

Reinforcement Learning with World Models in the Era of Scale: From Online Imagination to World Action Models

Wojciech Dąbrowicki

`wojciech.dabrowicki@pwr.edu.pl`

Wrocław University of Science and Technology

Andrzej Rusiecki

Wrocław University of Science and Technology

Research Article

Keywords: World Models, Model-Based Reinforcement Learning, Representation Learning, Offline Reinforcement Learning

Posted Date: October 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11327969/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Reinforcement Learning with World Models in the Era of Scale: From Online Imagination to World Action Models

Wojciech Dąbrowicki(0009-0004-2052-4052)^{1*†}
and Andrzej Rusiecki(0000-0003-2239-1076)^{1*†}

¹Department of Computer Science, Wrocław University of Science and Technology, Wybrzeże Stanisława Wyspiańskiego 27, Wrocław, 50-370, Lower Silesia, Poland.

*Corresponding author(s). E-mail(s): wojciech.dabrowicki@pwr.edu.pl;
andrzej.rusiecki@pwr.edu.pl;

[†]These authors contributed equally to this work.

Abstract

In World Model Reinforcement Learning (WM-RL), agents learn a parameterized model of environmental dynamics that serves as the basis for planning, imagination-based policy training, or synthetic data generation in decision making tasks. From 2019 to 2026, a drastic phase transition took place within this field. Initially started as an effort to improve efficiency in visual control tasks, WM-RL has become the core of interdisciplinary AI infrastructure. Today, internet-scale video-based generative world foundation models allow the autonomous vehicle pipeline to learn generalist policies from unlabeled human video, facilitating applications in medical therapy and molecular dynamics. To review the entire trajectory of WM-RL research, it is necessary to synthesize key sources that summarize the advances made throughout the years. Three prominent paradigm shifts are observed during this period. These include the shift from online Model-Based Reinforcement Learning (MBRL) to pretrain-first imagination-only policy training; architecture integration from decoupled worlds and policies to unified World Action Models; and the generalization of specific algorithms to physical AI infrastructure for all MBRL applications. Regarding architecture, WM-RL follows several generations of innovations. They include PlaNet, Recurrent State-Space Model (RSSM), the Dreamer family, value-equivalent implicit models, Transformer-based world models, diffusion dynamics, and the abstract-space Joint-Embedding Predictive Architecture (JEPA)

predictor. In this review, we examine WM-RL from multiple aspects, ranging from general model-based reinforcement learning, robotic manipulation, navigation, autonomous driving, multi-agent systems, large language model agents, healthcare, and games. The usage of world models varies widely depending on the application areas, while the validated architectures are discussed accordingly.

Keywords: World Models, Model-Based Reinforcement Learning, Representation Learning, Offline Reinforcement Learning

1 Introduction

The ability to mentally simulate the consequences of potential decisions before executing any action has been widely accepted as one of the critical elements contributing to biological intelligence (Tolman 1948). The world model reinforcement learning framework provides an explicit formulation of this intuition. Instead of learning from pure trial-and-error, agents maintain a model of environment dynamics represented by parameters and use it for planning, training with generated imaginary experiences, or predicting future states along hypothetical action sequences. Many theoretical benefits arise, including lower sample complexity, risk-free exploration using imagination, and the ability to consider plans outside of agents’ current experience ranges.

Modern instances of world models originate from seminal World Models papers (Ha and Schmidhuber 2018). These papers illustrated the potential of employing a variational autoencoder to compress sensory percepts while utilizing a long short-term memory network to predict time-series dynamics. The CMA-ES controller would perform all training within the simulated environment (“dream”). Though seemingly simple from architectural perspectives, this work makes significant contributions to world models. Specifically, neural networks were able to learn compact and generative dynamics models solely from raw pixel input. Furthermore, such models could enable agents to train and transfer skills meaningfully to the real world.

A series of groundbreaking developments has followed since then. PlaNet (Hafner et al. 2019b) contributed recurrent state-space models (RSSM) and latent-space model predictive control methods (MPC). Later, DreamerV1 (Hafner et al. 2019a) introduced a latent actor-critic replacement for MPC for full imagination-based end-to-end training. Discrete categorical latents enabled DreamerV2 (Hafner et al. 2020) to achieve super-human Atari scores. Concurrently, MuZero (Schrittwieser et al. 2020) illustrated that value-equivalent world models combined with Monte Carlo Tree Search (MCTS) could solve chess, Go, Atari, and Shogi without relying on prior domain-specific information. EfficientZero improved sampling efficiency through unsupervised self-representation learning of MuZero. The MBPO paper (Janner et al. 2019) established the model-based policy optimization paradigm with uncertainty-aware rollout truncation techniques as a strong online model-based RL baseline algorithm. MOPO (Yu et al. 2020) subsequently proposed using the same pessimistic uncertainty penalty methodology for offline reinforcement learning. The COMBO (Yu et al. 2021) technique unified offline conservative Q-learning algorithms with offline world model

rollouts. SimPLe (Kaiser et al. 2020) demonstrated that even pixel-space world models can significantly increase sampling efficiencies in Atari games. Finally, the IRIS (Micheli et al. 2022) architecture achieved state-of-the-art Atari 100k results using a discrete autoencoder coupled with an autoregressive Transformer world model.

In the 2025-2026 period, the scope of world-model RL systems reached unprecedented scales. World models learned from billions of frames (Agarwal et al. 2025; Assran et al. 2025; Gao et al. 2026), deployed in production autonomous vehicle pipelines (Jiang et al. 2026a; Russell et al. 2025), and provided core functionality for self-evolving embodied agents (Bolton et al. 2025). This review comprehensively highlights all stages along this path.

As shown in Figure 1, publications on world-model reinforcement learning increased by more than 30 times from 2017 to 2025, witnessing a sharp increase since 2023. Such development trends coincide well with the upcoming pretrain-first and industry-scale strategies that will be introduced below. Figure 1 is intended as an indicative measure of the field’s growth rather than an exhaustive account of the literature covered in this review, since many recent contributions appear first as preprints that are not indexed in Scopus.

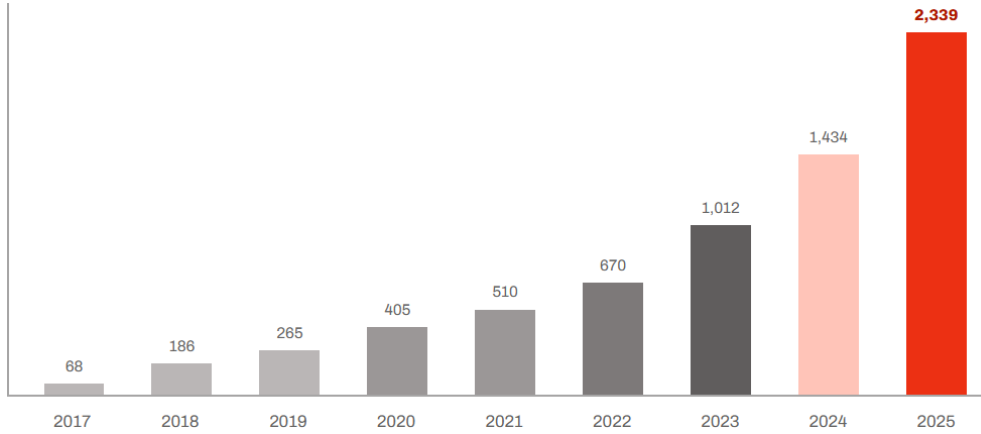


Figure 1 A systematic overview of the WM-RL publications in recent years reveals a dramatic surge in growth rate, increasing by more than thirty times from 2017 to 2025. Notably, there is an evident acceleration trend that begins in 2023, coinciding with the shift towards pretrain-first and scaled-up paradigms in WM-RL research. Publications during 2017-2025 were searched via Scopus using the query filter: (World Model) AND ((Reinforcement Learning) OR (RL))

In this paper, we define the concept of world models in reinforcement learning (WM-RL) and discuss the key features of current approaches based on functional paradigms and representation spaces. Further, we will discuss several critical differences between world modeling and other related areas. This survey presents our results of a comprehensive literature investigation by organizing WM-RL research under different criteria: the evolution of world model architectures from 2019 to 2026; application cases, including robotics, autonomous driving, multi-agent systems, and games;

and remaining research challenges. The structure of this paper is organized as follows. First, in Section 2, we provide formal definitions of world model RL problems. Specifically, value-equivalent world models and Joint Embedding Probabilistic Automata (JEPAs) are introduced. Next, we introduce functional paradigms and representation spaces, which serve as classification criteria for organizing the related literature. The evolution path of world model architectures is then summarized in Section 3. The path includes Gaussian processes, ensemble dynamics, RSSM and Dreamer family architectures, transformers, diffusion-based, decoder-free, object-centric, and state-space models. Sections 4, 5, and 6 discuss offline model-based reinforcement learning, exploration, and hierarchy and temporal abstraction, respectively. Application cases of world models cover robotic manipulation (Section 7); navigation and locomotion (Section 8); autonomous driving (Section 9); multi-agent systems (Section 10); large language model agents (Section 11); games (Section 12); and health sciences and discovery (Section 13). Finally, we conduct discussions of world action models (Section 14) and critical factors such as safety and uncertainty quantification (Section 15). Seven open research challenges are presented in Section 16, based on which the key findings are proposed in Section 17. This paper concludes in Section 18.

2 Formal Framework and Technical Background

2.1 The World Model RL Problem

Let us consider a POMDP (Partially Observable Markov Decision Process) with components $(\mathcal{S}, \mathcal{A}, \mathcal{O}, T, E, R, \gamma)$. The world model consists of an estimation function $(\hat{T}_\theta, \hat{R}_\theta, \hat{E}_\theta)$, along with an encoding function $q_\theta : \mathcal{H} \times \mathcal{O} \rightarrow \Delta(\hat{\mathcal{S}})$ that maps the deterministic hidden state and observation to a distribution over stochastic latents $\hat{\mathcal{S}}$. The learning of the policy takes place in this latent space. This framework offers several advantages: First, it allows computation in a lower dimensional latent space instead of dealing with observations in the high dimensional space. Second, it avoids real-world costs and risks during the imagination rollout.

The RSSM (Hafner et al. 2019b) decomposes the latent state z_t into stochastic s_t and deterministic h_t components:

$$h_t = f_\theta(h_{t-1}, s_{t-1}, a_{t-1}), \quad (\text{gated recurrent unit (GRU) transition}) \quad (1)$$

$$s_t \sim q_\theta(s_t | h_t, o_t), \quad (\text{posterior encoder}) \quad (2)$$

$$\hat{s}_t \sim p_\theta(\hat{s}_t | h_t). \quad (\text{prior dynamics}) \quad (3)$$

World model training maximizes the Evidence Lower Bound (ELBO):

$$\mathcal{L}(\theta) = \mathbb{E} \left[\sum_t \ln p_\theta(o_t | s_t, h_t) - \beta \text{KL}(q_\theta(s_t | h_t, o_t) \| p_\theta(\hat{s}_t | h_t)) \right] \quad (4)$$

Three improvements were introduced in DreamerV3 (Hafner et al. 2025a), allowing fixed-hyperparameter generalization for more than 150 tasks. Specifically, (i) *symlog prediction* was performed in DreamerV3, which means applying the function

$\text{symlog}(x) = \text{sgn}(x) \ln(|x| + 1)$ to the predictions for all prediction targets. In addition, (ii) *two-hot return* was encoded by the network in DreamerV3 for handling heavy-tailed return distribution, while (iii) *balanced KL with free bits* prevented posterior collapsing.

2.2 Value-Equivalence and Implicit Models

The value equivalence (VE) principle adopted by MuZero (Schrittwieser et al. 2020) and TD-MPC (Hansen et al. 2022, 2024) indicates that a world model does not need to predict observations; instead, it must learn values and policies equivalent to those of the true world. In their paper, "TD-MPC2: Scalable, robust world models for continuous control" (Hansen et al. 2024), the authors train an implicit world model with 317 million parameters, without a decoder, on 80 continuous control tasks under a single set of hyperparameters, suggesting that a decoder-free architecture can devote its representational capacity to value-relevant prediction. In addition, EfficientZero (Ye et al. 2021) adds a consistency loss function inspired by self-supervision to MuZero and obtains impressive results on Atari games with limited interactions (100k). UniZero (Pu et al. 2024) distinguishes latent states and contexts in transformers for building world models in order to conduct deep Monte Carlo Tree Search in POMDPs.

2.3 The Joint-Embedding Predictive Architecture

LeCun’s JEPA framework (LeCun 2022) predicts future *representations* instead of future *observations*. Specifically, given the context encoder f_θ , predictor g_ϕ , and target encoder $\bar{f}_\theta$ (exponential moving average of f_θ), the objective loss function can be defined as follows:

$$\mathcal{L}_{\text{JEPA}} = \mathbb{E}[\|g_\phi(f_\theta(x_{\text{ctx}}), \text{mask}) - \text{sg}(\bar{f}_\theta(x_{\text{target}}))\|^2]. \quad (5)$$

In this formula, the stop-gradient operation $\text{sg}(\cdot)$ aims to avoid representation collapse issues during training. The application of this framework is broad and diverse, extending to video predictions with short sequences (Bardes et al. 2024) or scaling up to more than one billion parameters with pretraining on over one million hours of internet video (Assran et al. 2025). After action-conditioned post-training on fewer than 62 hours of unlabeled robot video, the model supports zero-shot robot planning.

2.4 Three Functional Paradigms

We organized the literature by how the world model is functionally used:

1. **Online MBRL.** WM is trained online alongside the policy; the policy is trained via imagination rollouts. Standard Dreamer/TD-MPC/MuZero pipeline. Dominant pre-2025.
2. **Pretrain-first, offline imagination.** WM is pretrained on large-scale (often unlabeled) data; the policy is trained entirely inside the pretrained WM with no further real-environment interaction. Dominant 2025–2026.
3. **WM-as-simulator.** WM is used for policy evaluation, synthetic data generation, or safety validation. The policy may be trained by any method (RL,

(Behavioral Cloning) BC, (Vision-Language-Action) VLA), but the WM replaces or augments the environment.

2.5 Representation Space Overview

Apart from the classification of world models based on the functional paradigms introduced above, it is meaningful to consider different world models based on the representation space used to forecast the dynamics. As illustrated in Figure 2, there are five types of representation spaces, with exemplary architectures summarized. A detailed discussion of pixel space/world models and embedding space/world models can be found in Section 3.2 and Section 3.5, respectively. For object-centric world models, refer to Section 3.6. For 3D geometric world models used for robotic manipulation, refer to Section 7.3. Lastly, the language/symbolic space/world models are introduced in Section 11.

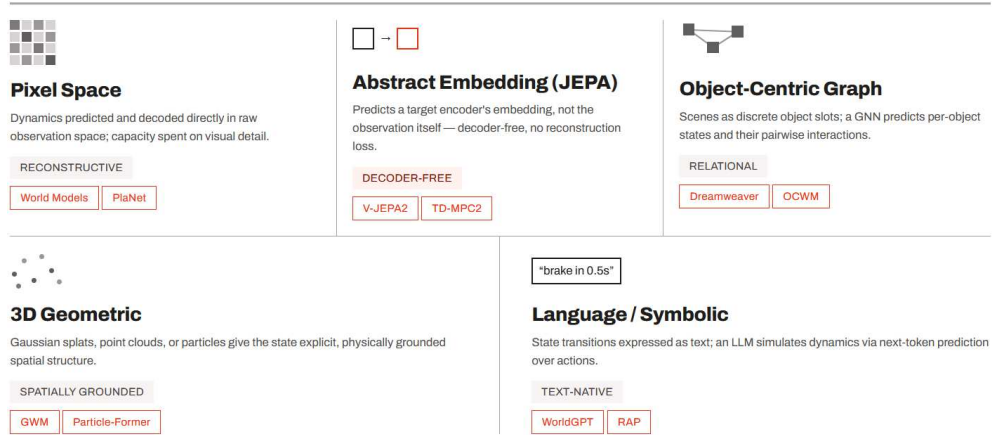


Figure 2 Taxonomy of representation spaces used in world model reinforcement learning, comprising pixel space, abstract embedding space, object-centric graph space, 3D geometric space, and language/symbolic space, with representative architectures shown for each category

3 Architectural Lineage of World Models

In recent years, world model architecture design has gone through several stages of improvement. Each stage builds upon past work by introducing new architectural components or techniques, such as different inductive bias assumptions and training techniques. As shown in Figure 3, this evolution started with dynamics modeling via Gaussian processes, progressed to the Dreamer architecture based on RSSM dynamics modeling, transitioned to decoder-free and transformer-based world model approaches, and eventually reached industry-scale combined world and action models in 2025-2026. We provide an overview of all evolutionary stages. Before diving into

world models with deep learning techniques, we discuss the foundational concepts behind world models.

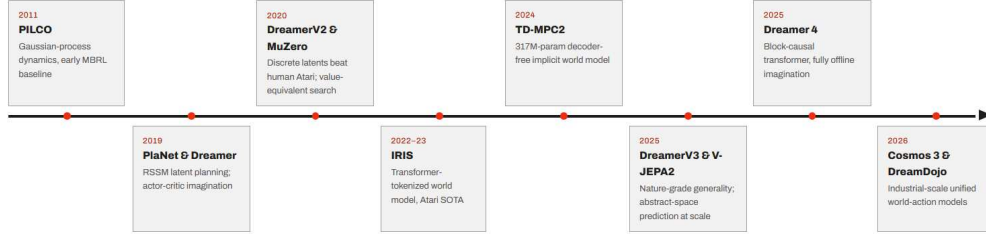


Figure 3 Chronological timeline of world model architectural development from PILCO (2011) to Cosmos 3 and DreamDojo (2026), highlighting key milestones including PlaNet and Dreamer, DreamerV2 and MuZero, IRIS, TD-MPC2, DreamerV3 and V-JEPA2, and Dreamer 4

3.1 Early Foundations: Gaussian Processes, Ensembles, and Neural ODEs

Deep learning has enhanced the sample efficiency and scalability of world models in MBRL (model-based RL), which were previously limited by PILCO (Deisenroth and Rasmussen 2011), which uses Gaussian processes and analytically propagated gradients. For example, Chua et al. (2018) demonstrates that ensemble-based deep dynamics models, coupled with trajectory sampling through the Cross Entropy Method, yield competitive results against model-free algorithms in continuous control benchmarks with significantly reduced samples required. Since then, the ensemble method has become the *de facto* solution for MBRL tasks using state-based dynamics models. Kurutach et al. (2018) also illustrates that ensemble world models enable policy optimization in model-based policy learning, achieving performance close to trust region policy optimization (TRPO), yet with up to ten times fewer environment steps. The neural ordinary differential equation (Chen et al. 2018) world model is formulated in continuous-time space, facilitating its direct use in irregularly sampled sensor data fusion for robotics applications. Recently proposed physics-informed neural networks (PINN) (Ramesh and Ravindran 2023) as prior world models introduce explicit physical principles to guide the learned dynamics model in a deterministic manner, showing superior results when the dynamics have intrinsic structure available. Also, planning based on gradients (S V et al. 2023) offers an alternative to sampling-based planners such as CEM, by optimizing action sequences directly through the differentiable dynamics model.

3.2 Latent Visual World Models: PlaNet and the Dreamer Series

PlaNet (Hafner et al. 2019b), the pioneer of visual MBRL, adopts an RSSM encoder that compresses the raw observation images into latent space. It utilizes the cross-entropy method (CEM)-based MPC algorithm, reaching the same efficiency level as model-free reinforcement learning (MFRL) agents in DMControl tasks. Later, DreamerV1 (Hafner et al. 2019a) adopts latent actor-critic algorithms to optimize end-to-end policies using backpropagation through imagined paths. In their Atari ablations, the authors report that categorical latents outperform Gaussian ones and conjecture that this is because they better represent multimodal or discontinuous dynamics (Hafner et al. 2020). DreamerV2 (Hafner et al. 2020) introduces discrete categorical latents with straight-through gradients and was the first world-model agent to reach human-level Atari performance.

DreamerV3 (Hafner et al. 2025a) contributed three robustness techniques: the symlog transformation of prediction targets, two-hot encoding of reward and return predictions, and KL balancing with free bits. Its publication in *Nature* helped establish WM-RL as a mainstream approach. Lastly, Dreamer 4 (Hafner et al. 2025b) is a transformative milestone that replaces the gated recurrent unit (GRU)-based online learning framework with the block causal transformers’ offline imagination mechanism.

A series of variants tackle unique challenges associated with Dreamers. For instance, DreamerPro (Deng et al. 2022) innovation lies in learning prototype-based representations rather than reconstruction-based features. DreamerPro achieves enhanced world model resilience against visual distractors, whereas traditional WM agents suffer severe degradation in performance when reconstructing images. Another fascinating work includes Plan2Explore (Sekar et al. 2020). They implement the RSSM architecture to drive efficient exploration without relying on task-specific reward signals. Their key intuition involves using the disagreement signal from an ensemble of the RSSM network as the exploration policy. Moreover, MuDreamer (Burchi and Timofte 2024) proposes integrating MuZero’s value equivalence objective function into Dreamer’s actor-critic policy update, revealing mutual complementarity. TransDreamer (Chen et al. 2022) addresses Dreamer’s shortcomings in modeling long-term dependency by transforming the inner state transition dynamics of the GRU module. Lastly, ContextWM (Wu et al. 2023) pre-trains the context conditioning world model with vast video data for general-purpose application in various tasks, anticipating future pretrain-first paradigms.

3.3 Transformer-Based World Models

In recent years, we have witnessed transformative progress in world models in reinforcement learning (WM-RL), exemplified by the transition from GRU-based RSSM architectures to transformer-based dynamics. IRIS (Micheli et al. 2022) achieved a human-normalized score of 1.046 on Atari 100k by employing the dVAE tokenizer combined with an autoregressive GPT transformer. The success of IRIS marked the predominance of discrete token formulations of WMs in games. The recently proposed

TWM (Robine et al. 2023) unified tokens from observations, actions, and rewards under one transformer sequence model, reaching state-of-the-art results on Atari 100k and demonstrating that the joint modeling of multimodal sequences subsumed traditional encoder/dynamics/decoder decomposition approaches. Δ -IRIS (Micheli et al. 2024) shortened the sequences by tokenizing only the differences between frames instead of entire frames themselves, which facilitated even deeper architectures within the same computational constraints. Additionally, the UniZero (Pu et al. 2024) disentangled state variables from context in a transformer-based world model architecture, solving the problem of context bleeding during Monte Carlo Tree Searches in long horizons.

Dreamer 4 (Hafner et al. 2025b) unified previous works in developing a fully parallelizable transformer architecture with block causal attention, known as the block causal transformer. The shortcut-forcing loss trained the dynamical component to directly regress latent states up to horizons $k \in \{2, 4, 8, \dots\}$, reducing cumulative errors from iterative rollouts and facilitating computable shortcut paths for real-time inference in a single GPU implementation. Further theoretical analysis revealed that the concepts of teacher-forcing, shortcut-forcing, and multistep losses all belong to the family of multi-chunk predictions, wherein parallel supervision over chunks improved the accuracy and efficiency of training simultaneously (Xu et al. 2026a).

EMERALD (Burchi and Timofte 2025a) introduced the combination of MaskGIT decoding along with Categorical Variational Autoencoder and transformer-based world models, achieving highly efficient decoding compared to traditional autoregressive approaches while retaining the discretization property of IRIS. Another notable work was Learning Transformer-Based World Models with CPC (Burchi and Timofte 2025b), which incorporates contrastive predictive coding in the transformer world model architecture for improved temporal representation quality.

3.4 Diffusion-Based World Models

In recent years, denoising diffusion models for visual world modeling have been shown to promote efficient policy training in reinforcement learning for game playing. Specifically, Alonso et al. (2024) argues that the rich perceptual content of denoising diffusion-generated images improves reinforcement learning signals. Recently, Rigter et al. (2023) proposed the diffusion world model, a novel framework combining state transition predictions into one diffusion architecture for imaginative rollout and model-based planning. However, inference cost is a significant limitation, since sampling a single frame typically requires many denoising steps. DIAMOND mitigates this by using a small number of steps per frame. DreamerAD (Yang et al. 2026d) reduces the cost further for autonomous driving: shortcut forcing cuts sampling from 100 diffusion steps to a single step, with an overall speedup of about $80\times$, enabling practical reinforcement learning in latent space. The work by Cheng et al. (2026) explores the problem of the mismatch between search and value learning in world model based reinforcement learning, where the improvement of policy is dependent upon value functions that are obtained from a different, independent search policy, resulting in inconsistent training. In order to achieve consistency between the two, search and policy optimization are combined in a diffusion policy representation known as MBDPO,

which helps in scalable policy learning in world models. To support efficient reinforcement learning in large-scale diffusion world models, [Amigo et al. \(2026\)](#) proposed a framework of coupled local/global world models based on a decoupled first-order gradient (FoG) method.

3.5 Decoder-Free and Abstract-Space Prediction

A growing body of work supports the view that pixel reconstruction diverts world-model capacity toward task-irrelevant visual details. First, in experiments with TD-MPC2 ([Hansen et al. 2024](#)), good generalization can be achieved without the decoder component. Second, MuDreamer ([Burchi and Timofte 2024](#)) and DreamerPro ([Deng et al. 2022](#)) both outperform the reconstruction method Dreamer by eliminating or sidestepping pixel decoders. Third, NE-Dreamer ([Bredis et al. 2026](#)) substitutes reconstruction with encoder embedding predictions through causal temporal transformers. Through ablation analyses, it was discovered that the observed gains stem from temporal causality modeling instead of omitting the decoder alone. Additionally, the Denoised MDP framework ([Wang et al. 2023](#)) demonstrates that decomposing world models into reward-relevant and reward-irrelevant factors during learning can lead to better performance compared to the conventional reconstruction-based paradigm.

Among these results, V-JEPA 2 ([Assran et al. 2025](#)) is particularly notable: a model with more than one billion parameters, pretrained on over one million hours of internet video and fine-tuned on fewer than 62 hours of unlabeled robot video, supports zero-shot robot planning in abstract space. Further disentanglement through background and action-conditioned modeling in DDP-WM ([Yin et al. 2026](#)) significantly reduces the likelihood of policy degradation in complicated vision tasks.

3.6 Object-Centric World Models

However, there are still some inherent disadvantages associated with traditional monolithic World Models, such as the inability to handle multi-object scenarios properly, since all interactions are embedded in global latent spaces rather than in discrete objects individually. Object-centric world models (OCWMs) solve this problem by incorporating slot-based perception and graph neural network (GNN)-based dynamics. Specifically, in [Feng et al. \(2026\)](#), the interaction graph among object slots is learned in an online manner with GNN to predict the multi-object manipulations via force contact propagation. In [Mosbach et al. \(2025\)](#), the idea of slot object-centric latent dynamics is applied with RSSM-like training, leading to substantial gains in relational control tasks. [Wu et al. \(2022\)](#) uses a Transformer-based framework to make compositional predictions for multi-object environments using object slots as inputs. Moreover, as mentioned in [Baek et al. \(2025\)](#), Dreamweaver derives hierarchical, object-based representations from raw video data by using a recurrent block-slot unit. The idea behind this is that scenes are broken down into objects and their properties, which allows for the creation of new future scenarios by combining the attributes of previously seen objects. Recently, object-centric world models for causality-aware RL [Nishimoto and](#)

Matsubara (2026) proposed OCWM with a causality-discovery mechanism to discover causally-related objects per sub-task during Reinforcement Learning.

3.7 State-Space Models and Efficiency

SSM-MBRL (Krinner et al. 2025) proposes a technique to accelerate MBRL training using a structured State Space Model instead of a GRU model. It achieves complete parallelization using scan operators, resulting in a 10x faster training time and a 4x faster MBRL training overall for agile quadrotor flights compared to the original implementation. Hieros (Mattes et al. 2024) utilizes Structured State Space sequence (SSS) models for hierarchical temporal abstraction in imagination processes. The experimental results show that hierarchical SSM world models outperform flat RSSM world models in solving tasks with long horizons. THICK WM (Gumbsch et al. 2024) presents a solution for implementing temporal abstraction in hierarchical world models with multiple planning timescales. Thus, both PlaNet-style and Dreamer-style agents can simultaneously conduct plans on different scales. Sparse imagination for efficient visual WM planning (Chun et al. 2025) suggests utilizing sparse imagination for efficient planning in WMs, achieving policy training in a 4-8 times lower computational cost than the baseline counterparts while maintaining equivalent policy performance.

3.8 Post-Training of World Models

RLVR-World (Wu et al. 2025) has established a sequential pipeline consisting of pre-training, RLVR post-training, and policy training with the help of post-trained WMs. NVIDIA’s Cosmos-Predict2.5 (Ali et al. 2025), which is based on training using 200 million hand-curated video clips followed by post-training optimization through reinforcement learning, shows the value of reinforcement learning post-training for world models, resulting in significant gains in video quality and task instruction alignment. The policy-driven world model adaptation (Chen et al. 2025) takes this idea further by jointly optimizing the policy with WM through Stackelberg learning, improving WM robustness against attacks in the environment.

4 Offline Model-Based RL

Learning a policy solely from static offline datasets becomes important when collecting fresh experiences is unsafe or costly. This setting, known as offline model-based reinforcement learning, is a special case of WM-RL. Mitigating the issue of offline WM distributional shift is critical since the learned world model could produce considerable errors outside the observed data distributions. Thus, an erroneous world model could result in exploiting a high-value area due to mistakes made by the model rather than due to actual value regions, causing potential problems during deployment. MOPO (Yu et al. 2020) provides the initial theoretical foundation for offline model-based reinforcement learning by introducing an uncertainty penalty estimated with an ensemble to limit the policy to operating within the offline data support. By merging model-based rollouts and Conservative Q-learning (Kumar et al. 2020), COMBO (Yu et al. 2021) improves offline MBRL efficiency, allowing efficient use of

offline data by combining Q-learning and model-based rollouts within the distribution range. MOREL (Kidambi et al. 2020) creates a pessimistic MDP by labeling all unknown areas (areas with high disagreement among the ensemble predictions) as absorbing states with minimal rewards to prevent movement outside the offline data distribution range. RAMBO-RL (Rigter et al. 2022) trains a world model adversarially in an effort to maximize Q-value depression in out-of-distribution areas, producing tighter pessimism bounds compared to uncertainty penalties. CoWorld (Wang et al. 2024a) solves the offline-to-online transfer problem in offline WM-RL, allowing offline-pretrained WMs to be efficiently adapted through online interaction. MOTO (Rafailov et al. 2023) shows evidence of efficient pretraining and fine-tuning pipelines for robotic learning, achieving fast adaptation with offline-pretrained WMs. The uncertainty-aware robotic world model (Li et al. 2025a) advances offline MBRL to robotic systems, marking the first offline WM demonstration in robotic hardware settings. It introduces uncertainty-penalization to offline long-horizon MBRL and explicitly considers observation uncertainties in robotic systems. The two most extreme offline WM examples include Dreamer 4 (Hafner et al. 2025b) and DreamDojo (Gao et al. 2026). Both systems train and evaluate policies entirely within the world model, with no interaction with the real environment (Section 2.4 - 2). This does not remove distribution shift: model errors can no longer be corrected online and must be controlled through the coverage and quality of the pretraining data (see OP4).

5 Exploration and Representation Learning

5.1 Model-Based Exploration

We analyze the contribution of world models to effective exploration strategies in model-based reinforcement learning. Two critical points exist here: The use of world models offers the ability to design intrinsic motivations based on model prediction uncertainty; Efficient planning-based exploration in the model space reduces the need to incur exploration costs in real environments. Specifically, Plan2Explore (Sekar et al. 2020) leverages ensembles to measure disagreement in model predictions of the recurrent state-space model, providing intrinsic rewards for exploring regions with maximum ensemble disagreement. Intuitively, these states require more information for prediction purposes. Thus, such explorations result in task agnostic WM learning, followed by efficient task learning with minimal rewards using just a few shots. In addition, LEXA (Mendonca et al. 2021) combines RSSM with a goal-conditioned policy and a self-supervised goal sampling algorithm, resulting in efficient open-ended exploration and task learning via retrospective goal-labeling with minimal shots. The latest dreaming of many worlds framework (Prasanna et al. 2024) proposes a state-of-the-art solution that addresses one-shot learning challenges. The key component is the augmentation of the RSSM framework with a context variable indicating the current environment. As a consequence, the framework learns to dream over a wide variety of simulated environments. The ability to simulate multiple environments during training makes this framework suitable for addressing zero-shot generalization issues. Adaptive WM (Gospodinov et al. 2024) solves the non-stationarity challenge of real

environments. It maintains a parameter distribution over WMs and dynamically updates the distribution while keeping track of expectations in the policy optimization objective function.

5.2 Visual Representation Robustness

Visual WMs that learn using a reconstruction loss exhibit high sensitivity to visual distractions such as moving background scenes, changing lighting, or extra objects. The reason is that the reconstruction loss forces the network to expend resources learning unnecessary details related to visual reconstruction instead of concentrating on modeling task dynamics. To mitigate this challenge, [Deng et al. \(2022\)](#) proposed DreamerPro, which predicts cluster assignments from prototype representations learned in the RSSM recurrent state space. By doing so, DreamerPro reduces the capacity needed for visual reconstruction, leading to robustness against visual distractions. Contrastive unsupervised representations for reinforcement learning (CURL), which learns visual representations via contrastive learning for downstream reinforcement learning policies, outperforms traditional reconstruction-based techniques in producing effective representations ([Laskin et al. 2020](#)). [Wang et al. \(2023\)](#) introduced Denoised MDP, in which the authors factor the visual world model into controllable (dynamics impacted by actions) and uncontrollable (background) components and train the two streams separately, using only the former for policy training. Recently, DDP-WM ([Yin et al. 2026](#)) extended the above approach to disentangled dynamics prediction by distinguishing the intrinsic object dynamics from the exogenous scene. Notably, DDP-WM significantly improves policy performance in distracting visual environments compared to existing works.

In order to solve the aforementioned information conflict problems in distracting visual model-based RL, the recently proposed MInCo ([Sun et al. 2025](#)) introduces mutual information constraints, preventing distractors from being captured by the encoder in the world model. Thus, the method regularizes the learning objective by forcing the learned latent representation to focus exclusively on the task at hand rather than on any visual distractions. In addition, [Sun et al. \(2024\)](#) utilizes a hybrid recurrent state-space model (HRSSM) for a combination of spatiotemporal masking and the bisimulation concept along with latent reconstruction. This technique is designed to develop latent representations for filtering exogenous visual noise and achieving robust world model learning.

6 Hierarchical and Temporal Abstraction

The scalability issue of flat world models on long-horizon sparse-reward tasks remains unresolved. Flat world models struggle because the number of imagination steps needed for future state simulation increases linearly with the task horizon, making long-horizon reward learning prohibitively expensive, both computationally and in terms of data. Hierarchical WMs serve as a potential solution by parallelizing tasks at different time scales simultaneously.

THICK World Models, introduced by Thick WM ([Gumbsch et al. 2024](#)) at ICLR 2024, incorporate temporal abstraction within the world model architecture directly.

Specifically, the architecture consists of a high-level WM predicting the state outcome after k primitive steps ("thick" predictions) and a low-level WM making predictions one step at a time. As a result, model predictive control and Dreamer-type agents become capable of reasoning about future states across multiple time horizons efficiently without expanding the search tree exponentially. Another notable framework called Hieros (Mattes et al. 2024) utilizes SSS world models for abstraction during imagination, demonstrating significantly improved performance for hierarchical SSM-based world models compared to standard RSSM-based world models when dealing with multistep planning.

A specific instantiation of H-WM (Huang et al. 2026a) is designed for robotic applications involving task-motion planning. The architecture includes a semantic task-level WM predicting the accomplishment of sub-goals at a distance ranging from 10 to 100 primitive actions, while a low-level WM is responsible for prediction at every primitive step. Communication across different levels occurs via latent variables of the subgoal representation. In the SkiMo architecture (Shi et al. 2022), hierarchical world models take advantage of learned abstractions referred to as "skills", representing temporally prolonged actions. As a consequence, high-level planning can be conducted over sequences of skills instead of individual primitives. Additionally, HAIC (Li et al. 2026a), MaskWAM (Yu et al. 2026a), RepWAM (Wang et al. 2026a), and WEAVER (Jain et al. 2026) further extend the World Action Model (WAM) paradigm in the domain of manipulation and architectural specialization (expanded on in the following sections).

7 Robotic Manipulation

After presenting the architecture and types of world models, let us examine their applications one-by-one. As illustrated in Fig. 4, there are seven applications, namely, robotic manipulation, navigation & locomotion, autonomous driving, multi-agent systems, LLM agents, games, and healthcare & scientific discovery. Let us start with robotic manipulation.

7.1 The Data Bottleneck and Human Video Pretraining

DreamDojo (Gao et al. 2026) provides a substantial demonstration to tackle the data manipulation bottleneck issue. With 44,000 hours of egocentric human videos (the largest visual-world pretraining dataset in terms of storage size), sufficient physical priors can be obtained for robot manipulation through latent action supervision. In addition, V-JEPA2 (Assran et al. 2025) realizes zero-shot pick-and-place manipulation with merely 62 hours of unlabeled robot data. Furthermore, the World Models Can Leverage Human Videos for Dexterous Manipulation (Goswami et al. 2026) provides systematic evidence to prove that bridging the embodiment gap between human hands and robot grippers is feasible in a variety of contact-intensive tasks, as long as the same physical principles hold.

Structured World Models From Human Videos (Mendonca et al. 2023) provides early empirical evidence for the hypothesis. A structured world model trained on human manipulation videos can learn physically interpretable latent dynamics that are

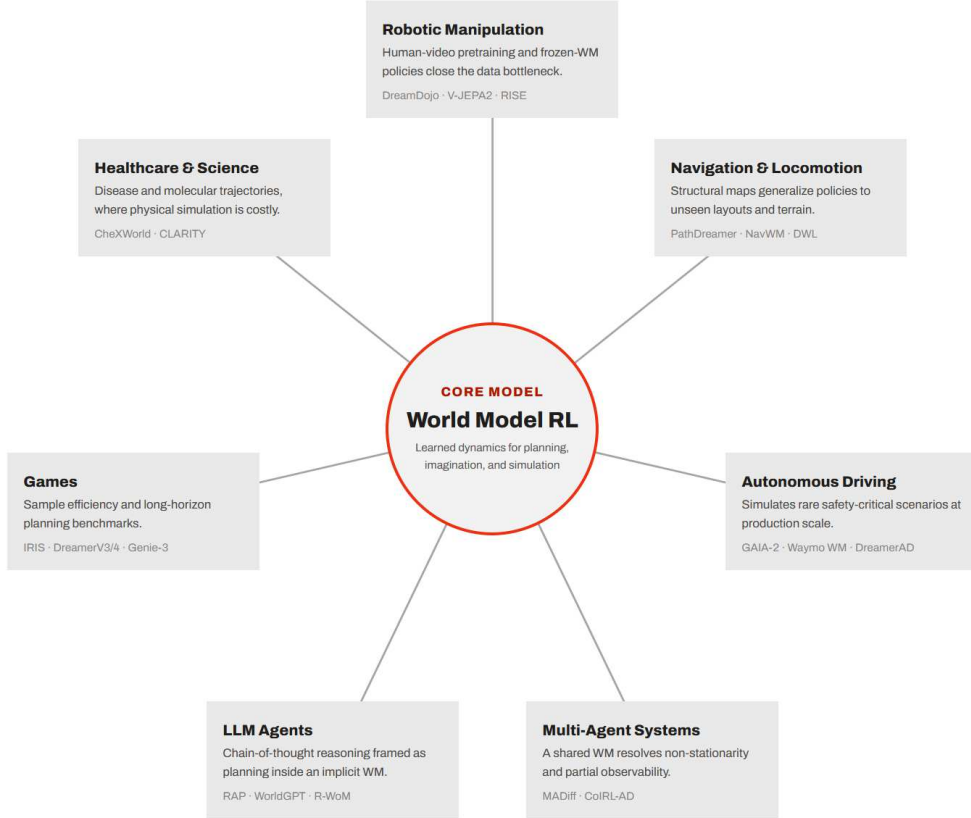


Figure 4 Application domains of world model reinforcement learning organized around a shared core model, spanning robotic manipulation, healthcare and scientific discovery, navigation and locomotion, autonomous driving, multi-agent systems, LLM agents, and games, with representative architectures listed for each domain

transferred to robot arm control tasks. AdaWorld: Learning Adaptable World Models with Latent Actions (Gao et al. 2025) leverages latent action inference techniques to adapt a Pretrained world model efficiently to novel robot manipulations and embodiments. Moreover, in the World Action Models are Zero-Shot Policies paper (Ye et al. 2026a), it is demonstrated that under certain conditions, a powerful enough world action model can serve as a zero-shot policy without extra reinforcement learning training.

7.2 World Models as RL Environments

Firstly, World-Gymnast (Sharma et al. 2026) presents extensive empirical evidence demonstrating the advantage of workflows involving WMs compared to physics simulators in producing higher-quality learned policies for RL environments. In particular, World-Gymnast conducts experiments to evaluate and compare the performance of

the learned policy under three paradigms: (i) manipulation in WMs; (ii) manipulation in physics simulators; and (iii) manipulation using imitation learning with experts’ demonstrations. The experiment shows that the learned policy in frozen WMs significantly surpasses others in the benchmark test set. Secondly, DiWA (Chandra et al. 2025) develops the technique of freezing WMs and leveraging the pre-trained WMs’ potentials by applying offline adaptation. Specifically, DiWA employs an offline adapter that produces the manipulation policy using the pre-trained WMs’ diffusion properties. Thirdly, World4RL (Jiang et al. 2025b) continues to improve the efficiency of fine-tuning WMs. However, unlike DiWA, which relies on single-step evaluation, World4RL uses the diffusion process-based WM as the environment simulator for policy refinement. Lastly, VLA-RFT (Li et al. 2025b) describes a detailed protocol for performing value-guided learning with RL in the WM framework. According to the protocol, the task includes episodes composed of actions followed by the verification of successful task completion, thus awarding the appropriate reward. The policy subsequently undergoes the process of RL training without having any direct contact with the task execution in reality.

Firstly, RISE (Yang et al. 2026a) introduces a compositional architecture for WMs that allows for developing policies that progressively improve themselves. Specifically, RISE composes novel scene WMs with the components learned for simpler scene compositions. The process involves collecting experiences by acting according to the learned policy, updating the components of the WM accordingly, and iteratively enhancing the policy through self-reinforcement. This addresses the problem of inadequate generalization to out-of-distribution examples associated with monolithic WMs. Secondly, RoboScape-R (Tang et al. 2026) estimates both the next state and the rewards of an action. As a result, it constitutes a full-fledged RL training environment solely implemented within the WM, obviating any need to interact with the real environment during the training process. Notably, a promising recent development is the comprehensive framework called MOTO (Rafailov et al. 2023). MOTO provides a protocol to realize the workflow from offline training of the RL policies within WMs to the online training phase. Another critical contribution is the finetuning of offline World Models in the Real World study (Feng et al. 2023).

7.3 3D and Geometric World Models

Standard world models of video data adopt 2D pixel-based representations. Thus, their limitation lies in handling tasks involving spatial 3D reasoning. In this part, we will review six recent works exploring different approaches to addressing this limitation. First, GWM (Lu et al. 2025) adopts the representation of 3D Gaussian splats for the world state and predicts their parameter transitions under actions, making the WM geometrically interpretable. Second, ContactGaussian-WM (Wang et al. 2026c) is the extended version of GWM with contact priors incorporated, achieving efficient manipulation of objects involving frequent contacts. Third, PointWorld (Huang et al. 2026b) extends the use of 3D point cloud-based WMs for broader manipulation in-the-wild scenarios. Compared to standard 2D video WMs, it exhibits stronger robustness against changes in viewpoints. Fourth, Particle-Former (Huang et al. 2025) deals with different kinds of materials simultaneously, namely rigid, deformable, and

granular objects, within one transformer WM based on particle representations. Fifth, 3DFlowAction (Zhi et al. 2025) utilizes the formulation of WM dynamics as 3D motion flows and facilitates cross-embodiment manipulation transfers. Finally, GAF (Chai et al. 2025) represents world dynamics as gaussian action fields, providing a continuous, differentiable WM for gradient-based manipulation planning.

7.4 Policy-World Model Unification for Manipulation

For the manipulation task class, the WAM paradigm involves significant interactions between prediction dynamics within the world model and action generator modules. As shown in Figure 3, VLAW (Guo et al. 2026) introduces a training approach wherein the policy generates the training episodes of the world model, whereas the world model provides imaginary episodes in great detail for the reinforcement learning of policies. Other research has extended this approach to online correction during policy evaluation. Here, the VLA policy proposes actions, sparsely imagining the outcome inside the world model; subsequently, the policy performs self-correction prior to executing actions. Notably, architectural advances have included architectures H-WM (Huang et al. 2026a), RISE (Yang et al. 2026a), HAIC (Li et al. 2026a), MaskWAM (Yu et al. 2026a), RepWAM (Wang et al. 2026a), and WEAVER frameworks (Jain et al. 2026).

8 Navigation and Locomotion

8.1 Visual Goal-Conditioned Navigation

Firstly, X-MOBILITY (Liu et al. 2024b) demonstrates the effectiveness of applying world models trained with a diversity of indoor navigation methods towards new environments that do not have prior access to the map information. Specifically, it achieves generalization to unseen floor plans and superior performance compared to reactive model-free navigation policies. Second, MUN (Duan et al. 2024) extends the previous findings to general goal-oriented navigation. In particular, it finds that a WM captures the structural information about the environment, such as wall geometry and corridor topology, which enables generalization to arbitrary goals without annotated waypoints. In addition to this, Shanks et al. (2025) has developed the concept of DreamerNav, an extended form of DreamerV3, with multimodal perception and curriculum learning in order to navigate quadruped robots in dynamically changing indoor environments. Additionally, they show that a policy learned solely through simulation works on two different quadruped platforms.

For indoor navigation, the embodied navigation foundation model (ENAV) (Zhang et al. 2025a), a recent state-of-the-art method for zero-shot indoor navigation, builds on top of foundation models. The key insight is using a pretrained foundation model for large-scale indoor navigation data to scaling zero-shot WM-based navigation via imagination-driven planning. Similarly, PathDreamer (Koh et al. 2021) applied the WM to indoor panoramic navigation early on, before recent efforts in foundation models. Specifically, among the first to apply the WM paradigm are panoramic image-based navigation tasks. Recently, NavWM (Bar et al. 2025) scales up video-based WM

pretraining for egocentric navigation by learning to predict future egocentric views with data from diverse outdoor navigation environments.

More advanced works involve navigating with a unified world-action model prediction. Specifically, WAM-Nav (Yang et al. 2026b) and NavWAM (Azuma et al. 2026) incorporate waypoint prediction in addition to world state prediction. In a study by Zanatta et al. (2026), a systematic analysis of the degradation or stability of the robustness of world models in environments of different variability is carried out using vision-based quadrotor navigation as a test platform. World models based on DreamerV3 are trained on different degrees of environmental randomness and are evaluated using cross-environment validation tests, which include not only self-supervised pretraining but also reinforcement learning (RL) fine-tuning. The authors show that the ability to generalize well during self-supervised pretraining is an effective predictor for sim-to-real transferability, since the world model with the highest cross-environment validation performance succeeded in transferring to a real quadrotor, while the model with the highest simulation-only performance in policy evaluation did not succeed in transferring.

8.2 Locomotion in Challenging Environments

Denoising world model learning (DWL) (Gu et al. 2024) provides experimental evidence demonstrating that world models learned using denoising achieve adequate accuracy for capturing contact dynamics across challenging terrains (e.g., staircases, inclinations, and deformations), surpassing existing RSSM-based world models on the task of locomotion policy learning across varying terrains. The state-space model-based MBRL (SSM-MBRL) formulation (Krinner et al. 2025) demonstrates that the state-space modeling approach achieves agile flight for quadrotors, attaining a tenfold improvement in the training efficiency of the world model without compromising flight/control policy performance. In addition, the world model-based perception (WMP) algorithm (Lai et al. 2025) illustrates how learning a world model using depth images alone is sufficient for achieving locomotion policies in contact-rich environments, such as the crawl and tilt terrains, where privileged learning algorithms using sparsely sampled terrain data fall short due to their inability to capture the geometrical details, through a one-stage process instead of the two-stage approach usually followed.

9 Autonomous Driving

9.1 The AV World Model Ecosystem

In particular, autonomous vehicles (AV) have become the application domain for which WM technology has seen its most accelerated progress, from academic exploration to practical industrial deployment. The key insight behind this technology is that rare safety-critical events could be imagined using the power of WM imagination, instead of being collected via actual fleets of vehicles, facilitating significant investment from industry players into AV WMs.

For instance, GAIA-1 (Hu et al. 2023) laid down a pioneering framework of generative AV WMs by conditionally leveraging the autoregressively modeled videos of ego-vehicles’ actions with the text descriptions. GAIA-2 (Russell et al. 2025) then expanded the applicability scope to cover multiple countries, namely the UK, USA, and Germany, ensuring multi-view spatiotemporally coherent scenarios using latent embeddings from proprietary driving models. DriveDreamer (Wang et al. 2024c) proposes a structured scenario construction approach that introduces high-definition maps and 3D box-level object annotations as conditional inputs to drive diffusion WMs for improved control and realism of generated driving scenarios. DriveDreamer-2 (Zhao et al. 2025) increases scenario diversity by introducing LLM-driven scenario descriptions during training. Drive-X (Shi et al. 2025a) presents an innovative idea of omni-scene modeling to equip AVs with generic world knowledge by training one unified WM backbone with various types of driving datasets collected in different sensor setups and geographical locations.

State-of-the-art innovations include Epona (Zhang et al. 2025b) that combines autoregressive modeling and diffusion sampling to achieve multi-minutes-long coherent video generation; HERMES (Zhou et al. 2025) that unifies scene understanding and 3D generation in one WM backbone; and UniFuture (Liang et al. 2026) that extends predictions to four-dimensional spaces (three dimensions and a temporal axis) for geometrically consistent future state estimation. The Waymo World Model (Jiang et al. 2026a) marks a production-stage deployment. It supports synchronized camera and LiDAR outputs, can convert dashcam footage into simulated scenarios, and can synthesize deliberately constructed edge cases.

Recent developments of AV WMs include Copilot4D (Zhang et al. 2024), which is showing at ICLR 2024 that discrete diffusion WMs are able to autonomously discover driving scene dynamics from LiDAR point clouds, thus alleviating the burdensome task of manual annotation in training datasets. ReconDreamer (Ni et al. 2024) tackles the problem of imperfect WM reconstruction quality by offering online restoration functions, making generated driving scenes closer to reality based on reference images. TrafficBots (Zhang et al. 2023) develops a special type of WMs dedicated to multi-agent traffic simulations.

9.2 RL and Planning Inside AV World Models

In Think2Drive (Li et al. 2024), latent space WM-RL is employed as an efficient technique for CARLA-v2, where RL training occurs solely within the latent space of the world model. On the NavSim v2 benchmark, DreamerAD (Yang et al. 2026d) resolves the latency issue during inference in the diffusion-based WM (about $80\times$ overall speedup due to shortcut forcing), along with introducing an autoregressive dense latent reward model, achieving state-of-the-art (SOTA) results at the time (87.7 Extended Predictive Driver Model Score (EPDMS)). Furthermore, Latent-WAM (Wang et al. 2026b) demonstrates that a small latent space world model architecture with 104M parameters and a spatial-aware compressive encoder achieves SOTA results on NavSim v2 (89.3 EPDMS) and Hugsim benchmarks. In WorldRFT (Yang et al. 2026c), the authors adopt reinforcement learning with verifiable rewards (RLVR)

after training for autonomous vehicle WM planning using predefined rule-based rewards (avoidance of collision, lane keeping, and passenger comfort). Additionally, in ResWorld (Zhang et al. 2026), ResWorld models future states by predicting residual transitions relative to the previous state rather than modeling the absolute future state transition to reduce cumulative errors in long-term predictions.

A set of research works focuses on solving the uncertainties in planning using latent world models. Learning Multiple Probabilistic Decisions from Latent World Model (Xiao et al. 2025) maintains a probabilistic distribution over future states in the WM; thus, planning aims to make decisions robust to such distributions. Adaptive WMs for Planning for Autonomous Vehicles (Vasudevan et al. 2025) retain several dynamic hypotheses about the environment while making decisions robust enough to account for uncertainties over hypotheses. AdaWM (Wang et al. 2025) online learns to adapt the WM dynamics under changing distributions.

9.3 Safety and Counterfactual Analysis

CounterScene (Jing et al. 2026) incorporates causal intervention methods in AV WM latent space to produce counterfactual trajectories ("what if braked 0.5 seconds earlier?") to ensure systematic testing for safety. VL-SAFE (Qu et al. 2025) converts language-based AV safety requirements into WM constraint predicates using a vision-language model. Dream to Drive (Nachkov et al. 2025) offers analytical AV WMs for a certain type of AV environment, allowing exact gradient computation for policy optimization rather than sample-based approximations. WorldVLMs (Englmeier et al. 2026) is an approach that leverages predictive WMs together with VLMs to provide a language interpretation of an AV’s decision-making process.

10 Multi-Agent World Models

Nonstationarity arises when agents’ strategies continually change throughout training. Partial observability occurs because none of the agents can observe the entire joint-state space of all agents simultaneously. Finally, the complexity of joint state-action spaces makes learning extremely difficult in multi-agent reinforcement learning (MARL) scenarios. World models offer a systematic way to overcome these three problems at once: a single shared WM could represent the complete joint-state space, simulate the opponents’ policies, and facilitate imagination-based cooperation without actual multi-agent interactions.

For example, Li et al. (2026b) proposed Puzzle it Out, published in AAMAS 2026. It presents a novel framework in which each local agent WM, learned from each agent’s own observation, is combined into a single global WM constrained by consistency conditions, effectively solving the problem of exponential joint state-action space growth. Experiments across 8 cooperative MARL environments significantly outperform 8 baseline algorithms. Another study, Grounded Answers (Liu et al. 2024c) from NeurIPS 2024, demonstrates that a generative world model can enable coordinated reasoning about multi-agent environments beyond individual WMs. Similarly, MADiff (Zhu et al. 2024) from NeurIPS applied diffusion sequence modeling along with joint state-action distribution modeling to perform multi-agent MBRL for coordinated

multi-robot manipulation based on imagination-based planning. Additionally, CoIRL-AD (Zheng et al. 2025) adopted competitive imitation reinforcement learning in a shared autonomous vehicle world model. Simultaneously training multiple AV agents within a shared WM yields robust defensive driving policies. Meanwhile, Multi-Task Interactive Robot Fleet Learning (Liu et al. 2024a) from CoRL employs visual world models trained using fleet-wide experiences. Agents could query the simulator’s experiences from other fleet members’ perspectives, making use of the overall fleet’s past experiences without centralized logging. Lastly, the Safe MARL (Jusup et al. 2024) paper extended ensemble uncertainty MBRL for multi-agent safe MARL.

11 LLM-as-World-Model for Agentic Tasks

11.1 Theoretical Grounding

A series of evidence demonstrates that large language models (LLMs) encapsulate significant amounts of procedural and causal knowledge encoded within parameters, constituting an intrinsic implicit world model. RAP (Hao et al. 2023) defines the LLM reasoning process as planning under the guidance of such a world model. It utilizes Monte Carlo Tree Search (MCTS) algorithms to simulate the transition of states by generating text output from LLMs. According to LLMs-Sim (Wang et al. 2024b), LLMs’ ability to simulate textual environments exhibits low accuracy in state transition predictions, particularly those involving arithmetic calculations, common sense, and scientific knowledge. Thus, LLMs fail to act as autonomous simulators at the moment, being prone to accumulating prediction mistakes in extensive rollout trials. The WorldGPT (Ge et al. 2024) experiment demonstrates that multimodal LLMs could serve as multimodal world models capable of state transition prediction with text-action pairs input in different areas.

11.2 Training LLM World Models

The RWML algorithm, developed by Yu et al. (2026b), directly trains LLMs to minimize the semantic gap between simulated and real-world states after one step ahead.

$$\mathcal{L}_{\text{RWML}} = \mathbb{E}_{(s,a,s') \sim \mathcal{D}} [d_{\phi}(g_{\theta}(s,a), \bar{s}')], \quad (6)$$

Here, the g_{θ} function denotes the large language model performing world modeling to predict the succeeding state, represented by $\bar{s}'$. On the other hand, the d_{ϕ} function indicates the distance measurement inside the pre-trained embedding space. It is noteworthy to highlight that this method does not involve any success reward signal for each task. Experimental results reveal improvements achieved by world model training, including 19.6 gains on the ALFWorld dataset and 6.9 gains on the τ^2 -Bench benchmark. The experiments prove the effectiveness of the explicit training strategy for enhancing language-model agent performance via world models. According to Mei et al. (2025a), grounding LLM-based world models using the retrieval-augmented generation technique reduces procedural hallucination issues during browser navigation. Ren et al. (2026) reports that fine-tuning the LLM world models based on structured

procedural knowledge via knowledgeable experience learning achieves approximately 15-point improvements in web navigation benchmarks.

11.3 MCTS Integration

Recent work by Shi et al. (2025b) shows that Monte Carlo Search, guided by state transition simulation by LLM, outperforms greedy chain-of-thought reasoning when applied in the context of games played through textual interactions. In addition, Path-Wise (Gungordu et al. 2026) employs LLM world model planners to assist in designing heuristics, allowing an agent to predict the effects of changes in its algorithms before implementing them.

11.4 Limitations and Complementary Niche

Despite recent progress, LLM world models face significant limitations. First, prediction quality degrades as the simulated horizon grows since errors compound over successive transitions (Wang et al. 2024b). Second, applying Monte Carlo Tree Search with LLM world models is computationally expensive because context length bounds the usable tree depth. Third, uncertainty calibration is inferior to that of probabilistic world models designed for this purpose. All these limitations indicate that LLM WMs fill a complementary role in comparison to neural WMs. Specifically, LLMs offer superior prior procedures for *digital environments*, including web browsing, the use of computer programs, and even executing scripts. Conversely, neural WMs excel at modeling the physical dynamics of visual controls. For instance, VDAWorld (O’Mahony et al. 2025) demonstrates successful hybridization, wherein the Vision-Language Model guides Neural WMs in capturing environmental abstractions for planning purposes.

12 Games and Interactive Environments

Game playing has always been considered the main playground for verifying the effectiveness of RL approaches. In WM-RL literature, this area continues to occupy an important position. Since 2019, the Atari 100k benchmark (Kaiser et al. 2020), consisting of 100,000 environment steps and around 2 hours of playtime, has set the bar for sample-efficient RL with pixel input. The SimPLe framework (Kaiser et al. 2020) marks the birth of Atari pixel-space WMs that improve sample efficiency by training a video prediction model and conducting rollouts using synthesized frames. In turn, IRIS (Micheli et al. 2022) achieves outstanding results on the Atari suite (human normalized score (HNS) 1.046; human-beating performance on 10 out of 26 games) by using a discrete token world model. Specifically, it uses dVAE for encoding visual observations and GPT-like transformers to model state transition dynamics. At the time, discrete autoregressive token WMs dominated Atari game benchmarks. TWM (Robine et al. 2023) integrates the representation of observations, actions, and rewards in one transformer, setting the state of the art (SOTA). Another notable approach, Δ -IRIS (Micheli et al. 2024), reduces the number of required tokens by predicting frame changes instead of whole frames. Meanwhile, DIAMOND (Alonso

et al. 2024) demonstrates that employing a diffusion WM helps generate perceptually better-quality images for policy improvement. Finally, EMERALD (Burchi and Timofte 2025a) combines discrete tokenization with MaskGIT-style parallel decoding, which speeds up inference relative to autoregressive token world models while retaining the advantages of IRIS-like discrete representations. The Minecraft gaming environment emerges as the preferred choice for testing agents’ long-horizon planning ability. Specifically, DreamerV3 (Hafner et al. 2025a) collects diamonds in a new Minecraft level using pure online RL from scratch. Furthermore, Dreamer4 (Hafner et al. 2025b) collects diamonds in Minecraft using video data collected offline with a success rate of 0.7%. It is the first time that an agent has achieved such results without interacting with the environment at all. The comparison baseline for this problem setting is VPT (Baker et al. 2022). Genie 3 (Hassabis et al. 2025) and SIMA-2 (Bolton et al. 2025) go beyond the fixed environments of existing benchmarks. Genie 3 generates diverse interactive 3D environments from text prompts in real time. SIMA-2 is trained and evaluated on commercial video games and on new game scenarios generated by Genie 3, using an “Infinite Training Loop” for continual improvement. Lastly, Oasis (Decart et al. 2024) offers an entire game engine built using a fully neural architecture. Namely, a Transformer network sequentially generates Minecraft games according to keyboard inputs without relying on any game engine logic code.

13 World Models in Healthcare and Scientific Discovery

A world model paradigm is suitable for many applications where conducting simulations using the principles of physics is computationally prohibitive and learning world models to assist predictions becomes necessary. Applications ranging from health care management to scientific discoveries include disease progression with different treatment policies, molecular dynamics under chemical intervention, and plasma control under actuator input. These problems belong to the family of dynamical systems that require computationally expensive simulation tasks.

The Medical World Model (Yang et al. 2025) proposes personalized tumor evolution states under different treatment policies through WMs. Yue et al. (2025) proposes CheXWorld, a self-supervised world-modeling approach for radiographs that learns representations of local anatomical structures, global anatomical layout, and domain variations, thereby improving downstream medical image classification and segmentation. Note that this is a representation-learning method rather than a sequential decision-making model. The CLARITY (Ding et al. 2025) framework proposes modeling the dynamics of context-aware disease state evolution as WMs for general-purpose clinical treatment policies. For clinical AI applications, Beyond Generative AI: WMs for Clinical Prediction, Counterfactuals, and Planning (Qazi et al. 2025) is the latest comprehensive survey on the topic of clinical decisions based on world models. Specifically, latent world models predict future physiological states, evaluate counterfactual treatment plans, and apply constrained reinforcement learning to learn safe clinical policies. A notable example of world models applied to fusion control is Char et al. (2023), which trains an offline policy inside a learned ensemble dynamics model

to track plasma quantities such as βN , ion rotation, and electron density on the DIII-D tokamak, a task previously reliant on physics-based simulators and rule-based controllers.

Within the realm of scientific discoveries, [Jing et al. \(2024\)](#) proposes modeling molecular dynamics trajectories using generative models and WMs for drug discovery tasks. All-Atom Diffusion Transformers ([Joshi et al. 2025](#)) further develop the idea into a broader field. By employing the generative model paradigm to predict future atomic configuration states, All-Atom Diffusion Transformers leverage the power of diffusion transformers to guide RL agents for property discovery. Deep Generative AI for Drug Discovery ([Kang et al. 2025](#)) introduces a framework for self-improving drug discovery using molecular dynamics prediction based on WMs for multi-target molecules.

14 World-Action Model Unification

In 2025–2026, a key structural development was the convergence of world models and policies into unified World Action Models (WAMs). By coupling dynamics prediction and action generation in a single model, a WAM can condition its actions on predicted futures and ground its predictions in the actions being taken.

NVIDIA Cosmos 3 ([Agarwal et al. 2026](#)) is an industrial-scale WAM that connects world simulation and action generation through a mixture-of-transformers architecture, pretrained on large multimodal datasets. World Action Models as Zero-Shot Policies ([Ye et al. 2026a](#)) claim that once a powerful WAM is created, it can serve directly as the final policy without RL fine-tuning. WoVR ([Jiang et al. 2026b](#)) proposes a framework for post-training reinforcement learning for vision-language action policies in world simulator environments. VLA (W) ([Guo et al. 2026](#)) explores iterative improvement: WAM produces simulation data for policy improvement; the refined policy utilizes more imagination through an updated WAM. HAIC ([Li et al. 2026a](#)) applies world action models for modeling full-body dynamics in humanoids interacting with objects. Whole-BodyVLA ([Jiang et al. 2025a](#)) proposes joint whole-body dynamics prediction, enabling unified loco-manipulation control with contacts. RISE’s Compositional World Model ([Yang et al. 2026a](#)) enables out-of-distribution scene generalization via compositional modeling of WAM modules.

For autonomous driving applications, [Wang et al. \(2026b\)](#)’s Latent-WAM demonstrates a highly compact 104M parameter WAM, reaching state-of-the-art autonomous vehicle planning through concurrent world state prediction and trajectory planning in a shared latent space representation.

15 Safety, Uncertainty, and Evaluation

15.1 Safe RL with World Models

Nightmare Dreamer ([Oseni et al. 2026](#)) implements safety through imagination. That is, the model uses imagination in the world model to simulate the trajectory outcomes that could lead to violations of safety constraints during policy training, preventing

Safety Gymnasium violations with 20x better sample efficiency than model-free approaches. Another promising direction is VL-SAFE (Qu et al. 2025), which allows natural language specifications for safety constraints by translating the natural language descriptions into formal world model constraints via Vision-Language Models. Safe MBRL with Ensemble Uncertainty (Yu et al. 2023) provides the theoretical foundation of guaranteeing safe MBRL using ensemble uncertainty and ensuring all constraints are satisfied.

15.2 Uncertainty Quantification

Optimistic world models (Mete et al. 2026) generalize the optimistic framework to deep WM architectures, providing theoretical guarantees on the regret for non-linear Markov decision processes. In addition, the paper "World Models That Know When They Don't Know" (Mei et al. 2025b) proposes using conformal prediction wrappers to provide calibrated uncertainty bounds without the need for expensive ensembles. On the other hand, Uncertainty-Aware Robotic World Models (Li et al. 2025a) incorporate an explicit modeling of state observation uncertainties in offline model-based reinforcement learning with real robots, allowing robust offline policy learning under noisy sensory data. Finally, Safe MARL (Jusup et al. 2024) with Ensemble Model-Based RL generalizes uncertainty-based safety to multi-agent systems.

15.3 Evaluation and Benchmarks

During the period from 2024 to 2026, several competing evaluation frameworks for world models emerged without reaching any agreement on the proper benchmark design. Specifically, WorldScore (Duan et al. 2025) introduced a unifying evaluation suite covering 3D, 4D, I2V, and T2V world models with more than 3,000 examples (metric axes including controllability, image/video quality, and dynamic interaction). WorldBench (Upadhyay et al. 2026) designs diagnostic evaluations targeting the physical attributes of world models. WorldArena (Shang et al. 2026) focuses on RL-related criteria measuring closed-loop training performance, synthetic data quality, and policy evaluation accuracy. WBench (Ying et al. 2026) evaluates multiple turns of interactive video WMs' memory consistency and controllable action responses. In addition, WorldMark (Xu et al. 2026b) includes various controllable interaction modalities for video WMs. Memory and action controllability of WMs in long-term (>60 steps) sequential interactions were systematically studied by MIND (Ye et al. 2026b). Moreover, WorldSimBench (Qin et al. 2024) investigates whether video generation world models meet simulator requirements concerning physics consistency. ACT-Bench (Arai et al. 2024) evaluates autonomous vehicle world models with regard to action-conditioned responses.

Methodologically speaking, the main problem is that the current benchmarks measure different qualities of world models without correlating to the actual downstream reinforcement learning policy performance.

16 Open Problems and Research Priorities

Despite substantial advancements discussed above, some fundamental issues still require further exploration. Based on the limitations observed in the reviewed literature, seven Open Problems (OP1-OP7) are summarized here as our identified future research directions. These include fundamental research problems related to the scaling and physical accuracy of world models (OP1-OP2); technical challenges of evaluation methodologies, cumulative errors, and limited action labels (OP3-OP5); and practical issues for safety assurance and multi-agent systems (OP6-OP7). We highlight that all these Open Problems are deeply interconnected. For instance, without a common evaluation framework (OP3), advancing the research of scaling law (OP1) will be challenging. Similarly, cumulative errors during imagination-based policy training (OP4) directly limit the provision of safety assurances (OP6).

Open Problem 1: Scaling Law of World Model - Currently, there is no established framework for efficiently allocating computing resources for parameters related to the world model, pre-training data volume, and RL training costs. Despite positive indications on scaling effects in terms of TD-MPC2, V-JEPA2, and [Pearce et al. \(2024\)](#) for pre-trained agents and world models, specific scaling laws for world models utilized in RL have yet to receive thorough investigation. Such research will lead to better allocation of resources amid the fast-growing field of WM-RL research. Further, it is unclear whether such laws hold for RL-relevant world models and downstream returns.

Open Problem 2: Physics Accuracy vs Visual Accuracy - WorldBench ([Upadhyay et al. 2026](#)) identifies consistent WM breakdown failures on the topics of contact mechanics, multi-agent interaction, fluid dynamics, and soft body simulations. Although techniques like contact-gaussian-WM and particle-formers address some narrow concerns, an integrated solution embedding physics prior information as a structured induction bias inside neural WMs, rather than relying on external oracles, remains elusive. It is crucial for safety-sensitive manipulations and self-driving vehicle applications in which precision in physics becomes a critical criterion during deployment.

Open Problem 3: Evaluation Unification - Eight competing benchmarks currently exist in WM research: WorldScore, WorldBench, WorldArena, WBench, WorldMark, MIND, WorldSimBench, and ACT-Bench. No unified evaluation protocol exists to gauge WM’s capability to predict RL policy efficiency. At the community-wide level, unification is analogous to BIG-bench’s ([Srivastava et al. 2022](#)) efforts in evaluating LLMs. This is the foremost research agenda with immediate applicability.

Open Problem 4: Accumulation of Error and Distribution Shift - Predictions from WM usually suffer compounding errors when the imagination horizon expands. Methods such as shortcut forcing, MBDPO, and curriculum training via bootstrapping can partially mitigate accumulation. Yet, it poses an inherent problem to be solved. More seriously, distribution shifts occur between the offline WM training corpus and online deployment when it comes to "Sec.2.4" paradigms. The offline-to-online transfer of WMs without suffering catastrophic forgetting, alongside the calibrated uncertainties in the presence of distribution shifts, remains an open problem.

Open Problem 5: Label Scarcity in Action Conditioning - Recent research shows that Dreamer 4 and DreamDojo learn policies from minimal, sparse action labels

during offline pre-training. Yet, there exists no clear characterization regarding the minimal requirements, the impact of label quality, and the scalability towards large embodiment discrepancies. A theoretical study on minimal-label action conditioning would bring practical implications for embodied AI deployments.

Open Problem 6: Safety Certification for WM-Guided Policies - Nightmare Dreamer (Oseni et al. 2026) enables safe exploration via imagination simulation; nevertheless, rigorous verification of the deployed WM-RL policy remains theoretically challenging and impractical due to combinatorial explosions. Safety certification via WM guided RL, combining conformal prediction bounds and constrained policy optimization, serves a pivotal role in surgeries, autonomous driving, and manufacturing automation.

Open Problem 7: Multi-Agent World Models Scalability - Puzzle it Out (Li et al. 2026b) presents scalable offline multi-agent reinforcement learning with up to five agents. However, principled multi-agent world model architectures with convergence guaranties, accounting for the nonstationary problem induced by simultaneous policy optimizations and exponentially large joint state-action spaces, are yet to be established.

17 Key Findings

Our synthesis of the 2019–2026 literature yields the following key findings:

- **From online to pretrain-first training.** The current mainstream framework has increasingly moved away from the online learning of world models and policies and has adopted the approach of pre-training world models with massive datasets and training policies exclusively through imagination (Dreamer/DreamDojo).
- **Abstract and decoder-free prediction is favored.** Across TD-MPC2, DreamerPro, MuDreamer, NE-Dreamer, and V-JEPA2, models that avoid pixel reconstruction match or outperform reconstruction-based ones, particularly under visual distractors. Ablations suggest that the gains stem from temporal causality modeling and not merely from removing the decoder.
- **Human video alleviates the data bottleneck.** Pre-training using extensive datasets containing tens of thousands of hours of egocentric human videos (DreamDojo) and internet-scale videos (V-JEPA2) allows for effective manipulation using limited robotic data.
- **World models and policies are converging.** World action models (WAM) involve action generation and dynamics prediction. If a world action model possesses sufficient robustness, then it can be applicable as a zero-shot policy, and world models can gradually complement or even replace traditional physics simulation environments.
- **Autonomous driving is the most mature application.** World models have moved from research prototypes to production pipelines (Waymo World Model, GAIA-2), mainly for synthesizing rare safety-critical scenarios.
- **LLMs and neural world models are complementary.** World models based on LLMs demonstrate promising prior knowledge in digital worlds; however, issues

arise in terms of degraded accuracy at long horizons. Traditional neural network-based world models surpass LLM-based models in modeling visual and physical phenomena; meanwhile, hybrid models are emerging in this direction.

- **Evaluation lags behind modeling.** Several benchmarks coexist; unfortunately, none of them verify any meaningful predictive validity concerning downstream RL policies.
- **Core challenges remain open.** Open problems include cumulative errors, distributional differences between offline training and online deployment, physics consistency, safety certification, and scaling laws.

18 Conclusion

From 2019 to 2026, there were paradigmatic transformations in WM-RL research, transitioning from the use of world models to improve the sample efficiency of visual control tasks to becoming a critical component of AI infrastructure for the physical world. These changes include three paradigm shifts corresponding to those mentioned in the Abstract section. First, the joint online training paradigm of the world model and policy in reinforcement learning transitions to a pretraining paradigm, where the world model is pre-trained on large scale datasets while the policy is only trained via imagination (Hafner et al. 2025b; Gao et al. 2026). Consequently, this eliminates the requirement for online interactions and allows scaling up training on Internet-scale datasets (Assran et al. 2025). Secondly, world models and policies are increasingly integrated as World Action Models (WAMs), sometimes functioning as zero-shot policies themselves (Ye et al. 2026a). Finally, algorithmic research is shifting towards creating infrastructures where the same algorithms are widely used. World foundation models (Agarwal et al. 2025; Ali et al. 2025; Agarwal et al. 2026) became advanced technology for several important applications, including autonomous driving (Jiang et al. 2026a; Russell et al. 2025), robotics (Gao et al. 2026), and healthcare (Yang et al. 2025).

Architectural research follows parallel transformations to meet the demands of WM-RL infrastructure. In addition to recurrent state space models (RSSM) with GRU-based dynamics (Hafner et al. 2019b), various architectures have emerged, including autoregressive transformers (Micheli et al. 2022), block causal transformers (Hafner et al. 2025b), decoder-less implicit models (Hansen et al. 2024), abstract-space predictors (Assran et al. 2025), and diffusion dynamics models (Alonso et al. 2024; Cheng et al. 2026). Despite significant differences in architectural designs, empirical results indicate that abstraction or implicitness performs better than pixel-by-pixel reconstruction when dealing with distracting information.

For future directions, we propose seven open problems (Sec. 16), which set forth an agenda after 2026. Firstly, scaling laws and physical consistency are highly interdependent because physical consistency imposes limits on the scalability of world models, and vice versa. These problems will likely require breakthrough research on novel theories. Secondly, evaluation unification provides a promising avenue for progress in WM-RL, but substantial collaboration within the community is needed to achieve this goal. Addressing open problems requires bridging the growing capability of models and the development of evaluation methodology, error control, and assurance technologies.

Statements and Declarations

Funding

No funding was received to assist with the preparation of this manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Ethics Approval and Consent to Participate

Not applicable. This work is a literature survey and did not involve human participants, animals, or primary data collection.

Consent for Publication

Not applicable.

Data Availability

Data sharing is not applicable to this article as no new data were generated or analysed during the current study. All sources discussed are cited in the reference list.

Materials Availability

Not applicable.

Code Availability

Not applicable. No original code was developed for this manuscript.

Author Contribution

W.D. and A.R. contributed equally to this work. Both authors conducted the literature review, wrote the manuscript, and read and approved the final version.

References

- Agarwal N, Ali A, Bala M, et al (2025) Cosmos world foundation model platform for physical ai. arXiv preprint arXiv:250103575 URL <https://arxiv.org/abs/2501.03575>
- Agarwal N, Ali A, Allen J, et al (2026) Cosmos 3: Omnimodal world models for physical ai. arXiv preprint arXiv:260602800 URL <https://arxiv.org/abs/2606.02800>
- Ali A, Bai J, Bala M, et al (2025) World simulation with video foundation models for physical ai. arXiv preprint arXiv:251100062 URL <https://arxiv.org/abs/2511.00062>
- Alonso E, Jelley A, Micheli V, et al (2024) Diffusion for world modeling: Visual details matter in atari. *Advances in Neural Information Processing Systems* 37:58757–58791
- Amigo J, Khorrambakht R, Mansard N, et al (2026) Coupled local and global world models for efficient first order rl. arXiv preprint arXiv:260206219 URL <https://arxiv.org/abs/2602.06219>
- Arai H, Ishihara K, Takahashi T, et al (2024) ACT-Bench: Towards action controllable world models for autonomous driving. arXiv preprint arXiv:241205337 URL <https://arxiv.org/abs/2412.05337>
- Assran M, Bardes A, Fan D, et al (2025) V-jepa 2: Self-supervised video models enable understanding, prediction and planning. URL <https://arxiv.org/abs/2506.09985>

- Azuma D, Miyanishi T, Sakamoto K, et al (2026) Navwam: A navigation world action model for goal-conditioned visual navigation. arXiv preprint arXiv:260613494 URL <https://arxiv.org/abs/2606.13494>
- Baek J, Wu YF, Singh G, et al (2025) Dreamweaver: Learning compositional world models from pixels. In: International Conference on Learning Representations (ICLR), URL <https://openreview.net/forum?id=e5mTvjXG9u>
- Baker B, Akkaya I, Zhokov P, et al (2022) Video PreTraining (VPT): Learning to act by watching unlabeled online videos. In: Advances in Neural Information Processing Systems (NeurIPS)
- Bar A, Zhou G, Tran D, et al (2025) Navigation world models. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 15791–15801
- Bardes A, Garrido Q, Ponce J, et al (2024) Revisiting feature prediction for learning visual representations from video. arXiv preprint arXiv:240408471 URL <https://arxiv.org/abs/2404.08471>
- Bolton A, Lerchner A, Cordell A, et al (2025) Sima 2: A generalist embodied agent for virtual worlds. arXiv preprint arXiv:251204797 URL <https://arxiv.org/abs/2512.04797>
- Bredis G, Balagansky N, Gavrilov D, et al (2026) Next embedding prediction makes world models stronger. arXiv preprint arXiv:260302765 URL <https://arxiv.org/abs/2603.02765>
- Burchi M, Timofte R (2024) Mudreamer: Learning predictive world models without reconstruction. arXiv preprint arXiv:240515083 URL <https://arxiv.org/abs/2405.15083>
- Burchi M, Timofte R (2025a) Accurate and efficient world modeling with masked latent transformers. In: International conference on machine learning, PMLR, pp 5894–5912
- Burchi M, Timofte R (2025b) Learning transformer-based world models with contrastive predictive coding. arXiv preprint arXiv:250304416 URL <https://arxiv.org/abs/2503.04416>
- Chai Y, Deng L, Shao R, et al (2025) GAF: Gaussian action field as a 4D representation for dynamic world modeling in robotic manipulation. arXiv preprint arXiv:250614135 URL <https://arxiv.org/abs/2506.14135>
- Chandra AL, Nematollahi I, Huang C, et al (2025) DiWA: Diffusion policy adaptation with world models. In: Conference on Robot Learning (CoRL), PMLR, pp 3378–3400

- Char I, Abbate J, Bardoczi L, et al (2023) Offline model-based reinforcement learning for tokamak control. In: Learning for Dynamics and Control Conference, PMLR, pp 1357–1372
- Chen C, Wu YF, Yoon J, et al (2022) Transdreamer: Reinforcement learning with transformer world models. arXiv preprint arXiv:220209481 URL <https://arxiv.org/abs/2202.09481>
- Chen J, Xu L, Venugopal A, et al (2025) Policy-driven world model adaptation for robust offline model-based reinforcement learning. arXiv preprint arXiv:250513709
- Chen RT, Rubanova Y, Bettencourt J, et al (2018) Neural ordinary differential equations. Advances in neural information processing systems 31:6572–6583
- Cheng X, Yuan W, Mu Z, et al (2026) Scaling world-model reinforcement learning through diffusion policy optimization. arXiv preprint arXiv:260526282 URL <https://arxiv.org/abs/2605.26282>
- Chua K, Calandra R, McAllister R, et al (2018) Deep reinforcement learning in a handful of trials using probabilistic dynamics models. Advances in neural information processing systems 31:4754–4765
- Chun J, Jeong Y, Kim T (2025) Sparse imagination for efficient visual world model planning. arXiv preprint arXiv:250601392 URL <https://arxiv.org/abs/2506.01392>
- Decart, Quevedo J, McIntyre Q, et al (2024) Oasis: A universe in a transformer. URL <https://oasis-model.github.io/>
- Deisenroth M, Rasmussen CE (2011) Pilco: A model-based and data-efficient approach to policy search. In: Proceedings of the 28th International Conference on machine learning (ICML-11). Omnipress, pp 465–472
- Deng F, Jang I, Ahn S (2022) Dreamerpro: Reconstruction-free model-based reinforcement learning with prototypical representations. In: International conference on machine learning, PMLR, pp 4956–4975
- Ding T, Zou Y, Chen C, et al (2025) CLARITY: Medical world model for guiding treatment decisions by modeling context-aware disease trajectories in latent space. arXiv preprint arXiv:251208029 URL <https://arxiv.org/abs/2512.08029>
- Duan H, Yu HX, Chen S, et al (2025) WorldScore: A unified evaluation benchmark for world generation. In: IEEE/CVF International Conference on Computer Vision (ICCV), pp 27713–27724, URL <https://arxiv.org/abs/2504.00983>
- Duan Y, Mao W, Zhu H (2024) Learning world models for unconstrained goal navigation. Advances in Neural Information Processing Systems 37:59236–59265

- Englmeier S, Winter K, Flohr FB (2026) Worldvlm: Combining world model forecasting and vision-language reasoning. arXiv preprint arXiv:260314497 URL <https://arxiv.org/abs/2603.14497>
- Feng F, Lippe P, Magliacane S (2026) Learning interactive world model for object-centric reinforcement learning. Advances in Neural Information Processing Systems 38:89827–89862
- Feng Y, Hansen N, Xiong Z, et al (2023) Finetuning offline world models in the real world. In: Proceedings of the 7th Conference on Robot Learning (CoRL), pp 425–445
- Gao S, Zhou S, Du Y, et al (2025) AdaWorld: Learning adaptable world models with latent actions. In: International Conference on Machine Learning (ICML)
- Gao S, Liang W, Zheng K, et al (2026) Dreamdojo: A generalist robot world model from large-scale human videos. arXiv preprint arXiv:260206949 URL <https://arxiv.org/abs/2602.06949>
- Ge Z, Huang H, Zhou M, et al (2024) WorldGPT: Empowering LLM as multi-modal world model. In: Proceedings of the 32nd ACM International Conference on Multimedia (ACM MM), pp 7346–7355
- Gospodinov E, Shaj V, Becker P, et al (2024) Adaptive world models: Learning behaviors by latent imagination under non-stationarity. arXiv preprint arXiv:241101342 URL <https://arxiv.org/abs/2411.01342>
- Goswami RG, Bar A, Fan D, et al (2026) World models for learning dexterous hand-object interactions from human videos. arXiv preprint arXiv:251213644 URL <https://arxiv.org/abs/2512.13644>
- Gu X, Wang YJ, Zhu X, et al (2024) Advancing humanoid locomotion: Mastering challenging terrains with denoising world model learning. arXiv preprint arXiv:240814472 URL <https://arxiv.org/abs/2408.14472>
- Gumbsch C, Sajid N, Martius G, et al (2024) Learning hierarchical world models with adaptive temporal abstractions from discrete latent dynamics (THICK). In: International Conference on Learning Representations (ICLR), URL <https://openreview.net/forum?id=TjCDNssXKU>
- Gungordu O, Xiong S, Fekri F (2026) Pathwise: Planning through world model for automated heuristic design via self-evolving llms. arXiv preprint arXiv:260120539 URL <https://arxiv.org/abs/2601.20539>
- Guo Y, Lee T, Shi LX, et al (2026) Vlaw: Iterative co-improvement of vision-language-action policy and world model. arXiv preprint arXiv:260212063 URL <https://arxiv.org/abs/2602.12063>

- Ha D, Schmidhuber J (2018) World models. arXiv preprint arXiv:1803.10122 URL <https://doi.org/10.48550/arXiv.1803.10122>
- Hafner D, Lillicrap T, Ba J, et al (2019a) Dream to control: Learning behaviors by latent imagination. arXiv preprint arXiv:1912.01603 URL <https://arxiv.org/abs/1912.01603>
- Hafner D, Lillicrap T, Fischer I, et al (2019b) Learning latent dynamics for planning from pixels. In: International conference on machine learning, PMLR, pp 2555–2565
- Hafner D, Lillicrap T, Norouzi M, et al (2020) Mastering atari with discrete world models. arXiv preprint arXiv:2010.02193 URL <https://arxiv.org/abs/2010.02193>
- Hafner D, Pasukonis J, Ba J, et al (2025a) Mastering diverse control tasks through world models. Nature 640(8059):647–653. URL <https://doi.org/10.1038/s41586-025-08744-2>
- Hafner D, Yan W, Lillicrap T (2025b) Training agents inside of scalable world models. arXiv preprint arXiv:2509.24527
- Hansen N, Wang X, Su H (2022) Temporal difference learning for model predictive control. arXiv preprint arXiv:2203.04955 URL <https://arxiv.org/abs/2203.04955>
- Hansen N, Su H, Wang X (2024) Td-mpc2: Scalable, robust world models for continuous control. In: International Conference on Learning Representations, pp 47376–47405
- Hao S, Gu Y, Ma H, et al (2023) Reasoning with language model is planning with world model (RAP). In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp 8154–8173, URL <https://doi.org/10.18653/v1/2023.emnlp-main.507>
- Hassabis D, Ghahramani Z, Hadsell R, et al (2025) Genie 3: A new frontier for world models. Google DeepMind Blog, research preview, August 2025
- Hu A, Russell L, Yeo H, et al (2023) GAIA-1: A generative world model for autonomous driving. arXiv preprint arXiv:2309.17080 URL <https://arxiv.org/abs/2309.17080>
- Huang J, Chen W, Li Z, et al (2026a) H-wm: Robotic task and motion planning guided by hierarchical world model. arXiv preprint arXiv:2602.11291 URL <https://arxiv.org/abs/2602.11291>
- Huang S, Chen Q, Zhang X, et al (2025) ParticleFormer: A 3D point cloud world model for multi-object, multi-material robotic manipulation. In: Conference on Robot Learning (CoRL), URL <https://arxiv.org/abs/2506.23126>

- Huang W, Chao YW, Mousavian A, et al (2026b) Pointworld: Scaling 3d world models for in-the-wild robotic manipulation. arXiv preprint arXiv:260103782 URL <https://arxiv.org/abs/2601.03782>
- Jain AK, Wu Y, Farebrother J, et al (2026) Weaver, better, faster, longer: An effective world model for robotic manipulation. arXiv preprint arXiv:260613672 URL <https://arxiv.org/abs/2606.13672>
- Janner M, Fu J, Zhang M, et al (2019) When to trust your model: Model-based policy optimization. In: Advances in Neural Information Processing Systems (NeurIPS), pp 12498–12509
- Jiang CM, Masotto X, Sun B (2026a) The waymo world model: A new frontier for autonomous driving simulation. Waypoint (Waymo Blog), URL <https://waymo.com/blog/2026/02/the-waymo-world-model-a-new-frontier-for-autonomous-driving-simulation/>
- Jiang H, Chen J, Bu Q, et al (2025a) Wholebodyvla: Towards unified latent vla for whole-body loco-manipulation control. arXiv preprint arXiv:251211047 URL <https://arxiv.org/abs/2512.11047>
- Jiang Z, Liu K, Qin Y, et al (2025b) World4RL: Diffusion world models for policy refinement with reinforcement learning for robotic manipulation. arXiv preprint arXiv:250919080 URL <https://arxiv.org/abs/2509.19080>
- Jiang Z, Zhou S, Jiang Y, et al (2026b) Wovr: World models as reliable simulators for post-training vla policies with rl. arXiv preprint arXiv:260213977 URL <https://arxiv.org/abs/2602.13977>
- Jing B, Stärk H, Jaakkola T, et al (2024) Generative modeling of molecular dynamics trajectories. In: Advances in Neural Information Processing Systems (NeurIPS)
- Jing B, Hao R, Zhou W, et al (2026) Counterscene: Counterfactual causal reasoning in generative world models for safety-critical closed-loop evaluation. arXiv preprint arXiv:260321104 URL <https://arxiv.org/abs/2603.21104>
- Joshi CK, Fu X, Liao YL, et al (2025) All-atom diffusion transformers: Unified generative modelling of molecules and materials. In: International Conference on Machine Learning (ICML), URL <https://arxiv.org/abs/2503.03965>
- Jusup M, Pásztor B, Janik T, et al (2024) Safe model-based multi-agent mean-field reinforcement learning. In: Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS), URL <https://arxiv.org/abs/2306.17052>
- Kaiser L, Babaeizadeh M, Milos P, et al (2020) Model-based reinforcement learning for Atari (SimPLe). In: International Conference on Learning Representations (ICLR),

URL <https://arxiv.org/abs/1903.00374>

- Kang SI, Shin JH, Wu BM, et al (2025) Deep generative ai for multi-target therapeutic design: Toward self-improving drug discovery framework. *International Journal of Molecular Sciences* 26(23):11443
- Kidambi R, Rajeswaran A, Netrapalli P, et al (2020) MOREL: Model-based offline reinforcement learning. In: *Advances in Neural Information Processing Systems (NeurIPS)*, pp 21810–21823
- Koh JY, Lee H, Yang Y, et al (2021) Pathdreamer: A world model for indoor navigation. In: *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp 14738–14748
- Krinner M, Aljalbout E, Romero A, et al (2025) Accelerating model-based reinforcement learning with state-space world models. *arXiv preprint arXiv:2502.20168* URL <https://arxiv.org/abs/2502.20168>
- Kumar A, Zhou A, Tucker G, et al (2020) Conservative Q-learning for offline reinforcement learning. In: *Advances in Neural Information Processing Systems (NeurIPS)*, pp 1179–1191
- Kurutach T, Clavera I, Duan Y, et al (2018) Model-ensemble trust-region policy optimization. *arXiv preprint arXiv:1802.10592* URL <https://arxiv.org/abs/1802.10592>
- Lai H, Cao J, Xu J, et al (2025) World model-based perception for visual legged locomotion. In: *2025 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp 11531–11537
- Laskin M, Srinivas A, Abbeel P (2020) CURL: Contrastive unsupervised representations for reinforcement learning. In: *Proceedings of the 37th International Conference on Machine Learning (ICML)*, PMLR, pp 5639–5650
- LeCun Y (2022) A path towards autonomous machine intelligence. *Open Review* 62(1):1–62
- Li C, Krause A, Hutter M (2025a) Uncertainty-aware robotic world model makes offline model-based reinforcement learning work on real robots. *arXiv preprint arXiv:2504.16680* URL <https://arxiv.org/abs/2504.16680>
- Li D, Chen X, Wu Q, et al (2026a) Haic: Humanoid agile object interaction control via dynamics-aware world model. *arXiv preprint arXiv:2602.11758* URL <https://arxiv.org/abs/2602.11758>
- Li H, Ding P, Suo R, et al (2025b) Vla-rft: Vision-language-action reinforcement fine-tuning with verified rewards in world simulators. *arXiv preprint arXiv:2510.00406*

URL <https://arxiv.org/abs/2510.00406>

- Li Q, Jia X, Wang S, et al (2024) Think2drive: Efficient reinforcement learning by thinking with latent world model for autonomous driving (in carla-v2). In: European conference on computer vision, Springer, pp 142–158
- Li S, Li X, Chen S, et al (2026b) Puzzle it out: Local-to-global world model for offline multi-agent reinforcement learning. In: Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), URL <https://arxiv.org/abs/2601.07463>
- Liang D, Zhang D, Zhou X, et al (2026) Unifuture: A 4d driving world model for future generation and perception. URL <https://arxiv.org/abs/2503.13587>
- Liu H, Zhang Y, Betala V, et al (2024a) Multi-task interactive robot fleet learning with visual world models. In: Conference on Robot Learning (CoRL), URL <https://arxiv.org/abs/2410.22689>
- Liu W, Zhao H, Li C, et al (2024b) X-MOBILITY: End-to-end generalizable navigation via world modeling. arXiv preprint arXiv:241017491 URL <https://arxiv.org/abs/2410.17491>
- Liu Z, Yang X, Sun S, et al (2024c) Grounded answers for multi-agent decision-making problem through generative world model. In: Advances in Neural Information Processing Systems (NeurIPS), pp 46622–46652, URL <https://arxiv.org/abs/2410.02664>
- Lu G, Jia B, Li P, et al (2025) GWM: Towards scalable gaussian world models for robotic manipulation. In: IEEE/CVF International Conference on Computer Vision (ICCV), pp 9263–9274
- Mattes P, Schlosser R, Herbrich R (2024) Hieros: Hierarchical imagination on structured state space sequence world models. In: Proceedings of the 41st International Conference on Machine Learning (ICML), pp 35079–35103, URL <https://arxiv.org/abs/2310.05167>
- Mei K, Guo J, Chang S, et al (2025a) R-wom: Retrieval-augmented world model for computer-use agents. arXiv preprint arXiv:251011892 URL <https://arxiv.org/abs/2510.11892>
- Mei Z, Yin T, Baker M, et al (2025b) World models that know when they don’t know: Controllable video generation with calibrated uncertainty. arXiv preprint arXiv:251205927 URL <https://arxiv.org/abs/2512.05927>
- Mendonca R, Rybkin O, Daniilidis K, et al (2021) Discovering and achieving goals via world models (LEXA). In: Advances in Neural Information Processing Systems (NeurIPS), pp 24379–24391

- Mendonca R, Bahl S, Pathak D (2023) Structured world models from human videos. In: Robotics: Science and Systems (RSS)
- Mete A, Sheikh SA, Lin TH, et al (2026) Optimistic world models: Efficient exploration in model-based deep reinforcement learning. arXiv preprint arXiv:260210044 URL <https://arxiv.org/abs/2602.10044>
- Micheli V, Alonso E, Fleuret F (2022) Transformers are sample-efficient world models. arXiv preprint arXiv:220900588 URL <https://arxiv.org/abs/2209.00588>
- Micheli V, Alonso E, Fleuret F (2024) Efficient world models with context-aware tokenization. In: International conference on machine learning, PMLR, pp 35623–35638
- Mosbach M, Ewertz JN, Villar-Corrales A, et al (2025) Sold: Slot object-centric latent dynamics models for relational manipulation learning from pixels. In: International conference on machine learning, PMLR, pp 44911–44935
- Nachkov A, Paudel DP, Zaech JN, et al (2025) Dream to drive: Model-based vehicle control using analytic world models. arXiv preprint arXiv:250210012 URL <https://arxiv.org/abs/2502.10012>
- Ni C, Zhao G, Wang X, et al (2024) ReconDreamer: Crafting world models for driving scene reconstruction via online restoration. arXiv preprint arXiv:241119548 URL <https://arxiv.org/abs/2411.19548>
- Nishimoto Y, Matsubara T (2026) Object-centric world models for causality-aware reinforcement learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp 24585–24593
- O’Mahony F, Cipolla R, Tewari A (2025) Vdaworld: World modelling via vlm-directed abstraction and simulation. arXiv preprint arXiv:251211061 URL <https://arxiv.org/abs/2512.11061>
- Oseni O, Wang S, Zhu J, et al (2026) Nightmare dreamer: Dreaming about unsafe states and planning ahead. arXiv preprint arXiv:260104686 URL <https://arxiv.org/abs/2601.04686>
- Pearce T, Rashid T, Bignell D, et al (2024) Scaling laws for pre-training agents and world models. arXiv preprint arXiv:241104434
- Prasanna S, Farid K, Rajan R, et al (2024) Dreaming of many worlds: Learning contextual world models aids zero-shot generalization. arXiv preprint arXiv:240310967 URL <https://arxiv.org/abs/2403.10967>
- Pu Y, Niu Y, Yang Z, et al (2024) Unizero: Generalized and efficient planning with scalable latent world models. arXiv preprint arXiv:240610667 URL <https://arxiv.org/abs/2406.10667>

[org/abs/2406.10667](https://arxiv.org/abs/2406.10667)

- Qazi MA, Nadeem M, Yaqub M (2025) Beyond generative ai: World models for clinical prediction, counterfactuals, and planning. arXiv preprint arXiv:251116333 URL <https://arxiv.org/abs/2511.16333>
- Qin Y, Shi Z, Yu J, et al (2024) Worldsimbench: Towards video generation models as world simulators. arXiv preprint arXiv:241018072 URL <https://arxiv.org/abs/2410.18072>
- Qu Y, Huang Z, Sheng Z, et al (2025) VL-SAFE: Vision-language guided safety-aware reinforcement learning with world models for autonomous driving. arXiv preprint arXiv:250516377 URL <https://arxiv.org/abs/2505.16377>
- Rafailov R, Hatch KB, Koley V, et al (2023) Moto: Offline pre-training to online fine-tuning for model-based robot learning. In: Conference on Robot Learning, PMLR, pp 3654–3671
- Ramesh A, Ravindran B (2023) Physics-informed model-based reinforcement learning. In: Learning for Dynamics and Control Conference, PMLR, pp 26–37
- Ren B, Yao Y, Sun R, et al (2026) Aligning agentic world models via knowledgeable experience learning. arXiv preprint arXiv:260113247 URL <https://arxiv.org/abs/2601.13247>
- Rigter M, Lacerda B, Hawes N (2022) RAMBO-RL: Robust adversarial model-based offline reinforcement learning. In: Advances in Neural Information Processing Systems (NeurIPS), pp 16082–16097
- Rigter M, Yamada J, Posner I (2023) World models via policy-guided trajectory diffusion. arXiv preprint arXiv:231208533 URL <https://arxiv.org/abs/2312.08533>
- Robine J, Höftmann M, Uelwer T, et al (2023) Transformer-based world models are happy with 100k interactions. arXiv preprint arXiv:230307109 URL <https://arxiv.org/abs/2303.07109>
- Russell L, Hu A, Bertoni L, et al (2025) GAIA-2: A controllable multi-view generative world model for autonomous driving. arXiv preprint arXiv:250320523 URL <https://arxiv.org/abs/2503.20523>
- S V J, Jalagam S, LeCun Y, et al (2023) Gradient-based planning with world models. arXiv preprint arXiv:231217227 URL <https://arxiv.org/abs/2312.17227>
- Schrittwieser J, Antonoglou I, Hubert T, et al (2020) Mastering atari, go, chess and shogi by planning with a learned model. Nature 588(7839):604–609. URL <https://doi.org/10.1038/s41586-020-03051-4>

- Sekar R, Rybkin O, Daniilidis K, et al (2020) Planning to explore via self-supervised world models. In: International conference on machine learning, PMLR, pp 8583–8592
- Shang Y, Li Z, Ma Y, et al (2026) Worldarena: A unified benchmark for evaluating perception and functional utility of embodied world models. arXiv preprint arXiv:260208971 URL <https://arxiv.org/abs/2602.08971>
- Shanks S, Embley-Riches J, Liu J, et al (2025) DreamerNav: Learning-based autonomous navigation in dynamic indoor environments using world models. *Frontiers in Robotics and AI* 12:1655171. URL <https://doi.org/10.3389/frobt.2025.1655171>
- Sharma AK, Sun Y, Lu N, et al (2026) World-gymnast: Training robots with reinforcement learning in a world model. arXiv preprint arXiv:260202454 URL <https://arxiv.org/abs/2602.02454>
- Shi C, Shi S, Sheng K, et al (2025a) DriveX: Omni scene modeling for learning generalizable world knowledge in autonomous driving. In: IEEE/CVF International Conference on Computer Vision (ICCV), URL <https://arxiv.org/abs/2505.19239>
- Shi LX, Lim JJ, Lee Y (2022) Skill-based model-based reinforcement learning (SkiMo). In: Conference on Robot Learning (CoRL), pp 2262–2272
- Shi Z, Fang M, Chen L (2025b) Monte carlo planning with large language model for text-based game agents. In: International Conference on Learning Representations (ICLR)
- Srivastava A, Rastogi A, Rao A, et al (2022) Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. arXiv preprint arXiv:220604615
- Sun R, Zang H, Li X, et al (2024) Learning latent dynamic robust representations for world models. In: International Conference on Machine Learning (ICML), pp 47234–47260
- Sun S, Zhang H, Liu Z, et al (2025) Minco: Mitigating information conflicts in distracted visual model-based reinforcement learning. arXiv preprint arXiv:250404164 URL <https://arxiv.org/abs/2504.04164>
- Tang Y, Shang Y, Chen Y, et al (2026) Roboscape-r: Unified reward-observation world models for generalizable robotics training via rl. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 1472–1481
- Tolman EC (1948) Cognitive maps in rats and men. *Psychological review* 55(4):189–208. URL <https://doi.org/10.1037/h0061626>

- Upadhyay R, et al (2026) WorldBench: Disambiguating physics for diagnostic evaluation of world models. arXiv preprint arXiv:260121282 URL <https://arxiv.org/abs/2601.21282>
- Vasudevan AB, Peri N, Schneider J, et al (2025) Planning with adaptive world models for autonomous driving. In: IEEE International Conference on Robotics and Automation (ICRA)
- Wang H, Ye X, Tao F, et al (2025) Adawm: Adaptive world model based planning for autonomous driving. In: International Conference on Learning Representations, pp 85591–85615
- Wang J, Zhang Q, Yang S, et al (2026a) Repwam: World action modeling with representation visual-action tokenizers. arXiv preprint arXiv:260613674 URL <https://arxiv.org/abs/2606.13674>
- Wang L, Zheng Y, Chen Q, et al (2026b) Latent-wam: Latent world action modeling for end-to-end autonomous driving. arXiv preprint arXiv:260324581 URL <https://arxiv.org/abs/2603.24581>
- Wang M, Jin W, Cao K, et al (2026c) Contactgaussian-wm: Learning physics-grounded world model from videos. arXiv preprint arXiv:260211021 URL <https://arxiv.org/abs/2602.11021>
- Wang Q, Yang J, Wang Y, et al (2024a) CoWorld: Making offline RL online: Collaborative world models for offline visual RL. In: Advances in Neural Information Processing Systems (NeurIPS), URL <https://arxiv.org/abs/2305.15260>
- Wang R, Todd G, Xiao Z, et al (2024b) Can language models serve as text-based world simulators? In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp 1–17
- Wang T, Du SS, Torralba A, et al (2023) Denoised mdps: Learning world models better than the world itself. URL <https://arxiv.org/abs/2206.15477>
- Wang X, Zhu Z, Huang G, et al (2024c) DriveDreamer: Towards real-world-driven world models for autonomous driving. In: European Conference on Computer Vision (ECCV)
- Wu J, Ma H, Deng C, et al (2023) Pre-training contextualized world models with in-the-wild videos for reinforcement learning. URL <https://arxiv.org/abs/2305.18499>
- Wu J, Yin S, Feng N, et al (2025) RLVR-World: Training world models with reinforcement learning. In: Advances in Neural Information Processing Systems (NeurIPS), URL <https://arxiv.org/abs/2505.13934>

- Wu Z, Dvornik N, Greff K, et al (2022) Slotformer: Unsupervised visual dynamics simulation with object-centric models. arXiv preprint arXiv:221005861 URL <https://arxiv.org/abs/2210.05861>
- Xiao L, Liu JJ, Yang S, et al (2025) Learning multiple probabilistic decisions from latent world model in autonomous driving. In: IEEE International Conference on Robotics and Automation (ICRA)
- Xu G, Zhang Q, Zhou J, et al (2026a) Next forcing: Causal world modeling with multi-chunk prediction. arXiv preprint arXiv:260611187 URL <https://arxiv.org/abs/2606.11187>
- Xu X, Lin Z, He K, et al (2026b) WorldMark: A unified benchmark suite for interactive video world models. arXiv preprint arXiv:260421686 URL <https://arxiv.org/abs/2604.21686>
- Yang J, Lin K, Li J, et al (2026a) Rise: Self-improving robot policy with compositional world model. arXiv preprint arXiv:260211075 URL <https://arxiv.org/abs/2602.11075>
- Yang N, Huang Y, Peng K, et al (2026b) Wam-nav: Asymmetric latent world-action modeling for unified visual navigation. arXiv preprint arXiv:260604907 URL <https://arxiv.org/abs/2606.04907>
- Yang P, Lu B, Xia Z, et al (2026c) Worldrft: Latent world model planning with reinforcement fine-tuning for autonomous driving. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp 11649–11657
- Yang P, Zheng Y, Qian D, et al (2026d) Dreamerad: Efficient reinforcement learning via latent world model for autonomous driving. arXiv preprint arXiv:260324587 URL <https://arxiv.org/abs/2603.24587>
- Yang Y, Wang ZY, Liu Q, et al (2025) Medical world model: Generative simulation of tumor evolution for treatment planning. arXiv preprint arXiv:250602327 URL <https://arxiv.org/abs/2506.02327>
- Ye S, Ge Y, Zheng K, et al (2026a) World action models are zero-shot policies. arXiv preprint arXiv:260215922 URL <https://arxiv.org/abs/2602.15922>
- Ye W, Liu S, Kurutach T, et al (2021) Mastering atari games with limited data. Advances in neural information processing systems 34:25476–25488
- Ye Y, Lu X, Jiang Y, et al (2026b) MIND: Benchmarking memory consistency and action control in world models. arXiv preprint arXiv:260208025 URL <https://arxiv.org/abs/2602.08025>

- Yin S, Yin K, Chen W, et al (2026) Ddp-wm: Disentangled dynamics prediction for efficient world models. URL <https://arxiv.org/abs/2602.01780>
- Ying K, Hu H, Ren S, et al (2026) Wbench: A comprehensive multi-turn benchmark for interactive video world model evaluation. arXiv preprint arXiv:260525874 URL <https://arxiv.org/abs/2605.25874>
- Yu D, Zou W, Yang Y, et al (2023) Safe model-based reinforcement learning with an uncertainty-aware reachability certificate. IEEE Transactions on Automation Science and Engineering 21(3):4129–4142
- Yu H, Lin H, Zhang J, et al (2026a) Maskwam: Unifying mask prompting and prediction for world-action models. arXiv preprint arXiv:260613515 URL <https://arxiv.org/abs/2606.13515>
- Yu T, Thomas G, Yu L, et al (2020) MOPO: Model-based offline policy optimization. In: Advances in Neural Information Processing Systems (NeurIPS), pp 14129–14142
- Yu T, Kumar A, Rafailov R, et al (2021) COMBO: Conservative offline model-based policy optimization. In: Advances in Neural Information Processing Systems (NeurIPS), pp 28954–28967
- Yu X, Peng B, Xu R, et al (2026b) Reinforcement world model learning for LLM-based agents. arXiv preprint arXiv:260205842 URL <https://arxiv.org/abs/2602.05842>
- Yue Y, Wang Y, Tao C, et al (2025) CheXWorld: Exploring image world modeling for radiograph representation learning. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 20778–20788, URL <https://arxiv.org/abs/2504.13820>
- Zanatta L, Malczyk G, Alexis K (2026) Generalization of world models under environmental variability for vision-based quadrotor navigation. arXiv preprint arXiv:260605015 URL <https://arxiv.org/abs/2606.05015>
- Zhang J, Li A, Qi Y, et al (2025a) Embodied navigation foundation model. arXiv preprint arXiv:250912129 URL <https://arxiv.org/abs/2509.12129>
- Zhang J, Fu Z, Xu Z, et al (2026) ResWorld: Temporal residual world model for end-to-end autonomous driving. In: International Conference on Learning Representations (ICLR), URL <https://arxiv.org/abs/2602.10884>
- Zhang K, Tang Z, Hu X, et al (2025b) Epona: Autoregressive diffusion world model for autonomous driving. In: IEEE/CVF International Conference on Computer Vision (ICCV), URL <https://arxiv.org/abs/2506.24113>
- Zhang L, Xiong Y, Yang Z, et al (2024) Copilot4D: Learning unsupervised world models for autonomous driving via discrete diffusion. In: International Conference

on Learning Representations (ICLR)

- Zhang Z, Liniger A, Dai D, et al (2023) TrafficBots: Towards world models for autonomous driving simulation and motion prediction. In: IEEE International Conference on Robotics and Automation (ICRA), pp 1522–1529
- Zhao G, Wang X, Zhu Z, et al (2025) DriveDreamer-2: LLM-enhanced world models for diverse driving video generation. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp 10412–10420
- Zheng X, Yang Z, Chen Y, et al (2025) Coirl-ad: Collaborative-competitive imitation-reinforcement learning in latent world models for autonomous driving. arXiv preprint arXiv:251012560 URL <https://arxiv.org/abs/2510.12560>
- Zhi H, Chen P, Zhou S, et al (2025) 3DFlowAction: Learning cross-embodiment manipulation from 3d flow world model. arXiv preprint arXiv:250606199 URL <https://arxiv.org/abs/2506.06199>
- Zhou X, Liang D, Tu S, et al (2025) HERMES: A unified self-driving world model for simultaneous 3D scene understanding and generation. In: IEEE/CVF International Conference on Computer Vision (ICCV), pp 27817–27827
- Zhu Z, Liu M, Mao L, et al (2024) MADiff: Offline multi-agent learning with diffusion models. In: Advances in Neural Information Processing Systems (NeurIPS), pp 4177–4206