Verses' AXIOM: Independent Validation Report

Prepared by:
Akshay Deshpande (Project Research Lead, Soothsayer),
Haritima Chauhan (Project Research Co-Lead, Soothsayer), and
Varun Agrawal (Research Assistant, Verses).

Contents

Introduction

AXIOM (Active eXpanding Inference with Object-centric Models) is designed as a general-purpose framework, with its capabilities tested across ten diverse 2D arcade-style environments from the Gameworld-10k challenge [1]. In these settings, agents learn to control characters or objects across a range of tasks. These environments are specifically designed to evaluate an agent's general control capabilities, testing its ability to learn efficiently from limited data, sustain reasoning and planning over time, and handle delayed or sparse rewards.

Our validation assessed the theoretical foundations of AXIOM, its empirical performance metrics, and comparison with DreamerV3 [2]. This evaluation considered the underlying mixture models and the Active Inference framework within the target environments, along with a review of key hyperparameters and design components. AXIOM's efficiency stems from its integration of Variational Inference, which approximates Bayesian posteriors using human-like priors, and Active Inference, a framework where agents minimize expected free energy—or surprise—to reduce epistemic uncertainty and support pragmatic behavior [3, 4, 5]. Additionally, the use of core priors within a fast structure-learning mixture-model architecture, combined with gradient-free computations and Bayesian simplification mechanisms, further enhances learning efficiency [6].

Methodology

Performance evaluation was conducted at the individual environment level across three primary dimensions: reward structure (both stepwise and cumulative rewards), sample efficiency (measured by the number of steps required to reach benchmark performance), and computational efficiency (quantified as the training time per step). All experiments were performed using a single NVIDIA A100 GPU per environment. Each environment was evaluated over ten independent trials using distinct random seeds to ensure robustness and reproducibility.

Results

Rewards

Figure 1 shows the 1,000-step moving average scores, where AXIOM outperformed DreamerV3 across most environments. We report average payoffs for both models that align with those presented in [1], and all values fall within one standard deviation of the mean. Table 1 summarizes the cumulative and maximum cumulative rewards achieved by AXIOM and DreamerV3. For AXIOM, differences in cumulative rewards are statistically indistinguishable at the 1% significance level, and maximum gameplay scores remained within the 99% margin of error, with the exception of a single outlier.

Notably, DreamerV3 outperformed AXIOM in the Cross environment and showed marginal gains in Drive and Impact, though improvements in these cases were within 1%. In contrast, AXIOM delivered consistently higher average and peak scores in environments such as Explode, Fruits, Gold, and Hunt, with less variability across seeds.

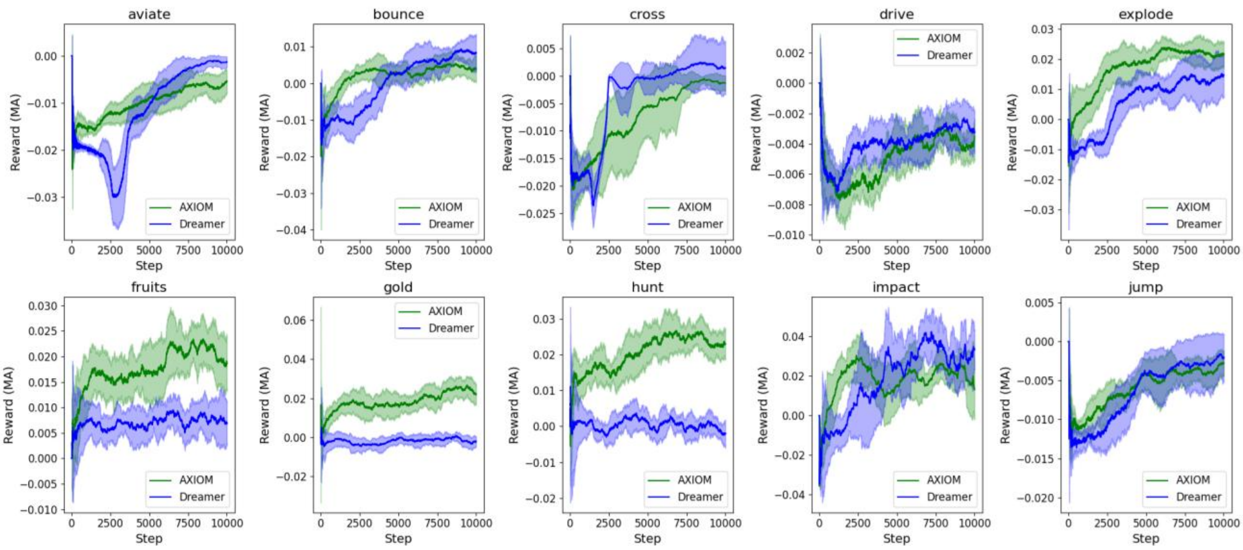

Figure 1: Moving Average (1K steps) Reward

Table 1: Cumulative Reward over 10K steps

Environment	AXIOM Cumulative Reward	AXIOM Max. Cumulative Reward	DreamerV3 Cumulative Reward	DreamerV3 Max. Cumulative Reward
Aviate	-95 (19)	-68	-108 (15)	-86
Bounce	27 (13)	48	10 (25)	45
Cross	-67 (35)	-32	-29 (18)	8
Drive	-49 (3)	-43	-39 (5)	-31
Explode	180 (30)	209	67 (28)	113
Fruits	181 (20)	209	70 (11)	88
Gold	190 (18)	212	-18 (11)	5
Hunt	209 (21)	242	3 (11)	28
Impact	188 (45)	250	204 (52)	297
Jump	-56 (9)	-40	-60 (20)	-36

Notes: Cumulative rewards are computed over 10,000 steps and averaged across 10 random seeds. Maximum reward values correspond to the highest-performing seed within each environment.

Training Time per Step

Table 2 reports the per-step training time for each environment. For AXIOM, the reported values may include additional runtime overhead and do not fully isolate the efficiency gains attributable to Bayesian model reduction. Nonetheless, the observed training times generally align with the combined model update and planning durations reported in [1], with the exception of the Cross environment.

AXIOM demonstrates significantly lower training times in nearly all environments, often operating at less than half the cost of DreamerV3. This computational efficiency is achieved despite using fewer model parameters.

Table 2: Training Time (ms/step)

Environment	AXIOM	DreamerV3
Aviate	477.06	993.1
Bounce	395.35	984.83
Cross	856.6	982.44
Drive	457.57	975.49
Explode	358.42	991.28
Fruits	433.76	980.27
Gold	537.11	1413.84
Hunt	592.3	980.95
Impact	304.66	981
Jump	334.38	982.11

Notes: Per-step time averaged across 10 seeds.

Sample Efficiency

Given the diverse nature of the environments, competency is not uniformly transferable across tasks. Therefore, in addition to evaluating overall baseline performance as defined by the VERSES team, we also assessed performance against task-specific benchmarks, defined as the mean reward over the final 1,000 steps. Table 3 presents the number of interactions required by AXIOM and DreamerV3 to reach both the per-task benchmark and the overall baseline threshold.

In our runs, AXIOM converged to its final performance in fewer than 2,950 steps in seven out of ten environments, with the fastest convergence observed at just 947 steps. In the remaining three environments, convergence was achieved within 5,800 steps. These results are averaged over 10 random seed trials. While DreamerV3 occasionally converged more quickly, this behavior was limited to a few seeds, and it often required more than 10,000 or 12,000 steps to reach comparable performance.

Conclusion

Our independent validation confirms that AXIOM represents a promising advancement, demonstrating strong performance across diverse 2D environments from the Gameworld-10k benchmark. It showed notable improvements in payoffs, sample efficiency, and computational speed compared to baseline DreamerV3. These gains were achieved using fewer model parameters, underscoring AXIOM's potential for scalable and resource-efficient learning.

Table 3: Interactions to achieve threshold performance

Environment	Benchmark Reward	Benchmark		Overall Baseline	
		AXIOM	DreamerV3	AXIOM	DreamerV3
Aviate	-0.0065	5613 (1483)	5594 [10]	7175 (1729)	6515 [10]
Bounce	0.0041	2659 (1794)	5269 [10]	675 (450)	2348 [10]
Cross	-0.0011	5767 (1572)	2475 [10]	4447 (2115)	2437 [10]
Drive	-0.0042	947 (1331)	784 [10]	2873 (2544)	1848 [10]
Explode	0.0213	2902 (1983)	6506 [8]	302 (405)	1127 [10]
Fruits	0.0198	2074 (2105)	2432 [4]	116 (81)	118 [10]
Gold	0.0245	1691 (1974)	27 [3]	61 (35)	184 [10]
Hunt	0.0226	1455 (1250)	28 [4]	98 (64)	255 [10]
Impact	0.0202	1389 (777)	2489 [10]	353 (318)	919 [10]
Jump	-0.0035	4525 (1694)	4922 [8]	4525 (1694)	4922 [8]
Average of steps over all environments (AXIOM)		2902 (1783)		2063 (2542)	

Notes: Columns (2) and (3) report the average steps required to reach the environment-specific benchmark. Columns (4) and (5) report the average steps required by the overall baseline following Verses' definition. Baseline competence is calculated as the overall average performance across all environments, defined as (Mean − Standard Deviation), yielding a value of −0.0034. Standard deviations (for AXIOM) and number of successful seeds (for DreamerV3) are reported in parentheses and square brackets, respectively. All results are averaged over 10 random seeds.

AXIOM converged to its final performance in fewer than 2,950 steps in most environments. While DreamerV3 occasionally converged more quickly, these instances were limited and often required over 10,000 steps to achieve comparable outcomes. In terms of rewards, AXIOM outperformed DreamerV3 in the majority of environments, particularly in high-variance and sparse-feedback scenarios such as Explode, Fruits, Gold, and Hunt. Moreover, AXIOM achieved this level of performance with fewer model parameters and substantially lower per-step training time, often operating at less than half the cost of DreamerV3. This efficiency reflects the strength of its underlying framework, combining Variational and Active Inference with structure-learning mixture models and Bayesian simplification.

AXIOM's strong performance and low resource requirements make it a strong candidate for mainstream applications. We recommend continued evaluation of how AXIOM's principles can be generalized and applied in more complex domains with real world.

References

[1] Heins, C., Van de Maele, T., Tschantz, A., Linander, H., Markovic, D., Salvatori, T., Pezzato, C., Catal, O., Wei, R., Koudahl, M., Perin, M., Friston, K., Verbelen, T., & Buckley, C. L. (2025). *AXIOM: Learning to play games in minutes with expanding object-centric models* (arXiv:2505.24784) [Preprint]. arXiv. https://arxiv.org/abs/2505.24784

[2] Hafner, D., Pasukonis, J., Ba, J., & Lillicrap, T. (2023). *Mastering diverse domains through world models* (arXiv:2301.04104) [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2301.04104

[3] Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: The free energy principle in mind, brain, and behavior*. MIT Press.

[4] Hoffman, M. D., Blei, D. M., Wang, C., & Paisley, J. (2013). Stochastic variational inference. *Journal of Machine Learning Research*, 14(1), 1303–1347.

[5] Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518), 859–877.

[6] Friston, K., Parr, T., & Zeidman, P. (2018). *Bayesian model reduction* (arXiv:1805.07092) [Preprint]. arXiv. https://arxiv.org/abs/1805.07092