated in Fig. 6, which reports the average proportion of time spent in exploration every 100 iterations. Exploration periods are consistently shorter when using the LLM-guided cognitive policy, reflecting a more effective exploitation of previously acquired knowledge. During the initial stage, both the LLM-guided and the novelty based exploration policies

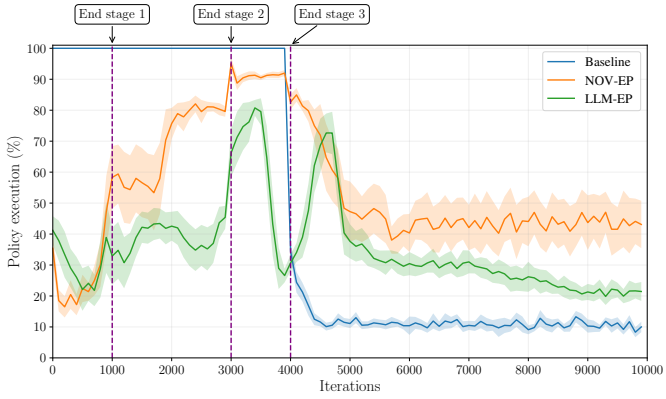

Fig. 6. Average percentage of use of the exploration policy every 100 iterations for each approach (95 % confidence interval).

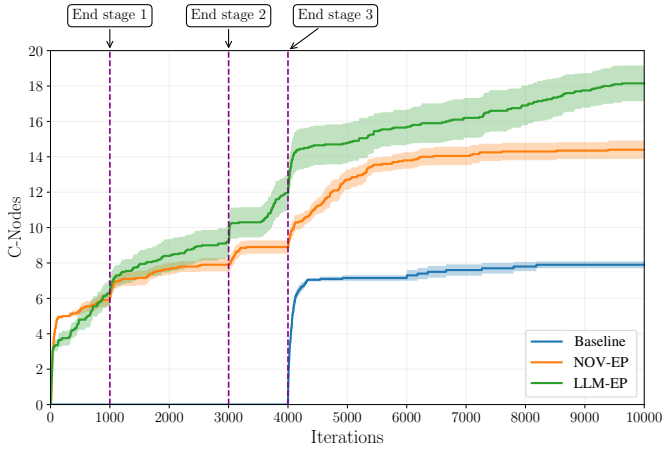

Fig. 7. Average number of C-Nodes generated during the experiment by each approach (95 % confidence interval).

are used with similar frequency, as no prior knowledge is available. In subsequent stages, however, the LLM-guided approach rapidly generates task-relevant knowledge, reducing the need for prolonged exploratory behavior until the onset of the next learning stage.

Fig. 7 provides additional insight into the grounding achieved by the different approaches by showing the average number of contextual nodes (C-Nodes which encode pseudo-rules) generated throughout the experiment. In the baseline condition, which only receives rewards in the final stage, the classification task is solved using approximately eight C-Nodes. In contrast, autonomous learning through exploration leads to the creation of a larger number of C-Nodes, as the robot discovers multiple ways to achieve the same outcomes. While both exploration strategies acquire the core knowledge required for task completion, the LLM-guided approach consistently generates a richer set of contextual structures, reaching approximately 18 C-Nodes compared to 14 for novelty based exploration. These additional structures provide alternative pathways that can be exploited in future tasks and novel situations, supporting adaptability over the robot's lifetime.

TABLE II
INTERPRETATION OF THE DISCOVERED GOALS.

Goal	Description
G_0	Fruit placed on the table center
G_1	Fruit classified properly
G_2	Button light turned on
G_3	Fruit grasped with left gripper
G_4	Fruit grasped with right gripper
G_5	Weighing scale activated
G_6	Fruit grasped with right gripper
G_7	Fruit grasped with left gripper
G_8	Fruit grasped on any gripper
G_9	Fruit grasped on the same side as the scale
G_{10}	Good fruit tested
G_{11}	Bad fruit tested
G_{12}	Button light turned off

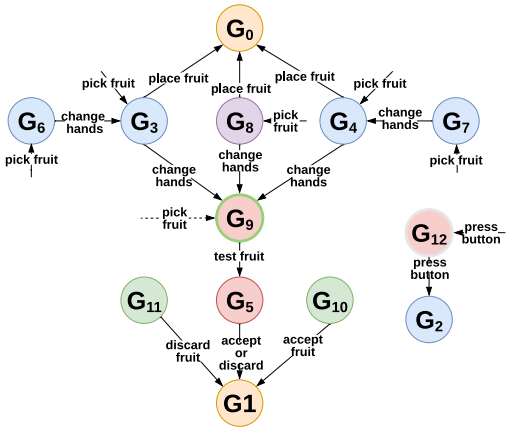

Fig. 8. Goal-Subgoal graph. Goals learned in the first stage are in blue, in the second stage in purple, in the third stage in red, and in the fourth stage in green. G_9 has been learned in the third stage with LLM exploration and in the fourth in the case of novelty based exploration. G_{12} has been learned only with LLM exploration.

Finally, the internal structure of the learned knowledge is illustrated in Fig. 8, which depicts the goal-subgoal graph. Nodes correspond to goals learned at different stages of the experiment, and edges represent the policies required to transition between them. This graph offers a compact and interpretable representation of the robot's decision-making process directly extracted from the C-node structure, highlighting the capacity for explainability of the underlying cognitive architecture. The relationships between goals emerge from experience gathered during LLM-guided trials, rather than being manually specified. Moreover, the graph clearly demonstrates knowledge reuse across stages, as goals learned earlier are leveraged to support more complex behaviors in later phases. For illustration purposes, an a posteriori semantic description of each goal is provided in Table II, and additional details on the goal-creation mechanism of the cognitive architecture can be found in [27].

V. Conclusions and Future Work

This paper presented an approach for integrating Large Language Models (LLMs) as a cultural and social knowledge repository within a structured cognitive architecture, aimed at supporting open-ended and lifelong learning in autonomous robots. Rather than assigning LLMs a direct control role, the proposed method uses them to guide exploration when the robot lacks prior contextual knowledge, while preserving grounding through embodied sensorimotor interaction.

To this end, we extended the e-MDB cognitive architecture with an LLM-based cognitive policy that injects culturally grounded priors to bias exploration and reduce search complexity. LLM-generated hypotheses about how to achieve purposes are not assumed to be correct, but are validated through interaction with the environment. Successful behaviors are consolidated as contextual associative representations (pseudo-rules) stored in long-term memory. This representation supports explainability and enables the accumulation and reuse of grounded knowledge over the robot's lifetime.

The experimental results show that this framework significantly accelerates early learning in new, unknown domains, while preserving the ability to reuse previously acquired knowledge when new purposes arise. From an open-ended learning perspective, the approach addresses the bootstrapping problem of autonomous developmental systems by reducing the cost of exploration respecting grounding or interpretability.

Future work will focus on mitigating current limitations, including computational overhead and the risk of incorrect LLM priors, through improved prompt design, model selection, and verification mechanisms. We also plan to investigate adaptive strategies for regulating LLM guidance over time in more complex, long-horizon open-ended learning scenarios.

References

[1] T. B. Brown *et al.*, "Language models are few-shot learners," *Advances in Neural Information Processing Systems*, 2020.

[2] S. Tellex, N. Gopalan, H. Kress-Gazit, and C. Matuszek, "Robots that use language," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, no. 1, pp. 25–55, 2020.

[3] X. Zhang, Y. Ding, Y. Hayamizu, Z. Altaweel, Y. Zhu, Y. Zhu, P. Stone, C. Paxton, and S. Zhang, "Llm-grop: Visually grounded robot task and motion planning with large language models," *The International Journal of Robotics Research*, p. 02783649251378196, 2025.

[4] W. Huang, P. Abbeel, D. Pathak, and I. Mordatch, "Language models as zero-shot planners: Extracting actionable knowledge for embodied agents," in *ICML*. PMLR, 2022, pp. 9118–9147.

[5] D. Driess, F. Xia, M. S. Sajjadi, C. Lynch, A. Chowdhery, A. Wahid, J. Tompson, Q. Vuong, T. Yu, W. Huang *et al.*, "Palm-e: An embodied multimodal language model," 2023.

[6] S. Harnad, "The symbol grounding problem," *Physica D: Nonlinear Phenomena*, vol. 42, no. 1-3, pp. 335–346, 1990.

[7] W. Samek, "Explainable deep learning: concepts, methods, and new developments," in *Expl. Deep Learning AI*. Elsevier, 2023, pp. 7–33.

[8] T. Lesort *et al.*, "Continual learning for robotics," *Autonomous Robots*, 2020.

[9] S. Doncieux, D. Filliat, N. Díaz-Rodríguez, T. Hospedales, R. Duro, A. Coninx, D. M. Roijers, B. Girard, N. Perrin, and O. Sigaud, "Open-ended learning: a conceptual framework based on representational redescription," *Frontiers in neurorobotics*, vol. 12, p. 59, 2018.

[10] S. Thrun, "Lifelong learning algorithms," in *Learning to learn*. Springer, 1998, pp. 181–209.

[11] M. Ahn, A. Brohan, N. Brown, Y. Chebotar, O. Cortes, B. David, C. Finn, C. Fu, K. Gopalakrishnan, K. Hausman *et al.*, "Do as i can, not as i say: Grounding language in robotic affordances," *arXiv preprint arXiv:2204.01691*, 2022.

[12] I. Singh, V. Blukis, A. Mousavian, A. Goyal, D. Xu, J. Tremblay, D. Fox, J. Thomason, and A. Garg, "Progprompt: Generating situated robot task plans using large language models," in *2023 IEEE Int. Conf. on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 11 523–11 530.

[13] J. Liang, W. Huang, F. Xia, P. Xu, K. Hausman, B. Ichter, P. Florence, and A. Zeng, "Code as policies: Language model programs for embodied control," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 9493–9500.

[14] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, X. Chen, K. Choromanski, T. Ding, D. Driess, A. Dubey, C. Finn *et al.*, "Rt-2: Vision-language-action models transfer web knowledge to robotic control," *arXiv preprint arXiv:2307.15818*, 2023.

[15] T. Carta, C. Romac, T. Wolf, S. Lamprier, O. Sigaud, and P.-Y. Oudeyer, "Grounding large language models in interactive environments with online reinforcement learning," in *International Conference on Machine Learning*. PMLR, 2023, pp. 3676–3713.

[16] Y. Cao, H. Zhao, Y. Cheng, T. Shu, Y. Chen, G. Liu, G. Liang, J. Zhao, J. Yan, and Y. Li, "Survey on large language model-enhanced reinforcement learning: Concept, taxonomy, and methods," *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[17] I. Kotseruba and J. K. Tsotsos, "40 years of cognitive architectures: core cognitive abilities and practical applications," *Artificial Intelligence Review*, vol. 53, no. 1, pp. 17–94, 2020.

[18] G. Baldassarre, M. Mirolli *et al.*, *Intrinsically motivated learning in natural and artificial systems*. Springer, 2013.

[19] A. Romero, G. Baldassarre, R. Duro, and V. G. Santucci, "Autonomous discovery and learning of interdependent goals in non-stationary scenarios," in *2024 IEEE International Conference on Development and Learning (ICDL)*. IEEE, 2024, pp. 1–8.

[20] A. Romero, G. Baldassarre, R. J. Duro, and V. G. Santucci, "H-grail: A robotic motivational architecture to tackle open-ended learning challenges," *IEEE Transactions on Cognitive and Developmental Systems*, 2025.

[21] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska *et al.*, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the national academy of sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.

[22] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual lifelong learning with neural networks: A review," *Neural networks*, vol. 113, pp. 54–71, 2019.

[23] F. Bellas, R. J. Duro, A. Faiña, and D. Souto, "Multilevel darwinist brain (mdb): Artificial evolution in a cognitive architecture for real robots," *IEEE Transactions on autonomous mental development*, vol. 2, no. 4, pp. 340–354, 2010.

[24] J. A. Becerra, A. Romero, F. Bellas, and R. J. Duro, "Motivational engine and long-term memory coupling within a cognitive architecture for lifelong open-ended learning," *Neurocomputing*, vol. 452, pp. 341–354, 2021.

[25] A. Romero, F. Bellas, J. A. Becerra, and R. J. Duro, "Motivation as a tool for designing lifelong learning robots," *Integrated Computer-Aided Engineering*, vol. 27, no. 4, pp. 353–372, 2020.

[26] R. J. Duro, J. A. Becerra, J. Monroy, and F. Bellas, "Perceptual generalization and context in a network memory inspired long-term memory for artificial cognition," *International Journal of Neural Systems*, vol. 29, no. 06, p. 1850053, 2019.

[27] E. Fallas-Hernández, S. Martínez-Alonso, A. Romero, J. A. Becerra-Permuy, and R. J. Duro, "Autonomous generation of sub-goals for lifelong learning in robots," *arXiv preprint arXiv:2503.18914*, 2025.

[28] R. Salgado, A. Prieto, F. Bellas, L. Calvo-Varela, and R. Duro, "Motivational engine with autonomous sub-goal identification for the multilevel darwinist brain," *Biologically Inspired Cognitive Architectures*, vol. 17, pp. 1–11, 2016.

[29] P.-Y. Oudeyer, F. Kaplan, V. V. Hafner, and A. Whyte, "The playground experiment: Task-independent development of a curious robot," in *Proceedings of the AAAI spring symposium on developmental robotics*. Stanford, California, 2005, pp. 42–47.

[30] V. G. Santucci, G. Baldassarre, and M. Mirolli, "Which is the best intrinsic motivation signal for learning multiple skills?" *Frontiers in neurorobotics*, vol. 7, p. 22, 2013.