Toward Continually Growing World Models: An OaK-Inspired Architecture for Learned Abstraction and Planning

An Xu*
Technical University of Munich
Harbin, China
AnX.RTL2324@tum-asia.edu.sg

Abstract

Learned world models increasingly empower autonomous agents to predict transitions, imagine rollouts, and plan behaviors in complex environments. However, conventional continual world modeling paradigms predominantly formulate lifelong adaptation along a single parametric axis: updating neural network weights ($\theta_t \to \theta_{t+1}$) over a static, fixed-dimensional latent substrate and a fixed single-step temporal granularity. When deployed in non-stationary or open-ended environments, this parameter-only adaptation is vulnerable to loss of plasticity, capacity saturation, and compounding rollout errors over extended horizons. In this position paper, we argue that genuine continual world modeling requires introducing a second, structural adaptation axis ($\mathcal{K}_t \to \mathcal{K}_{t+1}$), wherein the model's predictive state abstractions, temporal abstractions, and option models continually grow, evaluate, and reorganize over time. Grounding this dual-axis formulation in Richard Sutton's OaK architecture and its FC-STOMP lifecycle (Feature Construction $\to$ SubTask $\to$ Option $\to$ Model $\to$ Planning), we formulate *Structural Continual World Modeling* (SCWM). Under SCWM, discovered state features dynamically expand the world model's predictive state representation, inducing reward-respecting subtasks, closed-loop options, and predictive option models, with retention and pruning governed by downstream planning utility. We formulate five falsifiable research hypotheses and an experimental agenda to guide the development of continually growing world models.

1 Introduction

Model-based reinforcement learning (MBRL) with learned latent world models has demonstrated notable success across visual continuous control and robotics [Ha and Schmidhuber, 2018, Hafner et al., 2019, 2020, 2023, Hansen et al., 2024b, Richens et al., 2025], synthesizing imagined trajectories in latent space for sample-efficient policy optimization and model predictive control (MPC).

Despite their predictive power, many modern world models retain a fixed representational topology and temporal formulation after model design/training. Practitioners fix a latent dimensionality $z_t \in \mathcal{Z} = \mathbb{R}^{d_z}$ and train single-step forward dynamics $p_\theta(z_{t+1}, r_{t+1} \mid z_t, a_t)$ over primitive actions $a_t \in \mathcal{A}$. A dominant formulation in continual world modeling focuses on parameter adaptation within a fixed predictive architecture: $\theta_{t+1} = \mathrm{Opt}(\theta_t, \mathcal{D}_{\text{new}})$ [Kessler et al., 2023, Yang et al., 2024, Liu et al., 2025].

However, this parameter-only view fails under qualitative distribution shifts. Under the *Big World Hypothesis* [Javed and Sutton, 2024, Lewandowski et al., 2025, Elelimy et al., 2025], the physical world is vastly larger than any bounded agent; no pre-allocated latent representation or single

*Idea Track paper submitted to the NeurIPS 2026 Workshop on Continual World Models.

Preprint.

temporal scale can encapsulate all future interactions. Forcing continual adaptation solely through gradient weight adjustments in a static network precipitates *loss of plasticity* [Dohare et al., 2024] and catastrophic forgetting [Fu et al., 2025]. Furthermore, single-step rollouts ($z_t \to \cdots \to z_{t+H}$) suffer from compounding errors and exponential search complexity over long horizons [Freed et al., 2023, Jiralerspong et al., 2023].

We advocate that continual world modeling requires expanding beyond the parameter adaptation axis ($\theta_t \to \theta_{t+1}$) to encompass a **structural adaptation axis** ($\mathcal{K}_t \to \mathcal{K}_{t+1}$). Rather than merely re-tuning weights, the world model must continually discover, evaluate, and prune predictive abstractions across state and time. Grounding this dual-axis vision in Sutton's **OaK architecture** [Sutton, 2025]—which synthesizes the Alberta Plan [Sutton et al., 2022], reward-respecting subtasks [Sutton et al., 2023], constructive representations [Javed et al., 2023, Javed and White, 2019], and options [Sutton et al., 1999, 1998] into the **FC-STOMP** progression—we formulate **Structural Continual World Modeling** (SCWM).

Key Contributions. **(1) Dual-Axis Framework:** We formalize SCWM, augmenting parameter-level adaptation ($\theta_t \to \theta_{t+1}$) with an explicit structural adaptation axis $(\theta_t, \mathcal{K}_t) \to (\theta_{t+1}, \mathcal{K}_{t+1})$. **(2) Constructive Dynamics:** We formalize how discovered features $\mathcal{F}_t$ directly expand the world model's predictive state representation ($\tilde{z}_t = \Phi_t(z_t)$), inducing reward-respecting subtasks with terminal stopping bonuses, predictive option models, and multi-scale planning. **(3) Actionable Agenda:** We formulate five falsifiable hypotheses (H1–H5) and experimental protocols evaluating planning depth, plasticity retention, and Pareto-optimal utility pruning.

2 Beyond Parameter-Level Continual World Modeling

Standard continual world modeling formalizes adaptation as parameter trajectory estimation under a stationary architecture:

$$\theta_{t+1} = \arg\min_{\theta} \mathbb{E}_{(o,a,o')\sim\mathcal{D}_{\text{new}}} \left[\mathcal{L}_{\text{WM}}(M_\theta(E_\phi(o), a), o')\right] + \Omega(\theta, \theta_t), \tag{1}$$

where Ω regularizes weight drift or penalizes rehearsal divergence [Kessler et al., 2023, Yang et al., 2024]. We identify three fundamental structural bottlenecks: **1. Fixed Representational Topology:** The encoder $E_\phi : \Omega \to \mathcal{Z}$ enforces a fixed coordinate dimensionality $\mathcal{Z} \subset \mathbb{R}^{d_z}$, unable to dynamically allocate additional predictive dimensions for novel interaction affordances. **2. Fixed Temporal Granularity:** Single-step dynamics $(z_t, a_t) \mapsto z_{t+1}$ require autoregressive rollouts $\hat{z}_{t+H} \sim \prod_{k=0}^{H-1} p_\theta(z_{t+k+1} \mid z_{t+k}, a_{t+k})$, compounding approximation errors and exploding search over $\mathcal{A}^H$ [Freed et al., 2023, Jiralerspong et al., 2023]. **3. Capacity Saturation:** Sustained non-stationary gradient updates cause unit saturation, gradient rank collapse, and plasticity decay [Dohare et al., 2024].

Definition 1 (**Structural Continual World Modeling**)**.** A *Structural Continual World Model* is defined by $\mathcal{W}_t = (\theta_t, \mathcal{K}_t)$, where θ_t denotes continuous parameters and $\mathcal{K}_t = (\mathcal{F}_t, \mathcal{G}_t, \mathcal{O}_t, \mathcal{M}_t)$ is a knowledge structure comprising: constructed features $\mathcal{F}_t = \{f_i : \mathcal{Z} \to \mathbb{R}\}$, subtasks $\mathcal{G}_t$, closed-loop options $\mathcal{O}_t = \{o_i = (\mathcal{I}_i, \pi_i, \beta_i)\}$, and predictive option models $\mathcal{M}_t = \{\mathcal{M}_{o_i}\}$. The predictive state representation $\tilde{z}_t \in \tilde{\mathcal{Z}}_t$ dynamically expands with $\mathcal{F}_t$:

$$\tilde{z}_t = \Phi_t(z_t) \triangleq \left[z_t^\top,\ f_1(z_t),\ \ldots,\ f_{|\mathcal{F}_t|}(z_t)\right]^\top \in \mathbb{R}^{d_z + |\mathcal{F}_t|}. \tag{2}$$

Lifelong learning is governed by joint parameter adaptation and structural knowledge evolution:

$$(\theta_t, \mathcal{K}_t) \xrightarrow{\text{Experience } e_{t:t+\Delta}} (\theta_{t+1}, \mathcal{K}_{t+1}), \quad \text{where } \mathcal{K}_{t+1} \neq \mathcal{K}_t \text{ in general.} \tag{3}$$

3 An OaK-Inspired Continually Growing World Model

We instantiate Definition 1 by systematically operationalizing Sutton's FC-STOMP lifecycle [Sutton, 2025] into latent world models (Figure 1, Table 1).

3.1 Base Latent Dynamics Foundation

Given observation stream $o_{\leq t}$, base encoder E_ϕ extracts belief embedding $z_t = E_\phi(o_{\leq t})$. Incorporating discovered features $\mathcal{F}_t$, the world model operates over the extended predictive state $\tilde{z}_t = \Phi_t(z_t) \in \mathbb{R}^{d_z + |\mathcal{F}_t|}$ (Eq. 2), predicting 1-step transitions and rewards: $\hat{\tilde{z}}_{t+1}, \hat{r}_{t+1} \sim p_{\theta,t}(\tilde{z}_{t+1}, r_{t+1} \mid \tilde{z}_t, a_t)$ for $a_t \in \mathcal{A}$.

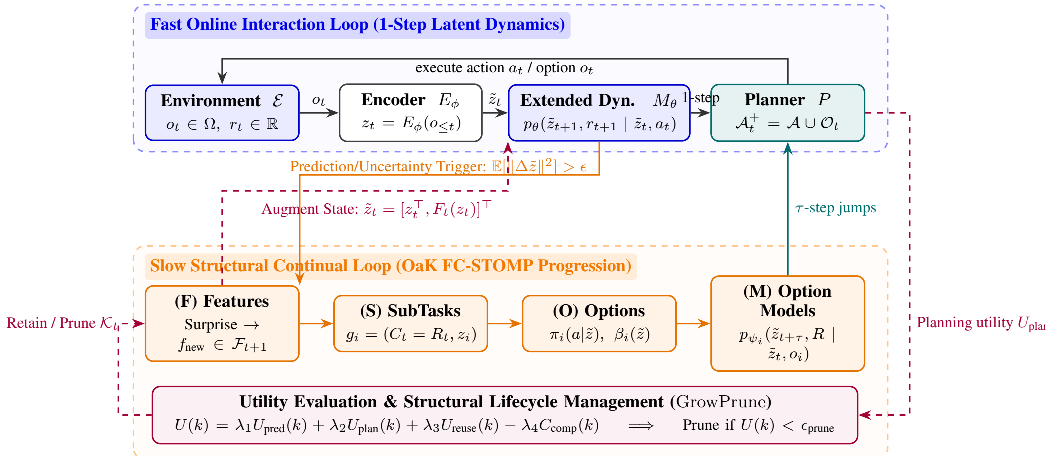

Figure 1: **Architectural overview of the OaK-inspired Continually Growing World Model.** A *Fast Base Dynamics Loop* handles 1-step primitive predictions over the extended state $\tilde{z}_t$, coupled with a *Slow Structural Adaptation Loop* executing the OaK FC-STOMP cycle. Systematic prediction surprises trigger feature construction (F), expanding $\tilde{z}_t$ and inducing subtasks (S), options (O), and predictive option models (M). Multi-scale planning (P) leverages both primitive and abstract transitions, returning empirical planning utility to retain or prune components.

3.2 Continual Feature Construction (F)

When systematic prediction surprises or epistemic uncertainties exceed a cognitive threshold ($\mathbb{E}[\|\tilde{z}_{t+1} - \hat{\tilde{z}}_{t+1}\|^2] > \epsilon_{\text{trigger}}$), candidate feature functions $f_{\text{new}} : \mathcal{Z} \to \mathbb{R}$ are generated:

$$f_i(z_t) = \sigma\left(\boldsymbol{w}_i^\top \psi(z_t) + b_i\right), \quad \mathcal{F}_{t+1} = \mathcal{F}_t \cup \{f_{\text{new}}\}, \tag{4}$$

where $\psi(z_t)$ represents constructive hidden units [Javed et al., 2023, Kearney et al., 2019]. Adding f_{new} expands $\tilde{z}_t \to \tilde{z}_t^{(t+1)}$, structurally expanding the computational basis of the predictive state representation. Extending structural construction to recover predictive information omitted by the base encoder remains an important open problem.

3.3 Reward-Respecting Subtasks (S) and Options (O)

Following Sutton et al. [Sutton et al., 2023], each constructed feature $f_i \in \mathcal{F}_t$ induces a *reward-respecting subtask* $g_i = (C_t = R_t, z_i) \in \mathcal{G}_t$. The subtask preserves environmental reward $C_t = R_t$ as its ongoing cumulant, combined with a terminal stopping bonus upon option termination:

$$z_i(\tilde{z}_{\text{term}}) = \hat{v}(\tilde{z}_{\text{term}}) + (\bar{w}_i - w_i) f_i(z_{\text{term}}), \tag{5}$$

where $\hat{v}(\tilde{z})$ is the baseline value estimate and $(\bar{w}_i - w_i)$ modulates feature attainment. Solving subtask g_i yields an option $o_i = (\mathcal{I}_i, \pi_i, \beta_i) \in \mathcal{O}_t$ [Sutton et al., 1999], where $\pi_i(a \mid \tilde{z})$ is the policy, $\beta_i(\tilde{z}) \in [0, 1]$ is the termination condition, and one natural choice for initiation set is $\mathcal{I}_i = \{\tilde{z} \mid f_i(z) \ge \delta_{\text{init}}\}$.

3.4 Predictive Option Models (M)

For each option $o_i \in \mathcal{O}_t$, the agent fits a multi-step jump-ahead transition and cumulative return model $\mathcal{M}_{o_i} \in \mathcal{M}_t$ over extended states [Sutton et al., 1998, Freed et al., 2023, Rodriguez-Sanchez and Konidaris, 2024]:

$$\mathcal{M}_{o_i} : (\tilde{z}_t, o_i) \mapsto (\hat{\tilde{z}}_{t+\tau}, \hat{R}_{t:t+\tau}, \hat{\tau}) \sim p_{\psi_i}(\tilde{z}_{t+\tau}, R_{t:t+\tau}, \tau \mid \tilde{z}_t, o_i), \tag{6}$$

where $\tau \in \mathbb{N}_{\ge 1}$ is duration and $R_{t:t+\tau} = \sum_{k=0}^{\tau-1} \gamma^k r_{t+k+1}$ is cumulative return, trained off-policy via intra-option model learning [Sutton et al., 1998].

3.5 Planning Across Temporal Scales (P)

The planner searches over hybrid action space $\mathcal{A}_t^+ = \mathcal{A} \cup \mathcal{O}_t$, traversing primitive steps and multi-step jumps ($\tilde{z}_0 \xrightarrow{a_0} \dots \xrightarrow{o_k} \tilde{z}_{t+H}$), contracting graph diameter for deep horizon search [Jiralerspong et al., 2023, Precup et al., 1997].

Table 1: **Mapping Sutton's OaK FC-STOMP Lifecycle to Structural Continual World Models.**

OaK Component	World-Model Interpretation	Formal Representation / Mechanism
Base Substrate	Parametric latent dynamics baseline	$z_t = E_\phi(o_{\leq t}),\ \ p_\theta(\tilde{z}_{t+1}, r_{t+1} \mid \tilde{z}_t, a_t)$
Feature Construction (F)	Predictive state representation expansion	$\mathcal{F}_{t+1} = \mathcal{F}_t \cup \{f_{\text{new}}\},\ \ \tilde{z}_t = [z_t^\top, f_1(z_t), \dots]^\top$
SubTask (S)	Reward-respecting local control objective	$\mathcal{G}_{t+1} = \mathcal{G}_t \cup \{g_i\},\ \ g_i = (C_t = R_t, z_i)$
Option (O)	Closed-loop temporally extended behavior	$\mathcal{O}_{t+1} = \mathcal{O}_t \cup \{o_i\},\ \ o_i = (\mathcal{I}_i, \pi_i, \beta_i)$
Option Model (M)	Temporal jump-ahead predictive dynamics	$\mathcal{M}_{o_i} : (\tilde{z}_t, o_i) \mapsto (\hat{\tilde{z}}_{t+\tau}, \hat{R}_{t:t+\tau}, \hat{\tau}) \sim p_{\psi_i}$
Planning (P)	Multi-scale forward tree/trajectory search	$\mathcal{A}_t^+ = \mathcal{A} \cup \mathcal{O}_t$, Search over hybrid rollouts
Utility Feedback	Knowledge retention, revision, and pruning	$U(k) = \lambda_1 U_{\text{pred}} + \lambda_2 U_{\text{plan}} + \lambda_3 U_{\text{reuse}} - \lambda_4 C_{\text{comp}}$

3.6 Utility-Driven Growth and Pruning: Managing Complexity

To balance capacity and avoid search explosion in $\mathcal{A}_t^+$, abstraction utility $U(k)$ (Table 1) is audited across accuracy, planning utility, and reuse. Low-utility elements ($U(k) < \epsilon_{\text{prune}}$) are pruned: $\mathcal{K}_{t+1} = \text{GrowPrune}(\mathcal{K}_t, \mathcal{E}_{t:t+\Delta}, U)$, favoring a compact, high-utility abstraction inventory.

4 Research Hypotheses and Experimental Agenda

We articulate five falsifiable hypotheses: **H1 (Planning Horizon):** Planning over hybrid $\mathcal{A}_t^+ = \mathcal{A} \cup \mathcal{O}_t$ achieves deeper horizons and higher returns than 1-step planning (p_θ) under fixed compute B_{plan}. **H2 (Adaptation Speed):** Constructing option models accelerates policy adaptation under shifts $\mathcal{D}_1 \to \cdots \to \mathcal{D}_N$ relative to fixed-architecture tuning. **H3 (Growth Frontier):** Multi-objective utility pruning yields a superior Pareto frontier over unconstrained growth ($|\mathcal{K}_t| \to \infty$) or prediction-only pruning. **H4 (Plasticity Preservation):** Structurally allocating feature-option units better preserves representation plasticity (effective rank, dormant units) relative to fixed-capacity continual baselines under sustained non-stationarity [Dohare et al., 2024]. **H5 (Planning Utility):** Abstractions retained via planning utility (U_{plan}) outperform abstractions selected purely via reconstruction (U_{pred}).

Experimental Protocols. We propose evaluating across continual and long-horizon control benchmarks (Crafter, Meta-World, Continual DMC): **(Exp A: Horizon vs. Compute)** Compare 1-step MPC (p_θ) against Option-MPC ($\mathcal{M}_\theta \cup \mathcal{M}_\mathcal{O}$) under fixed simulation budgets. **(Exp B: Sequential Shift)** Benchmark forward/backward transfer across non-stationary dynamics against Continual-Dreamer [Kessler et al., 2023], Online-WM [Liu et al., 2025], and DRAGO [Fu et al., 2025]. **(Exp C: Utility Ablation)** Compare unconstrained growth, random pruning, prediction-only pruning ($U = U_{\text{pred}}$), and multi-objective pruning ($U = U_{\text{pred}} + U_{\text{plan}} - C_{\text{comp}}$).

5 Discussion and Open Questions

Statistical Triggers. Differentiating stochastic environmental noise from systematic epistemic error is crucial; epistemic uncertainty metrics (ensemble divergence, latent variance) offer principled creation triggers. **Architecture-Compatible Expansion.** When predictive state $\tilde{z}_t$ dynamically expands ($|\mathcal{F}_t| \to |\mathcal{F}_t| + 1$), neural dynamics must allocate capacity while preserving previously learned transitions. Designing function-preserving structural expansion remains a central open problem for SCWM. **Relation to Concurrent Paradigms.** AgentOWL [Piriyakulkij et al., 2026] cumulatively acquires hierarchical options and updates an abstract world model, while SALT [Huang and Freed, 2026] learns state- and temporally-abstract dynamics from offline trajectories and Forecaster [Jiralerspong et al., 2023] focuses on temporally abstract planning. SCWM instead emphasizes the autonomous structural lifecycle of the predictive abstraction inventory itself—feature construction, subtask formation, option-model acquisition, utility evaluation, revision, and pruning—during lifelong interaction. In contrast to model-free hierarchical RL [Bacon et al., 2017] or fixed hierarchies [Schiewer et al., 2024, Hansen et al., 2024a], SCWM addresses ongoing structural emergence. **Theoretical Grounding.** The Big World Hypothesis [Javed and Sutton, 2024, Lewandowski et al., 2025, Elelimy et al., 2025, Kumar et al., 2023] motivates the view that bounded agents operating in vast worlds may require continually evolving predictive knowledge rather than a permanently fixed model structure. OaK provides one constructive architecture for operationalizing this view.

6 Conclusion

Continual world modeling must transcend parameter fine-tuning within static predictive topologies. By formalizing a structural adaptation axis alongside the parameter axis, *Structural Continual World Modeling* (SCWM) provides a principled foundation for continually growing world models: discovering, organizing, and retaining predictive abstractions worth having.

Acknowledgments and Disclosure of Funding

This research was supported by investigations into continual model-based reinforcement learning and the OaK architecture.

References

Pierre-Luc Bacon, Jean Harb, and Doina Precup. The option-critic architecture. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 31, pages 1726–1734, 2017.

Shibhansh Dohare, J. Fernando Hernandez-Garcia, Parash Rahman, Richard S. Sutton, and Mahmood R. Mahmood. Loss of plasticity in deep continual learning. *Nature*, 632:768–774, 2024. doi: 10.1038/s41586-024-07711-7.

Esraa Elelimy, David Szepesvari, Martha White, and Michael Bowling. Rethinking the foundations for continual reinforcement learning. *Reinforcement Learning Journal*, 2, 2025.

Benjamin Freed, Siddarth Venkatraman, Guillaume Adrien Sartoretti, Jeff Schneider, and Howie Choset. Learning temporally abstract world models without online experimentation. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, volume 202 of *PMLR*, pages 10271–10287, 2023.

Haotian Fu, Yixiang Sun, Michael L. Littman, and George Konidaris. Knowledge retention in continual model-based reinforcement learning. In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, volume 267 of *PMLR*, pages 17832–17851, 2025.

David Ha and Jürgen Schmidhuber. Recurrent world models facilitate policy evolution. *Advances in Neural Information Processing Systems (NeurIPS)*, 31, 2018.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, volume 97 of *PMLR*, pages 2555–2565, 2019.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations (ICLR)*, 2020.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*, 2023.

Nicklas Hansen, Hao Su, and Xiaolong Wang. Hierarchical world models as visual whole-body humanoid controllers. *arXiv preprint arXiv:2405.18418*, 2024a.

Nicklas Hansen, Hao Su, and Xiaolong Wang. TD-MPC2: Scalable, robust world models for continuous control. In *International Conference on Learning Representations (ICLR)*, 2024b.

William Huang and Benjamin Freed. Learning temporally and state-abstracted world models for long-horizon exploration. In *ICRA Workshop on Structure and Abstraction in Robot Learning*, 2026.

Khurram Javed and Richard S. Sutton. The big world hypothesis and its ramifications for artificial intelligence. In *RLC Workshop on Finding the Frame*, 2024.

Khurram Javed and Martha White. Meta-learning representations for continual learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Khurram Javed, Haseeb Shah, Richard S. Sutton, and Martha White. Scalable real-time recurrent learning using columnar-constructive networks. *Journal of Machine Learning Research*, 24(256): 1–34, 2023.

Thomas Jiralerspong, Frederik Träuble, Yoshua Bengio, and Martin Jagersand. Forecaster: Towards temporally abstract tree-search planning from pixels. *arXiv preprint arXiv:2310.09997*, 2023.

Alex Kearney, Vincent Liu, Richard S. Sutton, and Patrick M. Pilarski. Learning feature relevance through step size adaptation in temporal-difference learning. *arXiv preprint arXiv:1903.03252*, 2019.

Samuel Kessler, Jack Parker-Holder, Philip J. Ball, Stefan Zohren, and Stephen J. Roberts. The effectiveness of world models for continual reinforcement learning. In *Conference on Lifelong Learning Agents (CoLLAs)*, volume 232 of *PMLR*, pages 400–419, 2023.

Saurabh Kumar, Henrik Marklund, and Benjamin Van Roy. Continual learning as computationally constrained reinforcement learning. *arXiv preprint arXiv:2307.04345*, 2023.

Alex Lewandowski, Aditya A. Ramesh, Edan Meyer, Dale Schuurmans, and Marlos C. Machado. The world is bigger! a computationally-embedded perspective on the big world hypothesis. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 38, 2025.

Zichen Liu, Guoji Fu, Chao Du, Wee Sun Lee, and Min Lin. Continual reinforcement learning by planning with online world models. In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, volume 267 of *PMLR*, pages 38397–38423, 2025.

P. Piriyakulkij, S. Lehrach, K. Ellis, and K. Murphy. Joint learning of hierarchical neural options and abstract world model. *arXiv preprint arXiv:2602.02799*, 2026.

Doina Precup, Richard S. Sutton, and Satinder Singh. Planning with closed-loop macro actions. In *AAAI Fall Symposium on Model-directed Autonomous Systems*, pages 70–76, 1997.

Jonathan Richens, Tom Everitt, and David Abel. General agents need world models. In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, volume 267 of *PMLR*, pages 42110–42145, 2025.

Rafael Rodriguez-Sanchez and George Konidaris. Learning abstract world models for value-preserving planning with options. *Reinforcement Learning Journal*, 4, 2024.

M. Schiewer, A. Subramoney, and L. Wiskott. Exploring the limits of hierarchical world models in reinforcement learning. *Scientific Reports*, 14(1):26856, 2024. doi: 10.1038/s41598-024-76719-w.

Richard S. Sutton. The OaK architecture: A vision of superintelligence from experience. Invited Talk at NeurIPS and Keynote at Reinforcement Learning Conference (RLC), 2025.

Richard S. Sutton, Doina Precup, and Satinder Singh. Intra-option learning about temporally abstract actions. In *Proceedings of the 15th International Conference on Machine Learning (ICML)*, pages 556–564, 1998.

Richard S. Sutton, Doina Precup, and Satinder Singh. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112(1–2):181–211, 1999.

Richard S. Sutton, Michael Bowling, and Patrick M. Pilarski. The Alberta plan for AI research. *arXiv preprint arXiv:2208.11173*, 2022.

Richard S. Sutton, Marlos C. Machado, G. Zacharias Holland, David Szepesvári, Finbarr Timbers, Brian Tanner, and Adam White. Reward-respecting subtasks for model-based reinforcement learning. *Artificial Intelligence*, 324:104001, 2023. doi: 10.1016/j.artint.2023.104001.

Yifei Yang, Guozheng Ma, and Zhen Wang. Augmenting replay in world models for continual reinforcement learning. *arXiv preprint arXiv:2401.16650*, 2024.