Uncertainty-Aware Planning with Generative World- and Language-Models via Monte Carlo Tree Search

Magí Dalmau-Moreno[1,2], Néstor García[1], and Vicenç Gómez[2]
[1]Eurecat, Technology Centre of Catalonia, Barcelona, Spain
[2]Universitat Pompeu Fabra, Barcelona, Spain
{magi.dalmau, nestor.garcia}@eurecat.org, vicen.gomez@upf.edu

***Abstract*—Robots acting in household environments must learn to plan long-horizon tasks in the presence of perceptual uncertainty, sparse rewards, and imperfect models of dynamics. While Monte Carlo Tree Search (MCTS) is a powerful tool for sequential decision making, its classical assumptions of an accurate simulator and well-shaped rewards do not hold in realistic robotic settings. In this work, we present an uncertainty-aware MCTS framework that combines a learned generative world model for imagined rollouts, a vision–language model (VLM) for progress-based shaping, and multimodal LLM (M-LLM) action priors. A hybrid upper confidence bound (UCB) integrates uncertainty from the world model, the VLM scorer, and the prior policy to balance exploration and risk aversion. In AI2-THOR long-horizon household tasks (15–25 steps), preliminary results suggest promising trends in success rate and planning efficiency compared to ablations (world-model only, shaping only, or priors only). While these findings are limited to simulation and remain to be validated more thoroughly, they illustrate a potential path toward safer and more effective deployment of learned generative models in robotics.**

I. Introduction

Robotic agents acting in everyday household environments must plan and execute long-horizon tasks under uncertainty. Consider a service robot that prepares a cup of tea: it must localize and grasp the kettle, fill it with water, switch on the stove, wait until boiling, and finally pour the water into a cup. Each step requires reasoning over a partially observable, high-dimensional environment, where progress is only indirectly observable and small execution errors can accumulate. While classical model-based planning methods such as Monte Carlo Tree Search (MCTS) have proven highly effective in structured domains such as board games [1], their success hinges on access to an accurate simulator and well-defined reward signals. In robotic household scenarios, these assumptions rarely hold: accurate physics models are unavailable, rewards are sparse, and the state space is perceptual and continuous.

Recent advances in *generative world models* offer a promising alternative: instead of relying on exact simulators, an agent can *imagine* possible futures by rolling out trajectories from a learned dynamics model [2, 3].However, world models are imperfect and prone to compounding errors, particularly over long horizons. This makes value estimation from imagined rollouts highly unreliable. In parallel, progress in *vision–language models* (VLMs) has enabled robust prediction of semantic properties from images, including estimates of task progress or completion [4]. Such signals are attractive for robotics because they provide dense, human-aligned feedback without manual reward engineering. Yet, these estimates are noisy and their reliability varies with context.

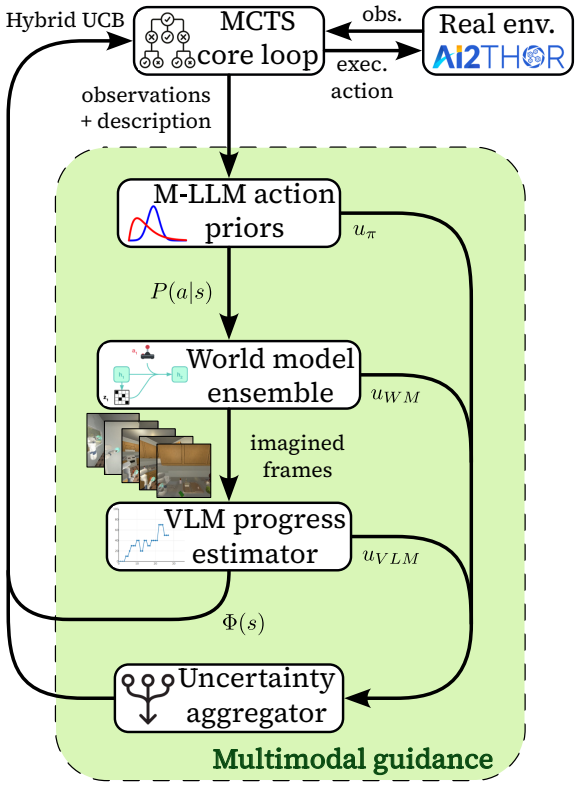

Fig. 1. Overview of the proposed framework. A multimodal large language model (M-LLM) provides semantic action priors $\pi(a|s)$ together with its uncertainty $u_\pi(s)$. A generative world model (WM) produces imagined rollouts $\hat{s}$, from which state transitions and associated uncertainty $u_{\text{WM}}(s,a)$ are estimated. A vision–language model (VLM) evaluates task progress by assigning a score $\Phi(s) \in [0,1]$ and an uncertainty estimate $u_{\text{VLM}}(s)$. These signals are fused in a Hybrid UCB formula within an uncertainty-aware Monte Carlo Tree Search (MCTS), balancing exploration and exploitation. The action selected at the root is then executed in the AI2-THOR environment, and the resulting observation updates the state before the next planning cycle.

In this work, we propose a planning framework that integrates generative imagination, multimodal perception, and uncertainty-aware search (see Figure 1). At its core, we extend MCTS with three key ingredients:

1) **Progress-based value shaping.** A VLM estimates task completion percentages from imagined states, which we

incorporate via potential-based reward shaping. In the ideal case, such shaping preserves optimality; in our setting with learned dynamics and noisy VLM signals, it serves as a dense surrogate signal that guides search.

2) **Multimodal priors for action guidance.** A large multimodal language model (M-LLM) conditions on the task description and current observation to provide a prior distribution over plausible next actions. This prior helps accelerate tree search by focusing exploration on semantically relevant branches.
3) **Uncertainty-aware exploration.** We introduce a hybrid Upper Confidence Bound (UCB) criterion that integrates uncertainty from the world model, the VLM scorer, and the prior policy. This enables the planner to interpolate between exploratory and risk-averse behavior, adapting naturally from training (where broad exploration is beneficial) to deployment (where caution is critical).

We evaluate our approach in the AI2-THOR household environment [5], treating it as the "real world" for robotic execution, while relying on a learned generative world model for imagined rollouts. Tasks such as tidying a room require compositional planning over 15–25 steps, are visually grounded, and have clear notions of progress. On these benchmarks, we observe promising trends toward higher success rates and improved planning efficiency relative to ablations. While these findings are limited to simulation and preliminary in scope, they suggest a potential path toward more risk-aware and efficient long-horizon planning for embodied agents in everyday settings.

II. Related Work

a) Model-based planning with MCTS: Monte Carlo Tree Search has emerged as a powerful planning algorithm in domains where a simulator is available, most famously in Go and other board games [1]. By combining tree search with value and policy priors, methods such as MuZero have extended MCTS to environments with learned dynamics models [3]. In robotics, MCTS has been applied to task and motion planning [6, 7], but practical deployment remains limited by the need for accurate simulators and well-shaped rewards. Our approach builds on this tradition but replaces the perfect simulator with a *generative world model* and supplements reward signals with vision–language feedback.

b) Generative world models for planning: Learning latent dynamics models enables agents to "imagine" futures without direct environment interaction [2, 8, 9]. Recent work has explored diffusion models [10, 11] and transformers for predicting realistic sequences of states. While powerful, these models accumulate error during long rollouts, leading to unreliable value estimates. Ensemble methods [12] and epistemic uncertainty quantification partially mitigate this issue, but they still require a robust mechanism for integrating uncertain imagined futures into planning. We address this by combining world-model imagination with task progress estimation and uncertainty-aware exploration.

c) Vision–language models as progress estimators: Large-scale VLMs have demonstrated strong capabilities in perceiving semantic properties of scenes and tasks from raw pixels [13, 14, 15]. In robotics, VLMs have been used for affordance reasoning and goal specification [4, 16], and more recently for monitoring task progress [17, 18]. Such estimates provide dense, human-aligned feedback, but are noisy and context-dependent. Our work incorporates VLM-derived progress signals via potential-based reward shaping, ensuring that they guide exploration without altering the optimal solution set.

d) Language priors for robotic planning: LLMs and M-LLMs have been applied to robotics as sources of semantic priors over action sequences, instructions, or skills [19, 20, 21]. These priors accelerate planning by biasing search towards linguistically plausible behaviours, but they are often brittle when used alone. In our framework, an M-LLM provides action priors that guide MCTS exploration while being balanced by model-based value estimation and uncertainty signals.

e) Uncertainty in model-based control: Accounting for uncertainty in learned dynamics has long been recognized as crucial for safe control [12, 22]. Recent planning algorithms incorporate epistemic uncertainty into tree search or rollout evaluation [23, 24]. In addition, risk-sensitive MCTS variants have been proposed in reinforcement learning [25, 26]. Our contribution is to combine uncertainty estimates from *three different sources*—the world model, the VLM scorer, and the prior policy—into a unified UCB criterion that interpolates between uncertainty-seeking exploration and risk-averse decision making.

f) Uncertainty-aware planning with inaccurate models: Recent work has also explored incorporating epistemic uncertainty into tree search for robotics tasks beyond classical motion planning. Faroni et al. [27] proposed an uncertainty-aware Monte Carlo Tree Search for robotized liquid pouring, where the planner leverages the variance of Gaussian Process models to bias exploration toward actions with lower predictive uncertainty. Their results demonstrate that explicitly reasoning about model inaccuracies can significantly improve task success rates even with minimal training data, highlighting the importance of integrating epistemic uncertainty into planning under imperfect dynamics.

g) Summary: In contrast to prior work, our framework unifies generative world models, VLM-based progress shaping, M-LLM priors, and uncertainty-aware exploration within a single MCTS algorithm. This integration enables long-horizon planning in visually rich household environments, where neither learned models nor vision–language signals alone are sufficiently reliable.

III. Methodology

We propose an uncertainty-aware extension of MCTS that integrates a generative world model, a vision–language progress estimator, and a multimodal language model prior policy. An overview of the framework is shown in Figure 1,

Algorithm 1 Uncertainty-aware MCTS with generative world models, VLM shaping, and M-LLM priors.

Require: Root state s_0, task description $\mathcal{T}$, world model $\hat{T}$, VLM scorer Φ, multimodal LLM prior π_{LLM}, search budget B, horizon H, discount γ, shaping weight β, terminal bootstrap α, uncertainty weights λ, hybrid schedule parameter ϕ.

1: **for** $b = 1 \dots B$ **do** ▷ Run B simulations

2: **Selection:**
Start at root node s_0.
While node s is fully expanded and not terminal:
Select action a maximizing hybrid UCB with Eq. 5.
Move to child node $s' = \hat{T}(s, a)$.

3: **Expansion:**
Expand node s by adding all legal actions.
Priors $P(a|s)$ obtained from $\pi_{\text{LLM}}(a|s, \mathcal{T})$.

4: **Rollout (Imagination + Shaping):**
Simulate trajectory $(s_0, a_0, \dots, s_H)$ using $\hat{T}$.
For each step t, compute discounted return with optional terminal bootstrap following Eq. 2.

5: **Uncertainty Estimation:**
Compute u_{WM} (world model variance), u_{VLM} (progress entropy/variance), u_π (policy entropy).
Combine as $u(s, a) = \lambda_{\text{WM}} u_{\text{WM}} + \lambda_{\text{VLM}} u_{\text{VLM}} + \lambda_\pi u_\pi$.

6: **Backup:**
For each (s, a) in the path:
$N(s, a) \leftarrow N(s, a) + 1$
$W(s, a) \leftarrow W(s, a) + G$
$Q(s, a) \leftarrow W(s, a) \,/\, N(s, a)$
Cache uncertainties $u(s, a)$ for next selection.

7: **Return:** Action a from root with highest visit count $N(s_0, a)$.

and the procedure is formalized in Algorithm 1. Each simulation selects a path using Hybrid UCB, expands the tree with priors from the M-LLM, rolls out a trajectory in imagination with VLM-shaped rewards, computes uncertainty estimates, and backs up returns and statistics. After B simulations, the planner executes the root action with highest visit count.

A. Problem setup

We consider an episodic decision-making problem modelled as an MDP $\mathcal{M} = (\mathcal{S}, \mathcal{A}, T, r, \gamma)$. States $s \in \mathcal{S}$ are high-dimensional observations (e.g., RGB frames and proprioceptive signals), actions $a \in \mathcal{A}$ are discrete manipulation and navigation primitives, and rewards $r(s, a, s')$ are sparse, typically provided only upon task completion. The agent has access to: (i) a generative world model $\hat{T}$ trained on real experience, (ii) a VLM that outputs task progress estimates $\Phi(s) \in [0, 1]$, and (iii) a M-LLM that provides action priors conditioned on task description $\mathcal{T}$ and current state.

B. Generative world model for imagination

Instead of relying on a perfect simulator, we generate imagined trajectories $(s_0, a_0, \dots, s_H)$ using a learned dynamics model $\hat{T}$. We employ either an ensemble of latent dynamics model. The world model additionally provides an epistemic uncertainty estimate $u_{\text{WM}}(s, a)$, derived from ensemble variance.

C. Progress-based value shaping

We query the VLM for a task completion score $\Phi(s)$, to provide dense feedback and define a shaped reward:

$$r'(s, a, s') = r(s, a, s') + \beta\big(\gamma\, \Phi(s') - \Phi(s)\big). \tag{1}$$

The telescoping property ensures that potential-based shaping does not alter the optimal policy. For rollouts of horizon H, we compute the discounted return

$$G = \sum_{t=0}^{H-1} \gamma^t\, r'(s_t, a_t, s_{t+1}) + \alpha\, \gamma^H\, \Phi(s_H), \tag{2}$$

where $\alpha \in [0, 0.3]$ optionally includes a small terminal bootstrap from the VLM. Note that the equivalence guarantees of potential-based shaping apply under the environment's true dynamics with a fixed, state-only potential; with learned dynamics and per-trajectory calibration, this acts as a heuristic that empirically improves search signals.

D. Multimodal LLM prior policy

At each node, we query a M-LLM conditioned on the task description and observation:

$$P(a \mid s, \mathcal{T}) = \pi_{\text{LLM}}(a \mid s, \mathcal{T}). \tag{3}$$

This prior is injected into MCTS expansion, biasing exploration towards semantically plausible actions. The entropy of this distribution $u_\pi(s) = -\sum_a P(a \mid s) \log P(a \mid s)$ serves as an uncertainty measure.

E. Uncertainty integration

We combine three uncertainty signals u_{WM} (world model variance), u_{VLM} (progress entropy/variance), u_π (policy entropy):

$$u(s, a) = \lambda_{\text{WM}}\, u_{\text{WM}}(s, a) + \lambda_{\text{VLM}}\, u_{\text{VLM}}(s) + \lambda_\pi\, u_\pi(s), \tag{4}$$

where λ are tunable weights. Here, u_{VLM} denotes the uncertainty of the progress scorer, estimated via predictive entropy or sampling variance.

F. Hybrid UCB for selection

During selection, each action is scored using a hybrid upper confidence bound:

$$\begin{aligned} \text{UCB}(s, a) = &\big(Q(s, a) - \rho(\phi)\, u(s, a)\big) \\ &+\; c_{\text{puct}} \frac{P(a|s)\sqrt{N(s)}}{1 + N(s, a)} + c_u(\phi)\, u(s, a), \end{aligned} \tag{5}$$

where $Q(s, a)$ is the estimated value, $N(s, a)$ visit counts, and $P(a|s)$ the LLM prior. The schedule parameter $\phi \in [0, 1]$ interpolates between exploration ($\phi = 0$) and risk-aversion ($\phi = 1$):

$$\rho(\phi) = \rho_{\max}\phi, \quad c_u(\phi) = c_{u,\max}(1 - \phi).$$

IV. EXPERIMENTS

We evaluate our approach on long-horizon household tasks in the AI2-THOR environment, treating it as the execution domain while relying on a generative world model for imagined rollouts. Our experiments are designed to answer three questions:

1) Does VLM-based potential shaping improve value estimation from imagined trajectories?
2) Do multimodal LLM priors accelerate MCTS exploration?
3) Does uncertainty-aware UCB improve robustness compared to risk-neutral or purely exploratory search?

A. Experimental setup

a) Environment: We use AI2-THOR as a proxy for real household environments. Agents perceive RGB observations and act via discrete primitives such as `PickUp`, `Put`, `Open`, `Close`, `ToggleOn`, and navigation actions. We treat these as low-level actions; tasks are specified in natural language (e.g., *"Set the table"*) and provided to the M-LLM.

b) Tasks: We consider long-horizon tasks, where agents must compose 15–25 sequential actions over multiple objects. For this preliminar evaluation we tested with five unseen tasks, naming: T1: Put a clean mug in the microwave, T2: Store a slice of bread in the fridge, T3: Place a knife from the drawer on the dining table, T4: Throw away an apple in the trash can, and T5: Put a pan on the stove. To achieve these goals, the robot needs to combine coherently actions such as opening and closing containers, picking up different objects, and placing them in specific target locations.

c) Generative world model: We trained an ensemble of models on collected trajectories from AI2-THOR. The model predicts next latent states and reconstructions from (s, a) pairs. Ensemble variance serves as u_{WM}. We finetuned the Unified World Model [28] on a dataset we generated from ALFRED tasks and rendered with AI2THOR Simulator, with RGB observations, per-step actions/masks, and next-frame targets. We initialized from publicly released UWM weights and trained with a two-stage curriculum: In *Phase 1*, we freeze perception and tune only the diffusion head to align action-conditioned dynamics to the new dataset; in *Phase 2*, we partially unfreeze early encoder layers and raise the action-free mix to adapt perception to the domain while leveraging unlabeled video for robustness.

d) VLM progress scorer: We query a vision–language model GPT-4o with a task description and an image to obtain progress estimates $\Phi(s) \in [0, 1]$, calibrated via monotonic regression per trajectory. While such calibration does not provide the state-only potential function assumed in the theoretical shaping guarantee, it serves as a practical heuristic to supply denser feedback. Uncertainty u_{VLM} is estimated via predictive entropy from multiple temperature-scaled samples.

e) Multimodal LLM Prior: We use a multimodal LLM to obtain action priors $P(a|s, \mathcal{T})$ from task description and observations. The entropy of this distribution $u_\pi(s)$ is included in the uncertainty aggregator.

TABLE I
PERFORMANCE ON LONG-HORIZON TASKS (15–25 STEPS).

Method	Success rate↑	Efficiency↓	Feasibility violations↓
MCTS + WM	24%	2,400	17%
MCTS + WM + VLM	34%	1,700	12%
MCTS + WM + LLM	31%	1,200	10%
Ours (Full)	49%	900	6%

f) Evaluation protocol: Each task is run for N episodes from randomized initial conditions. MCTS search is run with budget B simulations per decision, horizon H steps per rollout, and discount $\gamma = 0.99$. We compare to baselines and ablations described below.

B. Baselines

- **MCTS + WM:** Planning in imagination without shaping or priors.
- **MCTS + WM + VLM:** Planning in imagination with progress shaping but no LLM prior or uncertainty-aware UCB.
- **MCTS + WM + LLM:** Planning in imagination with semantic priors but no shaping or uncertainty-aware UCB.
- **Ours (Full):** World model + VLM shaping + LLM priors + uncertainty-aware UCB.

C. Metrics

- **Task success rate:** fraction of episodes completing the task within a maximum number of steps.
- **Planning efficiency:** average number of simulations required to select a successful plan.
- **Feasibility and robustness:** frequency of invalid actions (collisions, drops) and failure to recover from model errors.

D. Results

a) Quantitative: Table I reports success rates, efficiency, and correctness metrics on the considered long-horizon tasks across baselines.

b) Qualitative: We visualize imagined rollouts with progress estimates, show search tree expansion with and without LLM priors, and highlight how uncertainty guides exploration in ambiguous states. These examples illustrate how the different components of our framework contribute to more efficient and reliable planning in practice.

c) Ablation studies: We compare the effects of removing (i) VLM shaping, (ii) LLM priors, and (iii) uncertainty-aware UCB. Results highlight the contribution of each component.

E. Discussion of findings

Our evaluation focused on long-horizon tasks, since these scenarios best highlight the strengths of our framework: the overhead of imagination, progress shaping, and multimodal priors is outweighed by the need for deeper search. Results show that uncertainty-aware UCB provides the largest gains in visually ambiguous scenarios, while VLM shaping consistently improves planning reliability. LLM priors are most

beneficial in compositional tasks where high-level sequencing is critical.

V. Discussion and Future Work

a) Contributions: We introduced an uncertainty-aware MCTS framework that integrates generative world models, vision–language progress estimators, and multimodal language model priors. Our approach extends classical tree search with (i) potential-based reward shaping from VLM progress signals, (ii) semantic action priors from M-LLMs, and (iii) a hybrid UCB criterion that fuses uncertainty from world models, VLM scorers, and priors. We observe improvements on our tasks; however, results are preliminary and limited to AI2-THOR.

b) Limitations: Despite these encouraging results, some challenges remain. First, the fidelity of imagined rollouts is ultimately constrained by the generative world model, and compounding errors may still degrade value estimation in long-horizon tasks. Second, querying the VLM for progress signals introduces latency and cost, and its predictions may reflect biases from pretraining data. Third, our current action space is limited to discrete primitives; extending the framework to continuous control or higher-level skills will require additional abstractions. Finally, our evaluation is restricted to simulation; validating robustness, latency, and safety on physical robots is a crucial next step.

c) Future directions: Building on these insights, several promising directions emerge. A key challenge is *state canonicalization*: consolidating similar real and imagined states in latent space to avoid value fragmentation and improve sample efficiency. Another promising direction is *adaptive scheduling* of the hybrid UCB parameter ϕ, allowing the planner to automatically shift between exploration and risk aversion depending on real-time uncertainty. Extending the framework to hierarchical settings, where LLMs suggest high-level skills while MCTS refines low-level actions, could further enhance scalability. Finally, deploying the approach on real robotic platforms will provide valuable insights into robustness, latency, and safety in unstructured environments.

d) Broader impact: Our approach highlights how integrating generative world models with multimodal foundation models can enable robots to tackle long-horizon tasks in unstructured human environments with greater reliability. This has the potential to broaden the set of applications where robots can safely assist people, from household support to industrial collaboration, reducing the cognitive and physical burden on humans. At the same time, reliance on large-scale models raises important considerations: energy and data costs, biases inherited from pre-training corpora, and the risks of unsafe behaviour in safety-critical settings. By incorporating explicit uncertainty reasoning, the framework takes a step toward mitigating these risks-encouraging risk-aware planning, more transparent decision making, and safer deployment in everyday contexts. We see this as a promising direction for bridging the gap between powerful but opaque foundation models and the safety demands of embodied AI. Nevertheless, note that real-robot latency, perception drift, and compounding model error remain open

VI. Conclusion

We presented an uncertainty-aware extension of Monte Carlo Tree Search that integrates generative world models, vision–language progress scoring, and multimodal large language model priors. By combining potential-based reward shaping with a hybrid UCB criterion that accounts for model, perception, and policy uncertainty, our framework shows promise for robust long-horizon planning in simulation. Experiments in AI2-THOR show that this integration improves task success rates, planning efficiency, and robustness compared to baselines relying on either world models, VLM shaping, or LLM priors alone.

Looking forward, we see several opportunities to extend this work, including broader benchmarking, latent-space state canonicalization, hierarchical skill-level planning, and real-robot deployment. More broadly, our results highlight the promise of combining generative models and multimodal foundation models with principled planning algorithms, bringing us closer to embodied agents that can operate reliably in everyday settings.

VII. Acknowledgments

This work was financially supported by the European Commission's Horizon Europe Framework Program through the IntelliMan project under Grant Agreement No. 101070136; and by the Catalan Government through ACCIO-Eurecat (Project Flagship-RONOUS) funding grant. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the funding institutions. The funding institutions cannot be held responsible for them.

References

[1] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, and others, "Mastering the game of Go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[2] D. Hafner, T. Lillicrap, M. Norouzi, and J. Ba, "Dream to control: Learning behaviors by latent imagination," in *International Conference on Learning Representations (ICLR)*, 2020.

[3] J. Schrittwieser, I. Antonoglou, T. Hubert, K. Simonyan, L. Sifre, S. Schmitt, A. Guez, E. Lockhart, D. Hassabis, T. Graepel, and others, "Mastering Atari, Go, chess and shogi by planning with a learned model," *Nature*, vol. 588, no. 7839, pp. 604–609, 2020.

[4] M. Shridhar and others, "LLMs for embodied agents," in *Robotics: Science and Systems (RSS)*, 2023.

[5] E. Kolve, R. Mottaghi, W. Han, E. VanderBilt, L. Weihs, A. Herrasti, M. Deitke, K. Ehsani, D. Gordon, Y. Zhu, A. Kembhavi, A. Gupta, and A. Farhadi, "AI2-THOR: An Interactive 3D Environment for Visual AI," Aug. 2022. arXiv:1712.05474 [cs].

[6] R. Chitnis, T. Silver, B. Kim, L. Kaelbling, and T. Lozano-Perez, "CAMPs: Learning Context-Specific Abstractions for Efficient Planning in Factored MDPs," in *Proceedings of the 2020 Conference on Robot Learning* (J. Kober, F. Ramos, and C. Tomlin, eds.), vol. 155 of *Proceedings of Machine Learning Research*, pp. 64–79, PMLR, Nov. 2021.

[7] B. Kim, L. P. Kaelbling, and T. Lozano-Pérez, "Learning to Guide Task and Motion Planning Using Score-Space Representation," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2017.

[8] D. Hafner, T. Lillicrap, M. Norouzi, and J. Ba, "Learning latent dynamics for planning from pixels," *arXiv preprint arXiv:1811.04551*, 2019.

[9] D. Ha and J. Schmidhuber, "World models," in *NeurIPS Workshop on Advances in Neural Information Processing Systems*, 2018.

[10] M. Janner, Y. Li, J. B. Tenenbaum, and S. Levine, "Planning with diffusion for flexible behavior synthesis," in *International Conference on Machine Learning (ICML)*, 2022.

[11] X. Wang and others, "Diffusion Policies: Visuomotor Policy Learning via Action Diffusion," in *Robotics: Science and Systems (RSS)*, 2023.

[12] K. Chua, R. Calandra, R. McAllister, and S. Levine, "Deep reinforcement learning in a handful of trials using probabilistic dynamics models," in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 4754–4765, 2018.

[13] J.-B. Alayrac and others, "Flamingo: a Visual Language Model for Few-Shot Learning," *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[14] D. Driess, F. Xia, M. S. M. Sajjadi, C. Lynch, A. Chowdhery, B. Ichter, A. Wahid, J. Tompson, Q. Vuong, T. Yu, W. Huang, Y. Chebotar, P. Sermanet, D. Duckworth, S. Levine, V. Vanhoucke, K. Hausman, M. Toussaint, K. Greff, A. Zeng, I. Mordatch, and P. Florence, "PaLM-E: An Embodied Multimodal Language Model," in *Proceedings of the 40th International Conference on Machine Learning*, pp. 8469–8488, PMLR, July 2023. ISSN: 2640-3498.

[15] J. Gao, B. Sarkar, F. Xia, T. Xiao, J. Wu, B. Ichter, A. Majumdar, and D. Sadigh, "Physically Grounded Vision-Language Models for Robotic Manipulation," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 12462–12469, 2024.

[16] A. Brohan and others, "RT-1: Robotics transformer for real-world control at scale," *Robotics: Science and Systems (RSS)*, 2023.

[17] H. Liu, W. Chen, and L. Fei-Fei, "ProgPrompt: Progress Estimation for Instructional Videos via Multimodal Prompts," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[18] X. Huang, Y. Xu, and J. Wang, "Visual Progress Estimation with Vision-Language Models," *arXiv preprint arXiv:2401.01234*, 2024.

[19] M. Ahn, A. Brohan, Y. Chebotar, and others, "Do as I can, not as I say: Grounding language in robotic affordances," in *Proceedings of Robotics: Science and Systems (RSS)*, 2022.

[20] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, X. Chen, K. Choromanski, T. Ding, D. Driess, A. Dubey, C. Finn, P. Florence, C. Fu, M. G. Arenas, K. Gopalakrishnan, K. Han, K. Hausman, A. Herzog, J. Hsu, B. Ichter, A. Irpan, N. Joshi, R. Julian, D. Kalashnikov, Y. Kuang, I. Leal, L. Lee, T.-W. E. Lee, S. Levine, Y. Lu, H. Michalewski, I. Mordatch, K. Pertsch, K. Rao, K. Reymann, M. Ryoo, G. Salazar, P. Sanketi, P. Sermanet, J. Singh, A. Singh, R. Soricut, H. Tran, V. Vanhoucke, Q. Vuong, A. Wahid, S. Welker, P. Wohlhart, J. Wu, F. Xia, T. Xiao, P. Xu, S. Xu, T. Yu, and B. Zitkovich, "RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control," *arXiv preprint arXiv:2307.15818*, July 2023.

[21] W. Huang, F. Xia, T. Xiao, H. Chan, J. Liang, P. Florence, A. Zeng, J. Tompson, I. Mordatch, Y. Chebotar, P. Sermanet, N. Brown, T. Jackson, L. Luu, S. Levine, K. Hausman, and B. Ichter, "Inner Monologue: Embodied Reasoning through Planning with Language Models," in *Proceedings of Robotics: Science and Systems (RSS)*, arXiv, July 2022. arXiv:2207.05608 [cs].

[22] M. Deisenroth and C. E. Rasmussen, "PILCO: A model-based and data-efficient approach to policy search," in *Proceedings of the 28th International Conference on Machine Learning (ICML)*, pp. 465–472, 2011.

[23] T. M. Moerland, J. Broekens, and C. M. Jonker, "Thinking fast and slow with deep learning and tree search," *arXiv preprint arXiv:2002.02819*, 2020.

[24] A. Guez, T. Weber, I. Antonoglou, D. P. Reichert, and D. Silver, "Value uncertainty for efficient reinforcement learning," in *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 3849–3859, 2021.

[25] P. Thomas, G. Theocharous, and M. Ghavamzadeh, "High-confidence off-policy evaluation," in *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, pp. 3000–3006, 2015.

[26] M. K. Hryniewicki, "Monte Carlo Tree Search with Risk Aversion," in *International Conference on Artificial Intelligence and Soft Computing (ICAISC)*, pp. 480–491, 2020.

[27] M. Faroni, C. Odesco, A. Zanchettin, and P. Rocco, "Uncertainty-aware Planning with Inaccurate Models for Robotized Liquid Handling," July 2025. arXiv:2507.20861 [cs].

[28] C. Zhu, R. Yu, S. Feng, B. Burchfiel, P. Shah, and A. Gupta, "Unified World Models: Coupling Video and Action Diffusion for Pretraining on Large Robotic Datasets," May 2025. arXiv:2504.02792 [cs].