ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Unpacking the Role of Intrinsic Motivation in Elastic Decision Transformers: A Post-Hoc Analysis of Embedding Geometry and Performance

Leonardo Guiducci[1], Antonio Rizzo[1], Giovanna Maria Dimitri[2]

1- DISPOC, Universitá di Siena, Siena, Italy
Via Roma 56, 53100, Siena, Italy

2- SPS, Universitá degli Studi di Milano, Milan, Italy
Via Conservatorio 7, 20122, Milano, Italy

Abstract. Elastic Decision Transformers (EDTs) augmented with intrinsic motivation exhibit improved performance in offline reinforcement learning, yet the cognitive processes driving these gains remain unclear. We present a systematic post-hoc explainability framework that examines how intrinsic motivation influences learned embeddings through statistical characterization of covariance structure, vector magnitudes, and orthogonality. Our findings show that distinct intrinsic-motivation variants induce qualitatively different representational organizations: EDT-SIL (state-based) produces significantly more compact embedding spaces than baseline EDT, whereas EDT-TIL (transformer output-based) increases representational orthogonality. We identify environment-dependent correlations between embedding metrics and performance across locomotion domains. The results indicate that intrinsic motivation acts as a representational prior that shapes embedding geometry in cognitively meaningful ways, yielding environment-specific structures that support improved decision-making beyond simple exploration bonuses.

1 Introduction and Background

Reinforcement learning has progressed beyond purely reactive policy optimization toward models with greater generalization and adaptability, especially in offline settings where agents must learn effective policies entirely from previously collected data without any interaction during training.[1, 2, 3]. Elastic Decision Transformers (EDTs) [4] have emerged as a promising architecture that unifies sequence modeling with decision-making by leveraging Transformer [5] architectures to capture long-range dependencies and enable flexible policy behaviors under uncertainty. EDTs enhance standard Decision Transformers [6] through dynamic history length adjustment, enabling effective trajectory stitching for improved offline RL performance. Intrinsic motivation mechanisms, inspired by cognitive science theories of curiosity and novelty-seeking behavior [7, 8], have been incorporated into RL to encourage exploration when extrinsic rewards are sparse or poorly aligned with long-term success. In our previous work [9] we showed that integrating intrinsic rewards into EDTs yields improved performance across offline RL benchmarks. However, while the performance benefits

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

are empirically established, the representational mechanisms underlying these improvements remain largely unexplored, yet crucial [10].

In this paper we aim to answer this research question, analysing embedding geometry and performance gains, and investigating whether intrinsic motivation shapes learned representations in EDTs. We introduced a systematic post-hoc explainability framework using statistical analysis of the embeddings, in particular evaluating the: covariance trace, L2 norm and cosine similarity, to examine the geometry of the embeddings, and also analysing their distribution in the PCA space. Our analysis reveals that intrinsic motivation goes beyond being a simple exploration mechanism, acting as a representational prior that creates environment-specific organizational structures. We analyze the two EDT variants presented in [9]: *EDT-SIL*, where intrinsic loss operates on embedded states promoting compactness, and *EDT-TIL*, where it operates on transformer outputs enhancing orthogonality.

The novelty of the paper is manifold: we present a post-hoc explainability framework for analyzing embedding geometry changes and develop quantitative correlations between embedding properties and task performance across environments. We provide an in-depth analysis of the distinct representational structures of EDT variants, demonstrating that EDT-SIL produces significantly more compact representations than baseline EDT.

2 Methods

We analyze Elastic Decision Transformers (an architecture that processes trajectories as sequences of (state, action, reward) tuples, proposed in [4]) enhanced with intrinsic motivation mechanisms, which we proposed in [9] Our analysis focuses on two intrinsically-motivated EDT variants that incorporate Random Network Distillation (RND) [11] modules as auxiliary loss functions.

2.1 Intrinsic Motivation Variants

We examine two EDT variants that differ in where the intrinsic signal operates: **1. EDT-SIL (State Input Loss)**: in this case the architecture computes intrinsic loss directly from the embedded state representations, allowing the intrinsic signal to influence the state embedding layer and potentially encourage more structured representations. **2. EDT-TIL (Transformer Input Loss)** operates on transformer output representations, enabling the intrinsic signal to shape both embedding and transformer layers for more coherent sequential representations.

Following the RND framework [11], the intrinsic loss is computed as $L_{\text{int}} = |f_{\text{pred}}(x;\theta_{\text{pred}}) - f_{\text{target}}(x;\theta_{\text{target}})|_2^2$, where f_{target} is a fixed randomly-initialized network and f_{pred} is a trainable predictor network; their squared distance serves as a novelty signal. Here x represents either embedded states (SIL) or transformer outputs (TIL). The total loss combines the standard EDT objective with this intrinsic component: $L_{\text{overall}} = L_{\text{EDT}} + L_{\text{int}}$, enabling intrinsic motivation to enhance representation learning without disrupting the primary task objective.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

2.2 Post-Hoc Explainability Framework

Our primary contribution is a framework for analyzing how intrinsic motivation shapes learned representations. We did this by examining geometric and statistical properties of the embedding spaces, and focusing on three key metrics that capture complementary aspects of representational structure: *covariance trace*, which measures how variance is distributed across embedding dimensions and reflects the overall spread of the learned space; *L2 norm*, which quantifies vector magnitude as a proxy for representational compactness; and *cosine similarity*, which assesses the angular separation between embeddings and thus captures representational orthogonality. Moreover we projected the embeddings in the latent Principal Component Analysis space (PCA) and analyzed their position in the PCA projected space (highlighting their convex hull compactness). To establish quantitative relationships between representational properties and task performance, we computed Pearson correlations between embedding metrics and normalized performance scores across multiple seeds, identifying the most predictive metric for each environment-model combination.

3 Experiments and Results

We evaluate intrinsic motivation mechanisms in EDT, focusing on performance improvements and underlying representational changes. Using the standard EDT architecture [4] as a baseline. We compared against EDT-SIL and EDT-TIL variants across 4 continuous control tasks from the D4RL benchmark [12]: Ant, HalfCheetah, Hopper, and Walker2d. For both variants, we tested RND modules with 1, 3, and 10 layers, finding that the 3-layer configuration consistently yields the best performance.

We evaluated models on both medium datasets (∼1M transitions with cleaner trajectories) and medium-replay datasets (∼2M transitions with noisy replay buffer data) (with 5 random seeds for initializations). Performance was evaluated using Human-Normalized Scores (HNS), providing consistent scaling across environments:

$$\text{HNS} = \frac{\text{score} - \text{score\_random}}{\text{score\_human} - \text{score\_random}} \tag{1}$$

We analyzed 20,000 state embeddings from EDT models using geometric metrics. Our analysis confirms that EDT-SIL produces significantly more compact representations - it has 10% lower variance, 5% shorter vectors, and occupies 32% less volume in embedding space. The compactness likely explains why EDT-SIL performs better in certain tasks: compact representations generalize better from offline data. State embeddings were collected during the evaluation, executing the best performing model for each environment-dataset combinantion extracting embeddings at each step over single episodes with maximum 1000 steps. The metrics of covariance trace, L2 norm and cosine similarity were computed for the embeddings (results presented are the average across three repetitions for ensuring robustness of the results).

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

3.1 Performance Results

Table 1 presents performance results across both medium and medium-replay datasets, where intrinsic motivation variants demonstrate environment-specific effectiveness patterns. On medium datasets, EDT-TIL achieved the best performance in 2 out of 4 environments (Walker2d: 73.50 vs 68.50; Hopper: 59.63 vs 57.49/59.31). In medium-replay we observe something different. EDT-SIL shows improved performance in Hopper (84.67 vs 81.56), while EDT-TIL shows robust performance in HalfCheetah (38.60 vs 37.32) and Walker2d (65.06 vs 62.25). Interestingly, the baseline EDT achieves the best performance in Ant on medium-replay (85.51), suggesting that this environment may be less prone to intrinsic motivation on noisier datasets. Overall, improvements are modest and task-specific, with overlapping standard deviations in several cases, suggesting that intrinsic motivation creates complementary representational advantages suited to different environmental dynamics rather than yielding uniform gains. The higher variance in Ant and Walker2d further suggests that the intrinsic loss introduces additional optimization variability, more pronounced in environments with sensitive training dynamics.

Table 1: Performance comparison on Medium and Medium-Replay datasets. Human-normalized scores (HNS) show mean ± standard deviation across 5 seeds. The best results per each environment are highlighted in bold.

Dataset/Model	Ant	HalfCheetah	Hopper	Walker2d
Medium				
EDT	88.84±3.61	42.30±0.14	57.49±3.81	68.50±2.03
EDT-SIL	**90.49±5.01**	**42.46±0.12**	59.31±6.16	69.44±4.46
EDT-TIL	89.01±5.83	42.18±0.34	**59.63±2.35**	**73.50±4.29**
Medium-Replay				
EDT	**85.51±5.06**	37.32±2.46	81.56±9.96	62.25±5.21
EDT-SIL	84.02±3.72	37.64±2.44	**84.67±4.80**	57.21±8.54
EDT-TIL	83.72±4.13	**38.60±1.28**	81.72±9.27	**65.06±3.81**

3.2 Embedding Analysis Results

The analysis performed highlights clear environment-specific relationships between embedding properties and performance. Ant is characterized from reduced covariance (-0.907), suggesting that lower overall variance supports control, whereas HalfCheetah performs better with increased covariance (+0.850). Hopper shows improved performance with higher cosine similarity (+0.658), while Walker2d strongly favors representational orthogonality (-0.950). These contrasting patterns imply that intrinsic motivation mechanisms shape representations in ways tuned to the specific demands of each environment, consistent with principles of adaptive organization observed in biological systems. Across models, the two intrinsic motivation variants produce distinct and complementary representational effects. EDT-SIL consistently yields more compact embed-

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

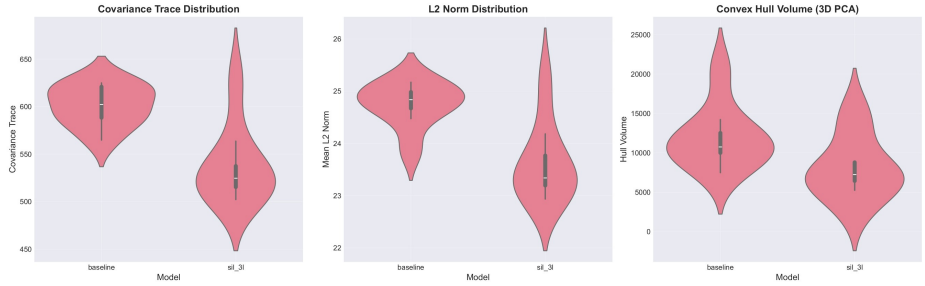

Fig. 1: Statistical validation of EDT-SIL compactness. Violin plots show distributions of (left) covariance trace, (center) L2 norm, and (right) convex hull volume across all environments and datasets (n=16 data points: 4 environments x 2 datasets x 2 models).

dings, reflected in lower covariance and vector magnitudes, whereas EDT-TIL enhances orthogonality by reducing cosine similarity in an environment-dependent manner. This division of representational roles mirrors hierarchical regulatory principles in biological neural systems, suggesting that intrinsic motivation may act as a representational prior that draws embedding geometry to support more effective learning. The broader variance in Figure 1 for the intrinsic variants reflects the optimization variability discussed in Section 3.1, confirming that the intrinsic loss affects both performance and representational geometry.

In Figure 1 we present the distribution of covariance trace, L2 norm and Convex Hull volume for the distribution of EDT baselines against EDT-SIL. The violin plots for all three metrics show a clear visual separation between the Baseline and EDT-SIL, with all independent measures consistently confirming the increased compactness of the EDT-SIL representations. This compactness of the EDT-SIL embeddings is present for all of the environments. In Figure 2 we show the distribution of the embeddings for the Medium/Hopper space, when the embeddings are projected into the PCA components space. As we can see the EDT-SIL cloud (the orange

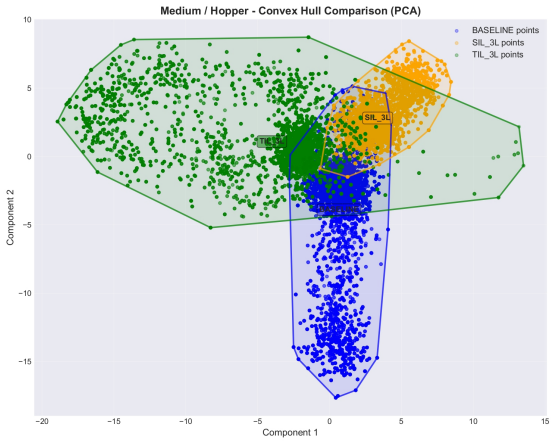

Fig. 2: Convex hull boundaries of state embeddings (2D PCA projection, Hopper medium). EDT-SIL (orange) occupies less volume than baseline (blue), confirming compact representations

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

one) shows significant more compactness rather than the other clusters (of EDT and EDT-TIL) confirming the hypothesis and the results observed previously.

4 Conclusions

We examine how incorporating intrinsic motivation into Elastic Decision Transformers (EDTs) alters their learned representations and how these changes relate to performance across continuous-control tasks. Both intrinsic-motivation variants, EDT-SIL and EDT-TIL, generally surpass the baseline EDT. Our post-hoc analysis shows that EDT-SIL produces more compact representations, while EDT-TIL increases representational orthogonality through lower cosine similarity an effect especially evident in Walker2d.

References

[1] Sergey Levine, Aviral Kumar, George Tucker, and Justin Fu. Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643*, 2020.

[2] Raghavendra P Kidambi, Aravind Rajeswaran, Praneeth Netrapalli, and Thorsten Joachims. Morel: Model-based offline reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

[3] Aviral Kumar, Aurick Zhou, George Tucker, and Sergey Levine. Conservative q-learning for offline reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

[4] Yueh-Hua Wu, Xiaolong Wang, and Masashi Hamaya. Elastic decision transformer. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[6] Lili Chen, Kevin Lu, Aravind Rajeswaran, Kimin Lee, Aditya Grover, Michael Laskin, Pieter Abbeel, Aravind Srinivas, and Igor Mordatch. Decision transformer: Reinforcement learning via sequence modeling. In *Advances in Neural Information Processing Systems*, volume 34, pages 15084–15097, 2021.

[7] Pierre-Yves Oudeyer and Frederic Kaplan. What is intrinsic motivation? a typology of computational approaches. *Frontiers in neurorobotics*, 1:108, 2007.

[8] Deepak Pathak, Pulkit Agrawal, Alexei A Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017.

[9] Leonardo Guiducci, Giovanna Maria Dimitri, Giulia Palma, and Antonio Rizzo. Introducing intrinsic motivation in elastic decision transformers. In *ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges (Belgium) and online event, 23–25 April 2025. i6doc.com publ.

[10] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[11] Yuri Burda, Harrison Edwards, Amos J. Storkey, and Oleg Klimov. Exploration by random network distillation. *CoRR*, abs/1810.12894, 2018.

[12] Justin Fu, Aviral Kumar, Ofir Nachum, George Tucker, and Sergey Levine. D4rl: Datasets for deep data-driven reinforcement learning. *arXiv preprint arXiv:2004.07219*, 2020.