Hierarchical Memory-Augmented World Models for Long-Horizon Robotic Task Planning in Dynamic Environments

Prekesh Vankatar
Department of Computer Science, George Mason University, Fairfax, VA, USA.
prakash.venkataraman@gmu.edu

Niklas Mendez
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
helloniklas@unr.edu

Keith Craig
School of Computing, Clemson University, Clemson, SC, USA.
craig1999@clemson.edu

Abstract

Long-horizon robotic task planning in dynamic, partially observed environments remains a formidable challenge due to the compounding of uncertainty, the combinatorial explosion of action sequences, and the need for rapid adaptation to unpredictable changes. Hierarchical memory-augmented world models offer a promising architectural response by structuring learned environment models into multiple temporal and spatial scales while equipping each level with dedicated memory components. This paper presents a system-level analysis of such architectures, emphasizing structural trade-offs between representational fidelity, memory capacity, planning depth, and computational tractability. We examine how hierarchical decomposition can decouple abstract task reasoning from fine-grained control, how external and action-aware memory modules can stabilize long-context predictions, and how these elements collectively support robustness in open-world deployments. The discussion extends beyond algorithmic design to encompass governance, fairness, sustainability, and infrastructure requirements, drawing on cross-domain comparisons from cloud robotics, large-scale machine learning systems, and socio-technical policy frameworks. We argue that the fusion of hierarchical organization with memory augmentation constitutes a foundational design principle for building reliable, scalable, and ethically deployable autonomous robotic agents, and we identify open questions concerning interoperability, auditability, and energy efficiency that must be addressed by the research community.

Keywords

Hierarchical world models; memory-augmented neural networks; long-horizon planning; robotic task planning; dynamic environments; actionable memory.

1. Introduction

Robotic systems operating in the open world must reliably sequence thousands of primitive actions to achieve goals that unfold over minutes or hours, all while contending with shifting object arrangements, novel obstacles, and incomplete sensory information. The classical planning paradigm that relies on hand-crafted symbolic models struggles to capture the rich perceptual and physical dynamics present in real environments, whereas pure model-free reinforcement learning often requires prohibitive amounts of interaction data to acquire long-range reasoning capabilities. In recent years, the concept of a learned world model—a predictive model of environment transitions and rewards—has reshaped the landscape of model-based planning by enabling agents to simulate future trajectories inside a compact latent space. The foundational demonstration by Ha and Schmidhuber [1] showed that a variational autoencoder paired with a recurrent dynamics model could generate synthetic rollouts sufficient for training control policies for visual tasks. This was swiftly followed by the Dreamer family of algorithms, which optimized policies directly through latent imagination, thereby reducing the need for costly real-world data collection and achieving state-of-the-art sample efficiency on continuous control benchmarks [2]. Such advances have invigorated a broader discourse on the path toward autonomous machine intelligence, with influential architectural blueprints proposing modular cognitive components—including world models, cost functions, and memory—that jointly enable reasoning across multiple time scales [3].

Despite these successes, flat world models that represent the entire environment state in a single monolithic latent vector encounter fundamental limitations when confronted with long-horizon tasks. The prediction error of a dynamics model tends to compound exponentially over the planning horizon, a phenomenon that becomes especially acute in stochastic and non-stationary settings. Furthermore, a flat representation lacks the inductive bias necessary to reason about subgoals, reusable skill sequences, or the hierarchical structure inherent in many real-world activities. This observation has motivated a growing body of work that integrates hierarchical structure with world model architectures. Simultaneously, the resurgence of differentiable memory mechanisms—originally developed for tasks such as algorithmic reasoning and few-shot learning—has opened the door to equipping world models with explicit storage and retrieval capabilities that can retain information far longer than the model's recurrent state. By combining hierarchical decomposition with memory augmentation, it becomes possible to design world models that not only plan across extended temporal horizons but also adapt their internal predictions in response to novel events without catastrophic forgetting of previously learned dynamics. This paper develops a system-level perspective on such hierarchical memory-augmented world models, scrutinizing the architectural choices that govern their effectiveness, the robustness properties that determine their real-world viability, and the broader infrastructure and governance dimensions that will shape their deployment.

2. Background and Related Work

The challenge of hierarchical planning has been approached from multiple disciplinary angles, including robotics, cognitive science, and artificial intelligence. In the domain of task-and-motion planning, hierarchical decomposition typically takes the form of a high-level symbolic task planner that operates over abstract predicates and a low-level motion planner that refines each symbolic action into continuous trajectories. Learned hierarchical models aim to acquire such decompositions automatically from data. Composable models of parameterized skills can be used to construct plans at varying levels of abstraction, where each skill is represented as a small Markov decision process that can be sequenced and adapted [4]. These learned skill representations significantly reduce the search space for planning and facilitate transfer across tasks that share a common skill vocabulary. At the same time, the problem of long-term memory in neural architectures has been addressed through the development of differentiable external memory systems. The Differentiable Neural Computer, for instance, equipped a controller network with a read-write memory matrix that could be accessed through attention operations, enabling the learning of complex algorithms from input-output examples [5]. Although originally demonstrated on synthetic reasoning tasks, the architectural principle of decoupling storage capacity from model depth has profound implications for world models, where the ability to store and recall precise facts about an environment over thousands of time steps can make the difference between successful task completion and failure.

The intersection of hierarchical control and skill discovery has further enriched the design space. Unsupervised skill discovery methods that maximize the predictability of state transitions while maintaining diversity have shown that agents can autonomously identify reusable behaviors that span several time steps without external reward signals [6]. These skills naturally form a two-level hierarchy: a high-level policy that selects among skills and a low-level controller that executes the chosen skill. When such skills are embedded within a learned world model, the high-level policy can plan in an abstract state space defined by skill termination conditions, thereby expanding the effective planning horizon. More recently, the introduction of action-aware memory constructs into world models has demonstrated considerable promise for interactive settings. By conditioning memory storage and retrieval on the agent's own actions, an architecture can tighten the coupling between what the agent does and what it remembers, leading to more coherent predictions under active exploration [7]. This line of work exemplifies a shift from purely observational world models toward interactive ones that are intrinsically suited for closed-loop planning. Complementary developments in sequence modeling have also influenced the design of planning systems, as evidenced by the Decision Transformer framework, which reformulates reinforcement learning as a conditional sequence generation problem and shows how large transformer models can produce coherent long-term behaviors when trained on diverse offline datasets [8]. Similarly, the reconceptualization of offline reinforcement learning as a high-dimensional sequence modeling task has opened new avenues for leveraging scalable architectures inspired by natural language processing [9].

A parallel thread of research emphasizes the construction of generalist agents that embed world knowledge, visual perception, and motor control within a single large model. The Gato architecture demonstrated that a single transformer could perform a wide variety of tasks—ranging from playing Atari games to controlling a robotic arm—by processing all modalities as a flat sequence of tokens, yet it did so without explicit hierarchical planning or memory augmentation [10]. In contrast, DreamerV3 introduced a principled approach to building world models that are robust across domains with minimal hyperparameter tuning, using symlog transformations to handle rewards of vastly different magnitudes [11]. At the level of skill priors, methods that learn a distribution over skill embeddings from offline experience have been used to accelerate reinforcement learning by providing a meaningful behavioral initialization for new tasks [12]. In parallel, exploration strategies that seek to match the state marginal distribution of a target task have been proposed to guide agents efficiently through large state spaces without random wandering [13]. The integration of causality into world models offers yet another frontier, where explicitly modeling causal dependencies between state variables can improve sample efficiency and generalization under distribution shift [14]. Finally, the availability of reusable visual representations pre-trained on large-scale video or human interaction data, such as R3M and RT-2, has begun to decouple perception from dynamics modeling, allowing world models to be trained on semantically rich feature spaces rather than raw pixels [15], [16]. Collectively, these advances set the stage for a systematic examination of the structural and governance aspects of hierarchical memory-augmented world models.

3. Hierarchical Memory-Augmented World Models: An Architectural Overview

A hierarchical memory-augmented world model can be understood as a multi-scale predictive system in which each level of abstraction maintains its own latent state, its own dynamics predictor, and its own memory module. The bottom level operates at the finest temporal resolution, often processing raw sensor data or compact latent encodings produced by a visual backbone, and its dynamics model predicts immediate next states given the current observation and action. The top level operates over a compressed representation that abstracts away minor fluctuations and encodes only those features that are relevant for long-range goal achievement. Between these extremes, one or more intermediate levels can exist, each with a progressively coarser temporal granularity and a more simplified representation of the environment. A key design choice is whether the levels communicate strictly in a top-down manner—with higher levels setting goals or constraints for lower levels—or whether bottom-up signals also propagate, allowing lower-level surprises to trigger re-planning at higher levels. Bidirectional information flow tends to produce more adaptive behavior but also introduces additional coupling that can destabilize training unless carefully regularized.

Memory modules at each level serve to store representations of past observations, action sequences, or even abstract subgoals that are expected to be useful for future prediction. These modules can be implemented as differentiable key-value stores, slot-based memory buffers, or as continuous attractor networks that maintain persistent representations over time. The architectural innovation of recent interactive world models is the incorporation of action-awareness into the memory read-write mechanisms. Rather than treating memory as a passive store of past observations, action-aware memory updates its contents based on the agent's current intentions and control signals [7]. This design tightens the synchronization between planning and prediction: when the agent is about to execute a complex manipulation sequence, for instance, the memory module can be primed to retain precise spatial configurations that would otherwise be overwritten if only recency-based storage heuristics were used. The memory hierarchy can thus be understood not merely as a means of extending the temporal horizon but as a mechanism for managing the representational bandwidth across planning levels.

The training of such architectures raises important questions about end-to-end versus staged optimization. End-to-end learning, where all levels and memory controllers are updated jointly from task rewards and prediction losses, offers the allure of simplicity and the potential for synergistic co-adaptation. However, it is notoriously difficult to propagate gradients through a temporal hierarchy that spans vastly different time scales, often leading to destabilizing gradient conflicts. Staged training, in which lower levels are first pre-trained to reconstruct observations and predict immediate transitions before higher levels are introduced, provides a more controlled curriculum but can result in suboptimal specialization that is hard to revise later. Hybrid strategies that interleave fine-tuning of individual levels with joint optimization have been explored, but no consensus has emerged regarding the optimal training schedule. The structural trade-offs that arise in the design of memory hierarchies within such world models are sufficiently nuanced to warrant dedicated analysis, which is provided in the following section.

4. Structural Trade-offs in Memory Hierarchy Design

The allocation of memory capacity across the hierarchical levels is perhaps the most consequential design decision in a memory-augmented world model. Concentrating most memory resources at the top level enables the system to maintain a detailed log of high-level decisions, subgoal sequences, and abstract environmental context over very long durations, but it risks starving the lower levels of the immediate contextual detail needed for precise execution. Conversely, endowing lower levels with large memory stores can improve short-horizon prediction accuracy and support reactive behaviors, but memory at fine temporal scales tends to become stale quickly and may be overwritten before it can inform higher-level reasoning. A balanced design typically involves a carefully tuned memory capacity allocation that mirrors the information content at each scale, often with an exponential taper as one moves from top to bottom. Implementation choices such as memory size, compression ratio of stored representations, and the presence of forget gates introduce additional degrees of freedom that interact with task characteristics, making it challenging to prescribe a one-size-fits-all configuration.

The addressing mechanism used to access memory is another axis along which important trade-offs emerge. Content-based attention, where memory slots are retrieved based on their similarity to a query vector, provides flexibility and can recall information that was stored far in the past without relying on exact positional indexing. However, such soft attention becomes computationally expensive as the memory bank grows and can suffer from interference when many encoded experiences share similar content signatures. Slot-based addressing, which assigns fixed memory slots to specific object-centric or spatial regions, avoids interference but requires a strong inductive bias about how the environment should be decomposed. Hybrid addressing schemes that combine location-based and content-based cues have been shown to improve efficiency and robustness, yet they increase the number of hyperparameters and the complexity of the memory controller. In the context of action-aware memory, the addressing mechanism must also be cognizant of the agent's behavioral state, introducing a triadic interaction between memory location, perceptual content, and motor intention that remains only partially understood.

The granularity of time abstraction in the hierarchy directly impacts both planning accuracy and computational cost. Levels that operate at coarse time resolutions can plan over extended horizons with few steps, but their predictions may be too coarse to resolve critical bottlenecks such as narrow passageways in navigation or precise alignment phases in assembly tasks. Levels that operate at fine time resolutions provide detailed predictions but make planning computationally demanding, as the search tree expands combinatorially with depth. Engineers must therefore strike a balance that aligns the temporal abstraction with the characteristic timescales of the environment dynamics and the task structure. One approach is to make the temporal abstraction adaptive, allowing higher levels to dynamically adjust their step size based on the predictive uncertainty they encounter. This adaptability introduces a meta-level learning problem that itself requires careful management of computational overhead and training stability.

5. Robustness and Adaptability in Dynamic Environments

Dynamic environments are characterized by continuous change that can be gradual, such as lighting variations over a day, or abrupt, such as a door being closed or an object being moved by a human collaborator. A hierarchical memory-augmented world model offers several structural advantages for maintaining robustness under such conditions. The hierarchical decomposition naturally isolates the effects of many local changes, because a disturbance at the finest temporal scale may leave the higher-level abstraction largely intact, allowing the agent to replan only at the affected lower levels while preserving the overarching high-level strategy. This insulation property reduces the reactive delay that would be incurred if the entire world model had to be recomputed from scratch. Memory modules further enhance robustness by providing a means to store and retrieve situational patterns that have been encountered before, effectively allowing the agent to retrieve prior solutions rather than synthesize them de novo on each occurrence. The action-aware memory paradigm is particularly relevant here, as it enables the agent to shield goal-relevant memories from being overwritten by the stream of ongoing sensory data, thereby maintaining strategic continuity even when the perceptual field changes dramatically.

The capacity for continual adaptation without catastrophic forgetting is a persistent challenge for any neural system deployed in an evolving world. In a hierarchical memory-augmented architecture, forgetting can be managed at multiple levels: the lower-level memory may be allowed to be relatively volatile, accommodating new perceptual experiences, while the upper-level memory is stabilized through slower update rules or explicit consolidation mechanisms. This biological analogy, reminiscent of the complementary learning systems theory in neuroscience, suggests a principled way to balance plasticity and stability across the hierarchy. When the agent encounters novel dynamics that persistently violate its predictions, it can trigger a controlled restructuring of the affected level's dynamics model and memory representation, a process that must be carefully regulated to avoid unlearning related skills. The integration of uncertainty estimation at each hierarchy level provides a natural signal for initiating such structural updates, as a sustained increase in prediction error at a given abstraction layer signals that the current model is no longer sufficient.

6. Governance, Fairness, and Policy Implications

The deployment of autonomous robots equipped with sophisticated world models in human-shared spaces raises a host of governance concerns that extend well beyond technical performance. A hierarchical planning system makes decisions at varying levels of abstraction, and the opacity of these decisions poses challenges for transparency and accountability. When a robot fails to complete a long-horizon task, determining whether the root cause lies in an incorrect high-level subgoal selection, a faulty memory retrieval, or a low-level control error becomes a non-trivial diagnostic problem. The integration of action-aware memory compounds this opacity, because the robot's internal memory state is no longer a passive record of observations but an actively constructed representation influenced by its own behavioral biases. This raises the bar for designing explainability interfaces that can expose the chain of reasoning to human operators and regulators.

Fairness considerations enter when such robotic systems are deployed in service roles that interact with diverse populations. If the memory-augmented world model is trained predominantly on data collected from specific demographic groups or architectural environments, the resulting planning behaviors may systematically disadvantage underrepresented populations. For instance, a domestic robot that has learned navigation strategies by observing motion patterns in spacious suburban homes may fail to construct appropriate high-level plans when deployed in a small urban apartment, leading to degraded service quality. The memory mechanism itself can amplify these disparities if it over-indexes on frequently encountered patterns and fails to retain rare but critical situations. Mitigating such biases requires deliberate data collection strategies, algorithmic fairness constraints integrated into the planning objectives, and continuous auditing of deployment outcomes. The broader field of responsible artificial intelligence has begun to develop frameworks for such auditing, but their application to embodied agents with hierarchical memory remains nascent [17], [18].

From a policy perspective, the long-horizon autonomy enabled by these world models invites regulatory attention to the scope of permissible autonomous decision-making. As robots become capable of generating and executing plans over many minutes or hours without human intervention, the question of at what point a human must be brought into the loop becomes pressing. Governance models from aviation automation, which mandate specific human oversight levels based on task criticality, may provide a useful template, but they require adaptation to the open-ended nature of robotic task planning. Moreover, the ability of memory-augmented systems to accumulate and leverage vast stores of experience over time raises unresolved questions about data ownership, consent, and the right to be forgotten, particularly when robots operate in private spaces such as homes and healthcare facilities.

7. Deployment Infrastructure and Sustainability

Bringing hierarchical memory-augmented world models out of simulation and onto physical robotic platforms requires substantial supporting infrastructure. These models are computationally intensive during both training and inference, demanding high-throughput accelerators and low-latency memory subsystems for real-time planning. Edge deployment remains challenging because the power and thermal budgets of mobile robots are limited, while offloading computation to cloud resources introduces latency and reliability risks that are unacceptable for safety-critical control loops. A practical deployment strategy often involves partitioning the hierarchy across the edge-cloud continuum: the fine-grained reactive level runs on local embedded hardware with lightweight memory structures optimized for fast retrieval, while the higher strategic levels run on more powerful remote servers that can host large memory banks and perform computationally expensive planning. This partitioned architecture raises its own engineering challenges, including secure, low-latency communication, graceful degradation when connectivity is lost, and version consistency between the models running at different sites. Drawing on lessons from cloud robotics, which has studied the orchestration of distributed robotic computation for over a decade, can inform the design of robust communication protocols and fault-tolerance strategies [19].

The sustainability implications of large-scale world model deployments are also coming under scrutiny. Training a single large world model with hierarchical memory and action-aware components can consume significant energy, equivalent to several tons of carbon dioxide emissions if conducted on non-renewable power grids. While a single robot may not incur such costs continuously, the cumulative footprint of a fleet of robots that periodically update their models through federated learning or centralized retraining can be considerable. The research community has begun to explore strategies for reducing this impact through model compression, efficient memory architectures, and the use of pre-trained spatio-temporal backbones that can be fine-tuned with lower resource expenditure [20]. In addition, the lifecycle of hardware accelerators that support these models must be considered, as the rapid obsolescence of specialized chips can lead to electronic waste and resource depletion. A holistic sustainability assessment for hierarchical memory-augmented world models must therefore account for the entire pipeline from data collection and training through to deployment, operation, and eventual decommissioning.

8. Conclusion

Hierarchical memory-augmented world models represent a compelling architectural synthesis that addresses the twin demands of long-horizon planning and robust adaptation in dynamic environments. By integrating multi-scale temporal abstraction with action-aware memory mechanisms, these systems can retain and strategically recall experiences over extended durations while maintaining the responsiveness required for fine-grained control. This paper has provided a systems-focused analysis of the structural trade-offs that arise in designing memory hierarchies, the training strategies that balance stability and plasticity, and the robustness properties that enable deployment in non-stationary worlds. Beyond technical architecture, we have argued that governance, fairness, and sustainability must be treated as first-class design considerations rather than afterthoughts. The opacity of hierarchical decision-making, the risk of encoded biases, and the energy footprint of large-scale models all demand careful attention from researchers, engineers, and policymakers alike. As the field moves toward increasingly autonomous robotic systems that operate in close proximity to humans, the principles articulated here—grounded in cross-domain lessons from cloud infrastructure, cognitive science, and regulatory theory—will be essential for ensuring that these technologies are not only powerful but also trustworthy and sustainable.

References

1. Ha, D., & Schmidhuber, J. (2018). World models. arXiv preprint arXiv:1803.10122.

2. Hafner, D., Lillicrap, T., Ba, J., & Norouzi, M. (2020). Dream to control: Learning behaviors by latent imagination. In International Conference on Learning Representations (ICLR).

3. LeCun, Y. (2022). A path towards autonomous machine intelligence. arXiv preprint arXiv:2207.10612.

4. Kaelbling, L. P., & Lozano-Pérez, T. (2017). Learning composable models of parameterized skills. In 2017 IEEE International Conference on Robotics and Automation (ICRA) (pp. 4015–4022). IEEE.

5. Graves, A., Wayne, G., Reynolds, M., Harley, T., Danihelka, I., Grabska-Barwińska, A., ... & Hassabis, D. (2016). Hybrid computing using a neural network with dynamic external memory. Nature, 538(7626), 471–476.

6. Sharma, A., Gu, S., Levine, S., Kumar, V., & Hausman, K. (2020). Dynamics-aware unsupervised discovery of skills. In International Conference on Learning Representations (ICLR).

7. Xiong, Z., Song, Y., Kang, H., Yan, Q., Jiang, L., Yang, J., ... & Jacobs, N. (2026). ActWorld: From Explorable to Interactive World Model via Action-Aware Memory. arXiv preprint arXiv:2606.17730.

8. Chen, L., Lu, K., Rajeswaran, A., Lee, K., Grover, A., Laskin, M., ... & Mordatch, I. (2021). Decision transformer: Reinforcement learning via sequence modeling. In Advances in Neural Information Processing Systems (NeurIPS).

9. Janner, M., Li, Q., & Levine, S. (2021). Offline reinforcement learning as one big sequence modeling problem. In Advances in Neural Information Processing Systems (NeurIPS).

10. Reed, S., Zolna, K., Parisotto, E., Colmenarejo, S. G., Novikov, A., Barth-Maron, G., ... & de Freitas, N. (2022). A generalist agent. arXiv preprint arXiv:2205.06175.

11. Hafner, D., Pasukonis, J., Ba, J., & Lillicrap, T. (2023). Mastering diverse domains through world models. arXiv preprint arXiv:2301.04104.

12. Pertsch, K., Lee, Y., & Lim, J. J. (2020). Accelerating reinforcement learning with learned skill priors. In Conference on Robot Learning (CoRL).

13. Lee, L., Eysenbach, B., Parisotto, E., Xing, E., Levine, S., & Salakhutdinov, R. (2019). Efficient exploration via state marginal matching. In Advances in Neural Information Processing Systems (NeurIPS).

14. Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2021). Dive into causal world models: A causal approach to robust reinforcement learning. In International Conference on Learning Representations (ICLR).

15. Nair, S., Rajeswaran, A., Kumar, V., Finn, C., & Gupta, A. (2022). R3M: A universal visual representation for robot manipulation. In Conference on Robot Learning (CoRL).

16. Brohan, A., Brown, N., Carbajal, J., Chebotar, Y., Chen, X., Choromanski, K., ... & Zitkovich, B. (2023). RT-2: Vision-language-action models transfer web knowledge to robotic control. arXiv preprint arXiv:2307.15818.

17. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. arXiv preprint arXiv:1606.06565.

18. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1–35.

19. Kehoe, B., Patil, S., Abbeel, P., & Goldberg, K. (2015). A survey of research on cloud robotics and automation. IEEE Transactions on Automation Science and Engineering, 12(2), 398–409.

20. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL).