

Action-Aware World Models for Long-Horizon Robotic Manipulation and Task Planning

Neil A. Koskinen

School of Computing, Clemson University, Clemson, SC, USA.
neilakoskinen@clemson.edu

Wenhao Long

Department of Computer Science, University of Central Florida, Orlando, FL, USA.
wenhaowork@ucf.edu

Hudson D. Reynolds

School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.
reynolds77@uc.edu

Abstract

Long-horizon robotic manipulation requires an agent to maintain coherent representations of objects, tasks, and environmental dynamics while executing extended sequences of physically grounded actions. Model-based reinforcement learning has emerged as a promising route toward sample-efficient planning, yet many existing world models remain largely passive predictors rather than action-aware reasoning systems that can support interactive task planning under uncertainty. This paper examines the architectural, infrastructural, and governance implications of action-aware world models for long-horizon robotic manipulation. It argues that effective systems must integrate predictive dynamics, memory, task specification, and action-conditioned attention within a unified architecture that balances model capacity, inference latency, data efficiency, and safety constraints. The discussion covers structural trade-offs in representational learning, the role of attention and memory in long-range credit assignment, the deployment challenges of embodied systems in partially observable environments, and the need for robust evaluation practices. The paper further considers fairness, accountability, and policy dimensions related to autonomous manipulation systems operating in human-shared spaces. By connecting learning-based planning, system engineering, and socio-technical governance, the analysis provides a multidimensional view of how action-aware world models can evolve from exploratory prediction tools into reliable components of next-generation robotic task planning infrastructures.

Keywords

action-aware world models; robotic manipulation; long-horizon planning; model-based reinforcement learning; embodied artificial intelligence; system governance; autonomous systems.

1. Introduction

Long-horizon robotic manipulation remains one of the most demanding challenges in embodied artificial intelligence because successful performance depends on coordinated perception, memory, physical interaction, and sequence-level reasoning across extended time scales. Unlike board games or simulated control benchmarks, manipulation tasks involve partial observability, deformable and granular materials, contact-rich dynamics, frequent

exogenous disturbances, and environment-specific constraints that are difficult to encode in advance. Model-based reinforcement learning has historically treated a predictive model of the environment as a central component for sample-efficient control and planning [1]. Early latent world model formulations demonstrated that high-dimensional observations could be compressed into lower-dimensional state representations to enable forward prediction and policy optimization without requiring a handcrafted simulator [2]. However, predictive accuracy alone does not guarantee usefulness for long-horizon task execution. Models must also retain task-relevant information, distinguish controllable actions from uncontrollable disturbances, and adapt when predictions diverge from observed reality.

Subsequent research has steadily moved from passive observation models toward learned dynamics models that support planning in compact latent spaces [3]. Discrete world model architectures have shown that categorical latent variables can reduce compounding error while supporting long-horizon credit assignment in visually rich environments [4]. Planning-oriented systems such as MuZero extended learned models to domains where an exact simulator is unavailable, combining reward, value, and policy prediction within a unified search process [5]. This line of work is conceptually rooted in earlier demonstrations that deep networks combined with tree search can achieve strong performance in complex sequential decision tasks [6]. Robotic manipulation, however, adds physical constraints and real-time execution demands that differ from simulated games. End-to-end visuomotor policies have produced compelling results for shorter manipulation behaviors by mapping raw perception directly to joint actions [7], but they often struggle when a task requires subtask ordering, recovery from failure, or reuse of past experience. Neural network dynamics models have therefore been explored as a way to enable model-based deep reinforcement learning with subsequent model-free fine-tuning [8], while probabilistic dynamics models have improved data efficiency by representing uncertainty explicitly [9]. Meta-learning approaches further support rapid adaptation across related tasks by learning representations and update rules that transfer to new settings [10].

More recently, the architectural landscape has been reshaped by attention mechanisms [11] and large pretrained language models [12]. Language-conditioned imitation learning has linked natural language instructions to low-level motor actions [13], while generalist agent formulations have attempted to unify image, language, and action modalities in a single policy [14]. Embodied multimodal models integrate visual perception, language understanding, and robotic control in ways that point toward more flexible instruction following [15]. These developments resonate with broader proposals for autonomous machine intelligence built on predictive world models, hierarchical planning, and self-supervised learning [16]. At the same time, foundation models raise systemic questions about robustness, evaluation, governance, and unintended social consequences [17]. Within this context, action-aware world models have been proposed as interactive memory systems that move beyond explorable representations toward action-conditioned update mechanisms [18]. Frontier model documentation further illustrates the practical complexity of aligning large-scale learning systems with reliable and context-sensitive behavior [19]. This paper develops a system-level analysis of these intersecting trends, focusing on the architectural trade-offs, infrastructure requirements, and governance considerations that will shape next-generation action-aware world models for long-horizon manipulation.

2. Background and Related Work

The conceptual foundation of model-based reinforcement learning can be traced to the idea that an agent can learn to plan by interleaving real experience with simulated experience generated by a learned model. This integration of learning, planning, and reacting is central to Dyna-style architectures and remains influential in contemporary robotic learning systems. World models extended this principle by learning compressed representations from high-dimensional sensory streams, enabling agents to imagine future trajectories without physically executing every exploratory action. The shift from raw pixel prediction to latent dynamics modeling was important because pixel-level reconstruction can waste representational capacity on visually salient but behaviorally irrelevant detail. Latent dynamics models instead optimize representations for the downstream planning objective, reducing the burden on the model and improving sample efficiency.

A recurring difficulty in latent prediction is that compounding error over long horizons can make imagined rollouts diverge sharply from actual outcomes. Discrete world models address part of this problem by restricting the latent space to a set of categorical states, which encourages the model to retain salient transitions rather than fine-grained but unstable visual texture. Planning with learned models has also benefited from value prediction, which allows the agent to evaluate imagined futures without simulating every step at full resolution. Search-based integration of learned models and value estimates has been highly effective in domains such as board games and video games, but robotic manipulation introduces continuous action spaces, contact dynamics, and safety-critical constraints that require careful adaptation. For manipulation, model-free visuomotor policies have historically been attractive because they avoid the need for accurate forward prediction and can learn reactive strategies directly from demonstrations or reward signals. However, they can be inefficient for long tasks that require multiple subtasks and explicit memory of past states.

The integration of language and vision has changed the design space for manipulation systems. Large pretrained language models provide strong priors over task semantics, object affordances, and plausible action sequences. Language-conditioned imitation learning uses such priors to map instructions to actions, while multimodal models jointly process images, text, and robot state. These models support generalization across tasks but also create new failure modes related to grounding, ambiguity, and the mismatch between linguistic abstraction and physical execution. In parallel, the rise of foundation models has prompted deeper examination of evaluation practices, robustness, and societal impact. The action-aware world model concept addresses an important gap in this literature by proposing that memory and representation updates should be explicitly conditioned on the actions an agent takes, rather than treating the world model as a passive predictor of sensory dynamics. This perspective connects predictive modeling, interactive memory, and task planning in a way that is directly relevant to long-horizon manipulation.

3. Action-Aware World Model Architecture

An action-aware world model for manipulation must serve multiple purposes simultaneously. It should predict how the environment changes in response to motor commands, maintain memory of task-relevant events over time, provide a representation from which a planner can evaluate candidate actions, and support uncertainty estimation when the environment deviates from previously encountered dynamics. These requirements create architectural tensions. A model that preserves fine-grained visual detail may be useful for detecting small object displacements but expensive to train and slow to query during online planning. A model that compresses aggressively may be fast and stable but may discard the very information needed

to distinguish similar objects or recover from contact-rich failures. The design challenge is to align representational capacity with task demands rather than optimizing only for reconstruction or prediction accuracy.

Attention mechanisms provide one pathway for managing this tension. Instead of treating all parts of a visual scene or all steps in a trajectory as equally important, attention allows the model to focus on action-relevant regions and time points. In manipulation, this might mean attending to the contact point between a gripper and an object during a grasp, or attending to the remembered location of a tool that was set aside several steps earlier. Attention also enables cross-modal integration, allowing language instructions, visual observations, and robot proprioception to inform one another within a shared representational space. The action-aware memory component can be implemented as a learned memory store that is updated by an action-conditioned transition function. When the agent takes an action, the memory update should reflect not only the new observation but also the intended effect of that action and the discrepancy between expected and actual outcomes. This helps the model maintain a coherent belief state even in partially observable settings where raw observations are ambiguous.

The architecture must also include a planning interface that can query the world model efficiently. In long-horizon manipulation, exhaustive search over possible action sequences is typically infeasible. Instead, planners rely on sampling, beam search, or hierarchical decomposition. The action-aware representation should therefore expose not only predicted next states but also affordance-like summaries that indicate which actions are physically plausible and which subtasks have been completed. Such summaries can be used by a high-level task planner to sequence subtasks while low-level controllers handle continuous motion. This hierarchical division reduces the burden on the world model and allows the system to operate with bounded inference latency. It also creates clear points for human inspection and intervention, which is important for safety and accountability.

4. Long-Horizon Planning and Memory Integration

Long-horizon manipulation tasks often require an agent to maintain a persistent understanding of objects that are temporarily out of view. For example, an agent may need to place an ingredient in a bowl, later retrieve a utensil from a drawer, and remember that the ingredient has already been added. Traditional reactive policies may fail in such settings because the relevant cue is no longer present in the current observation. World models that include an explicit memory mechanism can store task-relevant facts and retrieve them when needed. The action-aware property is particularly important because memory should be updated in light of what the agent did, not merely what the agent saw. If the agent believes it has poured a liquid but the visual observation is ambiguous, the action history can help disambiguate the state. Conversely, if an action fails, the discrepancy between predicted and actual outcomes should be recorded so that the model can revise its understanding of the environment.

Planning with such a model involves repeatedly imagining future trajectories, evaluating their expected outcomes, and selecting actions that balance task progress against risk. The model must be able to distinguish between uncertainty arising from incomplete observation and uncertainty arising from genuine stochasticity in the environment. In manipulation, contact dynamics are often stochastic at the low level but highly structured at the task level. A well-designed action-aware model can capture this structure by learning transition dynamics that are locally noisy but globally constrained by task semantics. This is related to commonsense reasoning about physical events, where large language models have been used to assess whether event sequences are plausible [20]. Although robotic manipulation requires more

precise physical grounding than language-based commonsense evaluation, similar principles apply: the model should develop expectations about coherent action sequences and flag transitions that violate those expectations.

Memory integration also affects the credit assignment problem. When a task fails at the final step, the agent must determine which earlier decision caused the failure. This is difficult in long sequences because the causal connection may span many time steps. An action-aware memory that records action-effect pairs can support retrospective analysis by enabling the agent to trace the chain of predictions and discrepancies. The planner can then adjust its behavior not by retraining from scratch but by revising the memory entries associated with the faulty transition. This capability is essential for continual learning in real-world settings where the robot must improve over time without discarding previously acquired competence.

5. System-Level Trade-offs and Infrastructure

Deploying action-aware world models in real robotic systems requires more than algorithmic innovation. It demands careful engineering of data pipelines, inference infrastructure, and model lifecycle management. Robotic manipulation datasets are expensive to collect because they require physical interaction with the world. Simulated data can reduce this cost, but the gap between simulation and reality remains a significant source of error. A major infrastructure challenge is therefore to maintain useful world model performance when the deployment environment differs from the training distribution. This challenge is not limited to visual appearance; it also includes contact parameters, object weights, surface friction, and the kinematic characteristics of different robot platforms. Systems must be designed with domain adaptation and continual fine-tuning in mind, rather than assuming that a model trained once will remain reliable.

Compute constraints are also central. Unlike offline language model training, robotic manipulation requires online inference at control frequencies that may range from tens to hundreds of hertz. A world model used inside a planning loop must be sufficiently fast to evaluate candidate actions without causing unacceptable delays. This creates pressure to reduce model size, quantize parameters, and optimize attention calculations. However, aggressive compression can degrade long-horizon coherence. A more balanced infrastructure strategy is to use a large, high-capacity model for slower task planning and a smaller, faster model for reactive control. The action-aware memory can serve as the interface between these two levels, providing the high-level planner with compact state summaries while allowing the low-level controller to operate on immediate sensory inputs. This separation also supports modular testing, where individual components can be evaluated for latency, accuracy, and failure modes.

Data governance is another systemic concern. Robotic manipulation systems operating in real environments collect vast amounts of visual and spatial information that may include human activity, private spaces, and identifiable objects. The infrastructure must include mechanisms for consent, data minimization, secure storage, and auditability. Moreover, because world models are often trained on datasets aggregated from multiple sources, issues of data quality, provenance, and bias become directly relevant to downstream behavior. A model trained predominantly on kitchen environments may fail in industrial settings, while a system that relies on demonstrations from a narrow demographic of operators may encode assumptions about how tasks should be performed. These are not merely technical limitations; they are structural properties of the data pipeline and must be addressed through deliberate system design and organizational policy.

6. Robustness, Fairness, and Safety Governance

The deployment of long-horizon robotic manipulation in human-shared environments raises safety concerns that differ in important ways from those in laboratory research. A robot that has learned to manipulate objects must be able to recognize when its world model is unreliable and pause or request help rather than executing a dangerous action. Robustness engineering for these systems must therefore include uncertainty calibration, anomaly detection, and safe fallback behaviors. An action-aware model can support these functions by comparing predicted outcomes with observed outcomes and flagging large discrepancies. If the model repeatedly fails to predict the effects of actions in a particular region of state space, the system should reduce its autonomy and rely on more conservative control strategies. This is consistent with broader safety research emphasizing the need for agents to recognize their own limitations and act accordingly [21].

Fairness considerations arise when manipulation systems are deployed in public or shared environments. A robot that performs a manipulation task differently depending on the visual appearance of a user or the layout of a particular home may reproduce structural inequalities. Model cards and similar documentation artifacts have been proposed to increase transparency about intended use, performance characteristics, and known limitations [22]. For embodied systems, such documentation must extend beyond aggregate accuracy metrics to include physical interaction risks, environmental assumptions, and the types of users for whom the system has been validated. The action-aware memory itself can carry biases because it learns from historically collected interaction data. If certain action sequences are underrepresented in the training distribution, the model may fail to plan effective behavior for those sequences during deployment. Addressing this requires not only algorithmic adjustment but also participatory design processes that include diverse stakeholders in the specification of tasks and evaluation criteria.

Governance frameworks must also account for the possibility of misuse. Robotic manipulation systems are dual-use technologies that can support beneficial applications such as assisted living, industrial automation, and disaster response, but they can also be repurposed for harmful physical intervention. The broader literature on malicious uses of artificial intelligence emphasizes the importance of forecasting potential harms, sharing threat information, and developing prevention and mitigation strategies [23]. In the context of action-aware world models, this means restricting access to models that encode sensitive manipulation capabilities, monitoring for unexpected deployment patterns, and establishing clear lines of accountability when autonomous systems cause harm. Effective governance will require coordination among researchers, manufacturers, insurers, and public regulators.

7. Deployment, Sustainability, and Policy Implications

Deployment of action-aware world models at scale depends on solving a range of integration problems that span software, hardware, and organizational boundaries. One such problem is the transfer of policies trained on offline datasets to online interaction. Offline reinforcement learning methods have shown that agents can be initialized from previously collected demonstrations and then fine-tuned with online data to accelerate learning while reducing risk [24]. In manipulation, this is especially valuable because it allows an action-aware world model to be pretrained on abundant but potentially suboptimal data, then adapted to a specific robot and task with relatively little physical interaction. Visual planning approaches that learn to predict future observations conditioned on candidate actions have also demonstrated the value of anticipatory representations for manipulation [25]. These methods can be combined

with action-aware memory to create systems that reason not only about immediate next observations but also about the longer-term consequences of action sequences.

Sustainability is an increasingly important criterion for evaluating these systems. Training large world models and repeatedly querying them during planning consume substantial computational energy. A robot deployed for years in a home or factory will accumulate an ongoing environmental footprint that depends on how efficiently its models are designed and updated. Techniques such as parameter sharing, knowledge distillation, and sparse attention can reduce inference energy without sacrificing task performance. The design of action-aware memory should also consider whether the system can selectively update only the most relevant memory entries rather than retraining the entire model after every new experience. This incremental updating is not only efficient but also more consistent with how physical systems should behave in continuously changing environments. From a policy standpoint, sustainability metrics should be included in technology assessments and procurement decisions, particularly when public institutions consider adopting autonomous manipulation systems at scale.

Finally, the policy implications of action-aware world models extend to labor, accessibility, and urban infrastructure. Robotic manipulation systems that can plan over long horizons may alter the organization of work in warehouses, hospitals, and homes. They may also improve independence for people with disabilities by automating physically demanding tasks. However, these benefits are not automatic. They depend on how the technology is governed, who controls the underlying data and models, and whether deployment decisions include affected communities. A responsible research agenda should therefore integrate technical development with empirical studies of user needs, safety outcomes, and societal impacts. This requires sustained interdisciplinary collaboration across robotics, machine learning, human-computer interaction, law, and public policy. Action-aware world models are not simply a technical advance; they are a new sociotechnical infrastructure whose design will shape the distribution of capabilities, risks, and responsibilities in physical environments.

8. Conclusion

Action-aware world models represent a significant architectural direction for long-horizon robotic manipulation and task planning. By integrating predictive dynamics, action-conditioned memory, attention, and hierarchical planning, these systems can move beyond the limitations of passive world models and reactive visuomotor policies. The core insight is that effective long-horizon behavior requires the model to track not only what the environment looks like but also what the agent did, what it expected to happen, and what actually occurred. This action-conditioned memory enables more robust state estimation, more coherent task sequencing, and safer failure recovery in partially observable environments. However, realizing these benefits requires careful attention to system-level trade-offs in representational capacity, inference latency, data infrastructure, and continual learning. It also requires robust governance frameworks that address safety, fairness, accountability, and sustainability. As robotic manipulation systems become more capable and more widely deployed, the design of their world models will increasingly become a matter of public interest. Future research should combine algorithmic advances with rigorous empirical evaluation in real-world settings, participatory specification of tasks, and transparent documentation of limitations. In this way, action-aware world models can evolve from promising research prototypes into dependable components of embodied intelligent systems that serve human needs while respecting human values and institutional safeguards.

References

1. Sutton, R. S. (1991). Dyna, an integrated architecture for learning, planning, and reacting. *ACM SIGART Bulletin*, 2(4), 160-163.
2. Ha, D., & Schmidhuber, J. (2018). World models. *arXiv preprint arXiv:1803.10122*.
3. Hafner, D., Lillicrap, T., Fischer, I., Villegas, R., Ha, D., Lee, H., & Davidson, J. (2019). Learning latent dynamics for planning from pixels. *Proceedings of the 36th International Conference on Machine Learning*, 2555-2565.
4. Hafner, D., Lillicrap, T., Norouzi, M., & Ba, J. (2021). Mastering Atari with discrete world models. *International Conference on Learning Representations*.
5. Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., ... & Silver, D. (2020). Mastering Atari, Go, chess and shogi by planning with a learned model. *Nature*, 588(7839), 604-609.
6. Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., ... & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484-489.
7. Levine, S., Finn, C., Darrell, T., & Abbeel, P. (2016). End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17(39), 1-40.
8. Nagabandi, A., Kahn, G., Fearing, R. S., & Levine, S. (2018). Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning. *IEEE International Conference on Robotics and Automation*, 7559-7566.
9. Chua, K., Calandra, R., McAllister, R., & Levine, S. (2018). Deep reinforcement learning in a handful of trials using probabilistic dynamics models. *Advances in Neural Information Processing Systems*, 31, 4754-4765.
10. Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. *Proceedings of the 34th International Conference on Machine Learning*, 1126-1135.
11. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998-6008.
12. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.
13. Lynch, C., & Sermanet, P. (2021). Language conditioned imitation learning over unstructured data. *Robotics: Science and Systems*, 1-12.
14. 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*.
15. Driess, D., Xia, F., Sajjadi, M. S. M., Lynch, C., Chowdhery, A., Ichter, B., ... & Florence, P. (2023). PaLM-E: An embodied multimodal language model. *Proceedings of the 40th International Conference on Machine Learning*, 8469-8488.
16. LeCun, Y. (2022). A path towards autonomous machine intelligence. *OpenReview preprint*.

17. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.
18. 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.
19. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & Zoph, B. (2023). GPT-4 technical report. arXiv preprint arXiv:2303.08774.
20. Zellers, R., Holtzman, A., Bisk, Y., Farhadi, A., & Choi, Y. (2019). HellaSwag: Can a machine really finish your sentence? Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 4791-4800.
21. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. arXiv preprint arXiv:1606.06565.
22. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency, 220-229.
23. Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. arXiv preprint arXiv:1802.07228.
24. Nair, A., Gupta, A., Dalal, M., & Levine, S. (2020). AWAC: Accelerating online reinforcement learning with offline datasets. arXiv preprint arXiv:2006.09359.
25. Ebert, F., Finn, C., Lee, A. X., & Levine, S. (2017). Self-supervised visual planning with temporal skip connections. Proceedings of the 1st Annual Conference on Robot Learning, 344-356.