Towards Cross-Game World Models with Latent Action Alignment from Web Gameplay Telemetry

Anonymous authors
Paper under double-blind review

Abstract

Building world models that transfer across multiple video games remains difficult because games differ not only in visuals and mechanics, but also in how player input is expressed and interpreted. Web-based games (HTML5 and WebGL) offer an underused frontier for telemetry collection because of their open delivery pipelines and standardized browser runtimes. However, raw player inputs, such as mouse and keyboard events, possess no intrinsic semantic invariance across different game engines or control schemes, disrupting cross-game transition learning. We present a position-and-methodology paper for learning cross-game world models from heterogeneous gameplay data. Rather than treating latent action learning as an engineering fix for messy telemetry, we argue that action abstraction is a structural requirement for transfer. We propose a hierarchical pipeline that uses browser actions as weak supervision to learn a unified discrete latent action space representing gameplay intent, and we frame web gameplay telemetry as a useful substrate for dataset-driven RL and transfer studies.

1 Introduction and Motivation

Model-based reinforcement learning (MBRL) has emerged as a powerful paradigm for training intelligent agents in complex simulated environments. By learning predictive models of environment dynamics, agents can plan, imagine trajectories, and optimize behavior with substantially greater sample efficiency than model-free alternatives. Foundational examples such as World Models, MuZero, and DreamerV3 illustrate how learned dynamics can support strong performance across domains ranging from arcade benchmarks to large open-ended environments (Ha & Schmidhuber, 2018; Schrittwieser et al., 2019; Hafner et al., 2023).

Despite these advances, reinforcement learning still lacks world models that transfer robustly across distinct games. One bottleneck is data scarcity at scale. Unlike natural language processing and computer vision, which benefit from internet-scale corpora, reinforcement learning in games remains anchored to curated benchmarks or domain-specific ecosystems such as the Arcade Learning Environment, Procgen, MineRL, and MineDojo (Bellemare et al., 2013; Cobbe et al., 2019; Guss et al., 2019; Fan et al., 2022).

We argue that web-based games, including HTML5, WebGL, and WebAssembly titles, provide an underexploited data frontier for world models. Browser execution offers a standardized runtime loop, open asset delivery, and relatively uniform access to rendered observations and native input streams, permitting telemetry collection across many game types without modifying each engine.

However, large-scale browser telemetry immediately exposes a central obstacle that standard reinforcement-learning benchmarks largely avoid: input heterogeneity. In a unified multi-game dataset, raw player actions are recorded as browser-native events such as mouse coordinates, clicks, and keyboard presses, yet those inputs have different meanings across games. The same key can indicate forward movement in a first-person game, upward traversal in a platformer, or a map overlay

Algorithm 1 Weakly supervised latent-action transfer

1: Collect browser telemetry trajectories $\tau = (s_1, a_{\text{raw},1}, \ldots, s_T)$ from a diverse source-domain corpus.
2: Encode observations into latent states $h_t = e_\psi(s_t)$.
3: Jointly train the action encoder f_ϕ and world model g_θ.
4: Infer latent intent tokens with $z_t = f_\phi(s_{t-\tau:t}, a_{\text{raw},t-\tau:t})$, where $z_t \in \mathcal{Z}_a$.
5: Optimize g_θ to predict $\hat{h}_{t+1}$ from (h_t, z_t).
6: Transfer to a new game domain: freeze θ and fine-tune only ϕ on limited telemetry so novel raw controls align with $\mathcal{Z}_a$.
7: Evaluate transfer using next-state prediction quality and adaptation efficiency.

in a strategy interface. Training a shared world model directly on such mechanically inconsistent action streams asks the transition model to absorb mutually incompatible semantics.

This motivates our core thesis: a latent action space is not merely a convenience for cleaning observational noise, but a necessary abstraction for cross-game world modeling. Raw mechanical inputs do not exhibit semantic invariance across game engines. A learned latent action layer is needed to translate surface-level controls into transferable behavioral primitives so that downstream dynamics can be modeled in a domain-agnostic space. This perspective is consistent with the reinforcement-learning literature on temporal abstraction, options, and learned skills (Sutton et al., 1999; Bacon et al., 2016; Eysenbach et al., 2018), as well as discrete learned codebooks for compressing reusable structure (van den Oord et al., 2017). Our claim is about the need for a shared action abstraction, not about one privileged estimator for that abstraction.

To that end, this paper presents an architectural blueprint for cross-game world modeling from web gameplay telemetry. Our proposal treats raw browser actions as a source of weak supervision for a temporal action encoder. The encoder maps recent state history and raw inputs into a bounded discrete latent action vocabulary. A downstream world model then predicts future latent state representations conditioned on latent intent rather than domain-specific keybindings. The intended contribution is a transfer interface and benchmark methodology, rather than a finished universal world model.

2 Technical Framework: Semantically Aligned Latent Dynamics

To achieve cross-game generalization, we propose a decoupled architecture that separates raw game-specific mechanics from the prediction of shared environmental dynamics. The pipeline has three stages: telemetry ingestion, weakly supervised latent action inference, and latent-conditioned world-model prediction. The operational logic is summarized in Algorithm 1.

2.1 Problem Formulation and Ingestion Layer

Let a game domain D_i in a multi-game distribution $\mathcal{D}$ be modeled as an underlying Markov decision process. Browser-collected trajectories yield asymmetric sequences of observations

$$\tau = (s_1, a_{\text{raw},1}, s_2, a_{\text{raw},2}, \ldots, s_T)\,, \tag{1}$$

where $s_t \in \mathcal{S}$ denotes a high-dimensional visual observation and $a_{\text{raw},t} \in \mathcal{A}_{\text{raw},i}$ denotes a high-frequency mechanical action emitted by the browser runtime for domain D_i.

The difficulty is that the raw action spaces differ across games in both representation and meaning. Formally, for two domains D_j and D_k, we generally have $\mathcal{A}_{\text{raw},j} \neq \mathcal{A}_{\text{raw},k}$, and even when tokens appear similar, their induced semantics may differ. As a result, directly fitting a shared transition model of the form

$$P(s_{t+1} \mid s_t, a_{\text{raw},t}) \tag{2}$$

across $\mathcal{D}$ is poorly posed.

2.2 Weakly Supervised Action Abstraction

To resolve this mismatch, we introduce an action abstraction network f_ϕ with parameters ϕ. Rather than relying on fully unsupervised action discovery, we use raw mechanical inputs as weak structural supervision. The encoder consumes a short temporal window of states and associated raw actions and emits a latent action token,

$$z_t = f_\phi\left(s_{t-\tau:t}, a_{\mathrm{raw},t-\tau:t}\right), \qquad z_t \in \mathcal{Z}_a, \tag{3}$$

where $\mathcal{Z}_a = \{1, 2, \ldots, K\}$ is a bounded discrete vocabulary of learned behavioral primitives.

Raw action events provide temporal anchors that indicate when interventions occur, even if the observed controls are not semantically aligned across domains. The objective is therefore to preserve intervention structure while encouraging the encoder to collapse mechanically different but functionally equivalent controls onto the same latent token. In a concrete implementation, z_t could be realized either as a deterministic quantized code or as a stochastic categorical latent variable. The present paper intentionally does not commit to a specific variational estimator, because the central research question is whether a shared latent action interface improves cross-game transfer at all.

2.3 Latent-Conditioned Transition Dynamics

Rather than predicting pixels directly, we assume a learned state representation $h_t = e_\psi(s_t)$ and let the downstream sequence model g_θ predict future latent states in intent space:

$$\hat{h}_{t+1} = g_\theta(h_t, z_t). \tag{4}$$

This is closer in spirit to contemporary world-model practice (Ha & Schmidhuber, 2018; Hafner et al., 2023; Schrittwieser et al., 2019) and encourages g_θ to learn reusable gameplay dynamics rather than memorizing surface-level input conventions. If desired, g_θ may also consume a lightweight domain-context token, for example from short text or metadata describing the game, but such conditioning should remain optional and be carefully ablated because strong game-identity embeddings can undermine the transfer objective. Under transfer to a new domain D_{new}, the world-model parameters θ can remain fixed while only the encoder f_ϕ adapts to map the new game's browser-native controls onto the established latent vocabulary $\mathcal{Z}_a$.

3 Transfer Learning Setting and Validation Agenda

The central empirical question behind the framework is whether weakly supervised latent actions improve adaptation efficiency when moving across games with mismatched visual styles and control layouts. We therefore focus on the transfer setting the framework is meant to address and on the evidence that would ultimately validate or falsify its core claim.

3.1 Training Protocol

The proposed training procedure has two phases.

Source-domain pre-training. The action encoder f_ϕ and the transition model g_θ are jointly optimized on a large and diverse corpus drawn from a source domain family, such as side-scrolling platformers or browser-native puzzle games. The goal is to learn both a useful latent action vocabulary $\mathcal{Z}_a$ and dynamics that are conditioned on those abstracted intents.

Target-domain adaptation. The model is transferred to a distinct target domain with different control conventions and visual assets. During this phase, the world model g_θ is frozen and only the lightweight encoder f_ϕ is updated using a small amount of target-domain telemetry. Success is measured by whether the frozen dynamics model remains predictive once the new control scheme has been aligned to the previously learned intent space.

3.2 Validation Criteria

If instantiated experimentally, the framework would naturally be assessed along two axes.

Latent predictive accuracy. We measure the discrepancy between predicted and realized target-domain transitions, for example via mean squared error over learned observation embeddings. Lower prediction error under limited adaptation data would indicate that the latent action layer preserves transferable structure.

Few-shot adaptation curves. We compare the proposed framework against two baselines: a target-domain model trained from scratch, and a transfer model conditioned directly on raw actions without latent alignment. The key quantity of interest is how quickly each method recovers predictive performance as target-domain supervision increases.

Transfer success criterion. The proposal should only be regarded as successful if a frozen or nearly frozen world model g_θ can be reused across games after limited target-domain adaptation of the encoder f_ϕ. If performance gains disappear once the raw-action baseline receives the same supervision budget, then the latent interface has not provided meaningful cross-game structure.

4 Open Research Challenges

The framework depends on whether a cross-game latent-action interface can carry reusable gameplay semantics across the distribution shifts highlighted in game generalization benchmarks (Cobbe et al., 2019).

Vocabulary scale and discretization. Choosing the size K of the latent vocabulary $\mathcal{Z}_a$ involves a central tradeoff. If K is too small, semantically distinct intents collapse together; if it is too large, the latent space fragments and adaptation becomes data-inefficient.

Information leakage. The encoder must discard superficial domain-specific styling while preserving true mechanical intent. If latent tokens or optional domain-context embeddings encode too much game identity, the downstream world model will overfit to source-domain appearance rather than transferable dynamics.

Temporal mismatch. Different games operate at different effective action frequencies. Puzzle games may involve sparse discrete interventions, whereas action-heavy titles may require dense cursor and camera control. Learning a common tokenization and aggregation strategy across such horizons remains an unresolved challenge.

Out-of-vocabulary intents. A fixed codebook $\mathcal{Z}_a$ can only express behavioral primitives seen in the source corpus. Target games with genuinely novel intents must be projected onto existing tokens, capping transfer unless the source corpus is broad and vocabulary extensibility is treated as first-class.

5 Discussion and Conclusion

We have outlined a research agenda for cross-game world models trained from web-based gameplay telemetry. The main claim is that input heterogeneity across games is not just an engineering nuisance but a structural barrier to transfer.

Our proposed solution is a hierarchical latent-action framework in which raw browser inputs act as weak supervision for learning a shared discrete intent space.

References

Pierre-Luc Bacon, Jean Harb, and Doina Precup. The option-critic architecture. *CoRR*, abs/1609.05140, 2016. URL http://arxiv.org/abs/1609.05140.

Marc G. Bellemare, Yavar Naddaf, Joel Veness, and Michael Bowling. The arcade learning environment: An evaluation platform for general agents. *Journal of Artificial Intelligence Research*, 47:253–279, 2013. DOI: 10.1613/JAIR.3912. URL https://doi.org/10.1613/jair.3912.

Karl Cobbe, Christopher Hesse, Jacob Hilton, and John Schulman. Leveraging procedural generation to benchmark reinforcement learning. *CoRR*, abs/1912.01588, 2019. URL http://arxiv.org/abs/1912.01588.

Benjamin Eysenbach, Abhishek Gupta, Julian Ibarz, and Sergey Levine. Diversity is all you need: Learning skills without a reward function. *CoRR*, abs/1802.06070, 2018. URL http://arxiv.org/abs/1802.06070.

Linxi Fan, Guanzhi Wang, Yunfan Jiang, Ajay Mandlekar, Yuncong Yang, Haoyi Zhu, Andrew Tang, De-An Huang, Yuke Zhu, and Anima Anandkumar. MineDojo: Building open-ended embodied agents with internet-scale knowledge. *CoRR*, abs/2206.08853, 2022. DOI: 10.48550/ARXIV.2206.08853. URL https://doi.org/10.48550/arXiv.2206.08853.

William H. Guss, Cayden R. Codel, Katja Hofmann, Brandon Houghton, Noburu Kuno, Stephanie Milani, Sharada P. Mohanty, Diego Perez Liebana, Ruslan Salakhutdinov, Nicholay Topin, Manuela Veloso, and Phillip Wang. The MineRL competition on sample efficient reinforcement learning using human priors. *CoRR*, abs/1904.10079, 2019. URL http://arxiv.org/abs/1904.10079.

David Ha and J"urgen Schmidhuber. World models. *CoRR*, abs/1803.10122, 2018. URL http://arxiv.org/abs/1803.10122.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy P. Lillicrap. Mastering diverse domains through world models. *CoRR*, abs/2301.04104, 2023. DOI: 10.48550/ARXIV.2301.04104. URL https://doi.org/10.48550/arXiv.2301.04104.

Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, Timothy P. Lillicrap, and David Silver. Mastering atari, go, chess and shogi by planning with a learned model. *CoRR*, abs/1911.08265, 2019. URL http://arxiv.org/abs/1911.08265.

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. DOI: 10.1016/S0004-3702(99)00052-1.

A"aron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 6306–6315, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/7a98af17e63a0ac09ce2e96d03992fbc-Abstract.html.