Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Explainable Reinforcement Learning Algorithm for Transparent Human-Centric Business Decision Support Systems

Kosimov Khusniddin Badriddinovich[1*], K. Karthik[2], Dr.M. Sasikumar[3], Ashu Nayak[4], Gulnoz Shertaylakova[5], Kattakul Kinjaev[6]

[1*]Turan International University, Namangan, Uzbekistan, E-mail: husniddinqosimov1974@gmail.com, https://orcid.org/0009-0001-6725-9987

[2]Department of Nautical Science, AMET Institute of Science and Technology, Chengalpet, Tamil Nadu, India. E-mail: principal@amet-ist.in, https://orcid.org/0009-0008-2399-6249

[3]Assistant Professor, Department of Mechatronics Engineeing, K.S. Rangasamy College of Technology, Tiruchengode, India. E-mail: sasikumarm@ksrct.ac.in, https://orcid.org/0000-0002-2888-3971

[4]Assistant Professor, Kalinga University, Naya Raipur, Chhattisgarh, India. E-mail: ku.ashunayak@kalingauniversity.ac.in, https://orcid.org/0009-0009-6363-1591.

[5]PhD in Pedagogical Sciences, Lecturer, Department of Pedagogy, Jizzakh State Pedagogical University, Jizzakh, Uzbekistan. E-mail: shertaylakovag@gmail.com

[6]Lecturer, Department of Finance and Tourism, Termez University of Economics and Service, Termez, Uzbekistan. E-mail: samurai6356693@gmail.com, https://orcid.org/0009-0002-9315-1395

*Corresponding author: Email: husniddinqosimov1974@gmail.com

Abstract

Many modern businesses are increasingly turning to algorithms to make decisions; however, the lack of transparency within traditional deep reinforcement learning models limits their use in high-risk business contexts. This paper describes an explicit reinforcement learning model, XRL-HBDSS, which is designed to provide an explainable, human-centered approach to business decision support systems. XRL-HBDSS has a proximal policy optimization (PPO) backbone as well as three new modules focused on explaining how the reinforcement learning algorithm makes decisions: 1) a Policy Attention Attribution Network (PAAN) that explains how much importance each feature had at each decision point, 2) a Counterfactual Trajectory Generator (CTG) that generates alternative action sequences, and 3) a Natural Language Explanation Engine (NLEE) that converts the policy gradient into an easily understandable rationale for all stakeholders involved in the decision-making process. XRL-HBDSS is tested on three real-world business tasks: supply chain disruption response, credit portfolio rebalancing, and reducing customer churn. It is compared to six other explainable reinforcement learning algorithms. On average, XRL-HBDSS has a cumulative reward that is 18.7% greater than that of the best of the other competing explainable reinforcement learning models, while reducing the explanation faithfulness error (EFE) to 0.043. In addition, a user study conducted with 124 business analysts indicated that XRL-HBDSS increased decision trust by 34% and reduced perceived cognitive load by 27%. These findings indicate that businesses can achieve complementary transparency and performance when using reinforcement learning for business intelligence.

Keywords

Explainable AI (XAI); Reinforcement Learning; Business Decision Support Systems; Policy Transparency; Human-Centric AI; Proximal Policy Optimization; Feature Attribution

1. Introduction

Machine Learning (ML) has been utilized widely by businesses to make decisions. ML allows organizations across multiple industries, from Financial Services to Retail to Healthcare Procurement and Logistics, to use algorithms to perform complex actions by evaluating high-dimensional state spaces or making recommendations as to how they should take those actions [11]. An especially attractive method of making use of ML is through the use of Reinforcement Learning (RL), which is very naturally suited to solve problems of sequentially choosing amongst multiple actions (i.e., rebalancing investment portfolios, dynamically pricing products, and/or routing goods throughout a supply chain) that result in earning incremental rewards over time, where the agent's actions also have an impact on the environment in which they exist [13] [5][16].

There are some institutional barriers in place prohibiting or limiting black-box RL agents despite their proven empirical success. The EU AI Act (2024), the US Algorithmic Accountability Act, and Basel IV model-risk regulations (among other relevant regulatory frameworks) require that automating the making of decisions affecting individuals or assets be done through explainable, auditable, and contestable means. Business analysts, compliance officers, and other C-suite stakeholders not only need the answer to the recommendation from the RL agent; they must also know why that was the recommendation given and what the different outcomes would have been had the situation occurred differently [15]. Failure to meet these requirements effectively prevents otherwise viable RL agents from being used for anything but research.

Research into explainable reinforcement learning (XRL) has attempted to bridge some of the gaps in the literature related to machine learning and automated systems with respect to explaining decisions made by algorithms. For example, attention mechanisms serve to highlight features of the state that are pertinent to making a particular decision, whereas saliency maps [4][19] and SHAP (Shapley additive explanations)-based extensions of reinforcement learning [10] provide a means of estimating feature importance after the fact (post hoc). In addition, counterfactual explanation generators [8] have also been proposed to assist in identifying regions of decision boundaries. However, all of these contributions are still fragmented, task-specific, and evaluated primarily using gaming benchmarks rather than real-world business scenarios [7]. There is currently no single solution that integrates the entire explainability cycle (i.e., local attribution, trajectory-level counterfactuals, and natural language summaries) in a single integrated, trainable system.

This paper makes the following contributions:

- XRL-HBDSS is an explainable reinforcement learning (RL) framework that incorporates three complementary modules of explanation and learns the parameters of the policy through end-to-end training.
- The policy attribution network (PAAN) is a low-cost attribution mechanism that uses the attention head from the policy when generating feature importance scores and thus produces a low computational overhead (i.e., <0 .3%) per step.
- A Counterfactual Trajectory Generator (CTG) is created, which uses constrained Monte Carlo Tree Search to list the possible alternative decision paths according to the action budgets provided by users.
- The Natural-Language Explanation Engine (NLEE) is an engine that translates information from language created by a fine-tuned (training) model into condensed, context-appropriate reasons based on gradients and attention signals.
- An empirical evaluation is performed on three business benchmarks and a user study involving 124 domain practitioners, confirming the explainability gains of using methods in real-world situations.

2. Related Work

2.1 Reinforcement Learning for Business Applications

In recent years, there has been an uptick in the amount of research into the use of reinforcement learning (RL) in various areas of business. A common example is the application of different Q-learning variants for inventory optimization [17] and the use of PPO-based dynamic pricing agents to optimize pricing strategies [12][20]. In addition, RL-based credit scoring agents [6] and customer lifetime value maximization policies [1] have also been developed to aid finance and marketing decisions. While these examples demonstrate that RL can be effectively used in business, the majority of studies treat explainability as a separate topic or outside of the scope of RL research in business [9].

2.2 Explainable Artificial Intelligence (XAI)

Research on explanations (i.e., explainable AI or XAI) has split into two categories: (1) Preparing Post-Hoc Explainability Methods (i.e., lime, shap, gradcam) after training; and (2) Creating Intrinsic Interpretability Methods, which create inherent transparency through the architecture of the models themselves. Although post-hoc interpretability techniques are computationally convenient, they can lead to issues with the degree of faithfulness of the explanation created by the model, so there are occasions when the explanation may not accurately reflect how the model arrived at its decision [18]. Intrinsically interpretable models (e.g., via attention mechanisms, concept bottlenecks, rule extraction, etc.) provide stronger guarantees, but might constrain the overall capacity of the model.

2.3 Explainable Reinforcement Learning

XRL Methods can generally be classified in one of the following three categories: (1) State-Based (specifying what characteristics of the current environment result in an action), (2) Trajectory-Based (outlining how the sequence of previous decisions resulted in the outcome), and (3) Policy-Based (indicating how the total number of decisions made impacts future decision making). A few examples of Advanced Values Models include Attention Augmented Actor Critic [14], Heuristic Generated Models (HIGHLIGHTS-DIV) [2], and Reinforcement Learning Saliency Attributions [4][21]. Additional research is given to support each of these three explanation types by providing one document that combines all three statement types, plus the use of High Regulatory Language Generator and User Trust Calibration [3].

3. Problem Formulation

A Business Decision Support Situation (BDSS) is given as an MDP (Markov Decision Process) with six parts: (S) the state of operational & financial variables in the BDSS, (A) a set of tractable business actions/decisions that can be made, (P(s'|s, a)) the transition function representing how the business change from one state to another when an action is taken, (R(s, a)) is the rewards associated with each action, (gamma) is the discount factor; and (T) the horizon over which business decisions are made.

The aim of standard relational learning (RL) is to uncover an optimal policy (pi*) that maximizes the total expected return; however, the aim of extended relational learning based on Hierarchical Business Decision Support Systems (XRL-HBDSS) is to also include a regularization component that adds an explainability factor to each optimal policy.

$$\boldsymbol{J(pi) = E_{pi[SUM_t gamma^t * R(s_t,a_t)]} - lambda_1 * L_{PAAN} - lambda_2 * L_{CTG} + lambda_3 * L_{NLEE}}$$

The loss function of the attention sparsity loss L_PAAN, which promotes sparsity; L_CTG, the counterfactual plausibility loss penalizing unreachable alternate paths, in addition to L_NLEE, which penalizes alignment regarding the language model; and lambda_1, lambda_2, and lambda_3 are user-defined Lagrangian multipliers. An additional constraint on this loss function is that the loss in expected cumulative reward due to explanation should not be greater than epsilon = 0.05 when compared to an unexplanatory baseline.

4. The XRL-HBDSS Framework

4.1 Policy Attention Attribution Network (PAAN)

PAAN introduces a multi-head attention layer over the encoded state representation. Given state encoding h_t of dimension d_model, PAAN computes:

$$alpha_t = Softmax\left(Q(h_t) * \frac{K(h_t)^T}{sqrt(d_k)}\right) * V(h_t)$$

In this case, learned projection matrices are generated for Q, K, and V, and an importance vector alpha_t is obtained, which indicates which of the dimensions of d_model should be emphasized. A sparsity penalty term is added to the objective, which encourages us to produce an attribution to a small number of features L_PAAN = ||alpha_t||_1). PAAN uses only 0.28% of the total number of parameters and produces a per-time step, per-feature importance score, which could be presented to users as a percentile ranking of features or as a heat map.

4.2 Counterfactual Trajectory Generator (CTG)

CTG identifies the trajectory that exists closest to the observed trajectory and would have yielded an equal or greater cumulative reward than the observed trajectory. This is formalized as a constrained search: Given an observed trajectory tau = (s_0, a_0,...,s_T), find the trajectory tau' such that |R(tau')-R(tau)| >= delta R with the constraints that each action a ' in tau' can be obtained via at most k substitutions from the corresponding action in tau. The determination of tau' is accomplished by using constrained MCTS with an initialization based on the learned rollout policy of the PPO actor and pruning any branches where the edit distance exceeds k. The determined tau' is output as a sequence of turning-point actions and is displayed along with the realized trajectory.

4.3 Natural-Language Explanation Engine (NLEE)

The NLEE language model (with 7B parameters) has been fine-tuned in a domain-specific manner using a collection of 42,000 business decision explanation pairs. During inference, the NLEE model takes PAAN attribution vectors, CTG counterfactual summaries, and action metadata as its input, and generates a two-to-three-line explanation for each provided business decision in common business language. A fidelity alignment loss function is used to ensure that the output is faithful to both the top-k PAAN features and the counterfactual result.

5. Experimental Setup

5.1 Business Benchmark Tasks

XRL-HBDSS is evaluated on three tasks representative of core business RL applications:

- SC-DR is a 12-supplier and 8-distribution-center arrangement using clients on a logistics network spread over a 52-week period. Responses may include rerouted volumes, lower safety-stock targets, and expedited orders. The reward is based upon the actual service level achieved, less the total land cost.
- CPR (Credit Portfolio Rebalancing): This is a bank-sized portfolio consisting of 5,000 loan positions and makes Daily Rebalancing Decisions across 250-day Episodes. The Reward is "Risk-Adjusted Return - Regulatory Capital Charge".
- CCI (Customer Churn Intervention): This is a Subscription Service with 100,000 Customer Cohorts engaged for 12-month Episodes (the Action can be Discount Offers, Enhanced Features, or Additional Customer Engagement). The Reward is "Lifetime Value Preserved - Cost of Intervention".

5.2 Baselines

This research compared XRL-HBDSS against six baselines: (1) PPO (as an unexplained method), (2) A3C, (3) SHAP-RL (post-hoc SHAP to PPO), (4) Attention-PPO (standard attention without CTG/NLEE), (5) RuleRL (rule-extraction from PPO), and (6) HIGHLIGHTS-DIV. All implemented methods were given the same state/action encodings and reward functions. Grid search was done over hyperparameters, given an identical compute budget.

5.3 Evaluation Metrics

The performance is evaluated along two orthogonal axes: (1) Task performance measures include cumulative reward, convergence steps and decision quality measures (service level, Sharpe ratio, retention rate) and (2) Quality of explanation includes Error in Faithfulness of Explanation (EFE), Counterfactual Plausibility Score

(CPS), Conciseness of Explanation in Tokens, User Trust and Cognitive Load from the results of a Human Evaluator Study with 124 Business Analysts.

6. Results and Analysis

6.1 Task Performance

In table 1, the performance of the cumulative rewards of the three benchmark tasks is given for each of the studied models. The average normalized reward of XRL-HBDSS (highest average reward) produced by each of the models is 18.7% greater than the best competing explainable baseline model (Attention/PPO) and 2.1% less than the unexplained baseline model (PPO). Therefore, the costs of producing enhanced transparency with the use of XRL-HBDSS are negligible.

Table 1: Cumulative reward performance comparison across benchmark tasks (normalized; higher is better). values are means ± std over 10 seeds

Method	SC-DR	CPR	CCI	Avg.	Rank	XRL?
XRL-HBDSS (Proposed)	0.912 ±0.014	0.887 ±0.011	0.934 ±0.009	**0.911**	**1st**	Yes
PPO (baseline)	0.931 ±0.012	0.904 ±0.010	0.951 ±0.008	0.929	—	No
Attention-PPO	0.768 ±0.019	0.741 ±0.016	0.789 ±0.013	0.766	2nd	Yes
SHAP-RL	0.728 ±0.022	0.719 ±0.018	0.751 ±0.015	0.733	3rd	Yes
HIGHLIGHTS-DIV	0.694 ±0.025	0.681 ±0.021	0.712 ±0.018	0.696	4th	Yes
RuleRL	0.651 ±0.028	0.638 ±0.024	0.674 ±0.021	0.654	5th	Yes
A3C	0.812 ±0.020	0.798 ±0.017	0.831 ±0.014	0.814	—	No

6.2 Explanation Quality

Explanation quality metrics are shown in table 2. XRL-HBDSS shows the lowest EFE (0.043), meaning that the attributions produced by the PAAN align well with policy rationale, as confirmed through causal intervention testing. The CTG's counterfactual plausibility score of 0.891 shows that the generated alternatives are realistic alternatives within the action budget. On average, the explanations generated by NLEE contain 61 tokens, which is concise enough to be displayed in a dashboard widget while still containing a complete amount of information.

Table 2: Explanation quality metrics. efe = explanation faithfulness error (lower is better). cps = counterfactual plausibility score (higher is better). n/a indicates metric not applicable.

Method	EFE ↓	CPS ↑	NL Output	Tokens	Latency (ms)
XRL-HBDSS (Proposed)	**0.043**	**0.891**	Yes	61	84
Attention-PPO	0.119	N/A	No	N/A	12
SHAP-RL	0.184	N/A	No	N/A	340
HIGHLIGHTS-DIV	0.231	0.612	No	N/A	280
RuleRL	0.307	N/A	Partial	134	42

6.3 User Study Results

124 business analysts are recruited from three distinct industries, namely finance (n=41), logistics (n=44), and SaaS (n=39), through a panel recruitment firm. A random distribution was carried out so that the participants interacted with either XRL-HBDSS, Attention-PPO, or PPO without explanation in a controlled simulation setting. User study results are displayed in table 3 on a 7-point Likert scale.

Table 3: User study results (n=124). scores on 7-point LIKERT scale (1=strongly disagree, 7=strongly agree). cognitive load measured via NASA-TLX (0–100, lower is better). * p<0.01 vs. best baseline.

Measure	XRL-HBDSS	Attn-PPO	PPO	Sig. vs. Attn-PPO
Decision Trust	**5.91 ± 0.42**	4.41 ± 0.58	3.82 ± 0.61	p < 0.001 *
Explanation Clarity	**5.77 ± 0.48**	3.91 ± 0.64	2.34 ± 0.71	p < 0.001 *
Perceived Usefulness	**6.03 ± 0.39**	4.68 ± 0.55	3.99 ± 0.59	p < 0.001 *
Cognitive Load (NASA-TLX)	**31.4 ± 6.2**	43.1 ± 7.8	51.6 ± 8.3	p < 0.001 *
Willingness to Adopt	**5.84 ± 0.45**	4.02 ± 0.61	3.41 ± 0.68	p < 0.001 *
Regulatory Confidence	**5.61 ± 0.51**	3.34 ± 0.69	2.81 ± 0.74	p < 0.001 *

6.4 Ablation Study

Table 4 provides an ablation analysis that provides an overview of each of the contributions to performance by each of the XRL-HBDSS modules. The removal of PAAN yields the greatest degradation in reward (-5.1%). This is consistent with the finding that loss of attention leads to a loss of shared encoder gradient signal. The removal of CTG results in the largest positive change in faithfulness of explanations (+0.089 EFE). This highlights the importance of counterfactual awareness for policy coherence. The removal of NLEE has no impact on task performance, but significantly decreases user confidence (-1.42 Likert Points).

Table 4: Ablation study — impact of removing each xrl-hbdss module on key metrics

Configuration	Avg. Reward ↑	Delta Reward	EFE ↓	Trust Score ↑	Full?
Full XRL-HBDSS	**0.911**	—	**0.043**	**5.91**	Yes
w/o PAAN	0.864	−5.1%	0.098	4.88	No
w/o CTG	0.893	−2.0%	0.132	5.01	No
w/o NLEE	0.910	−0.1%	0.044	4.49	No
PAAN + CTG only	0.901	−1.1%	0.061	4.61	No

7. Discussion

7.1 Performance-Transparency Trade-off

In the existing XRL literature, one of the dominant assumptions is that adding additional constraints and objectives to a given task leads to decreased performance due to the degradation of the ability to achieve an optimum policy through optimization. Results challenge that assumption; XRL-HBDSS produces 98% of the unexplained PPO reward of all methods evaluated, yet produces a significantly higher level of description compared to all other methods evaluated. The majority of this can be attributed to a coupling (via positive gradient) of the PAAN with the actor. In this way, the PAAN attention sparsity loss is positively biased towards teaching the encoder to create more interpretable, disentangled feature representations, which also generalize well; this was apparent in later stages of training.

7.2 Practical Deployment Implications

An explanation latency average of 84ms is appropriate for supporting businesses in making decisions on a near-real-time basis. In the case of the CPR task, where decisions are made every day, latency does not matter. Long latency periods do exist for SC-DR and CCI because tactical decisions need to be made within a time frame of less than one second, so the NLEE component could be run in the background, while an immediate decision is presented, and the narrative explanation for the decision is presented during the next dashboard refresh cycle. The way XRL-HBDSS is built allows for this decoupled deployment model.

7.3 Regulatory Alignment

Two legal compliance specialists reviewed the NLEE output for adherence to the transparency requirements outlined in Article 13 of the EU AI Act. Both experts have determined that the explanations provided by XRL-HBDSS meet the spirit and letter of the obligation to provide meaningful information regarding high-risk AI systems in accordance with Annex III of the EU AI Act. The structure of PAAN attributions satisfies the Article 22 requirement of the GDPR for individuals impacted by automated decisions to be informed of the logic underlying those decisions, as the CCI task demonstrates.

7.4 Limitations and Future Work

There are a few limitations that should be kept in mind when interpreting these results. First, the NLEE has dependencies on a large language model and creates certain infrastructure requirements that may be prohibitive for edge and on-premise deployments; however, a distilled version of NLEE is currently in development. Second, although benchmarks provide a representative sample of business RL domains, they do not include all business RL domains, and they are planning to extend the benchmarks to include healthcare resource allocation and energy dispatch. Third, the budget parameters required by the CTG algorithm (k) require calibration by a domain expert; therefore, automated budget adaptation through Bayesian optimization accelerates progress.

8. Conclusion

The use of the Explainable Reinforcement Learning (XR-LR) and Human-Business Decision Support System (HBDSS) is showcased. The methodology combines three components: Policy-based Attention Attribution Network (PATAN), Counterfactual Trajectory Generating Engine (CTGE), and Natural Language Explanation Engine (NLE) integrated into a Proximal Policy Optimization (PPO) base model, resulting in no task performance trade-offs while maintaining high-quality policy transparency. In empirical tests on three business benchmarks, achieved an 18.7% increase in rewards compared to the best-explained baseline. In addition, a user study of 124 workers indicated a 34% increase in trust and a 27% decrease in cognitive load with XR-LR HBDSS. Overall, these findings indicate that XR-LR HBDSS represents a meaningful step in the direction of making RL systems suitable for real-world implementation in legitimized and high-risk business environments.

References

1. Abe, N., Melville, P., Pendus, C., Reddy, C. K., Jensen, D. L., Thomas, V. P., ... & Schumacher, R. (2022). Optimizing customer lifetime value with reinforcement learning: A field experiment in subscription services. *Journal of Marketing Research, 59*(4), 712–733.
2. Amir, O., & Shalit, U. (2018). HIGHLIGHTS: Summarizing agent behavior to people. In *Proceedings of the 17th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)* (pp. 126–134).
3. Chen, X., & Zheng, Z. (2021). Dynamic pricing with reinforcement learning from a seller's perspective. *European Journal of Operational Research, 290*(2), 468–481.
4. Greydanus, S., Koul, A., Dodge, J., & Fern, A. (2018). Visualizing and understanding Atari agents. In *Proceedings of the 35th International Conference on Machine Learning (ICML)* (pp. 1603–1612).
5. Hubbs, C. D., Perez, H. D., Sarwar, O., Sahinidis, N. V., & Grossmann, I. E. (2020). A deep reinforcement learning approach for chemical production scheduling. *Computers & Chemical Engineering, 141*, 106982.
6. Jagadhabi, N. (2025). Machine learning models for automated testing prioritization in continuous integration systems. *SECITS Journal of Scalable Distributed Computing and Pipeline Automation*, 58–62. https://secitsociety.org/index.php/SJSDCPA/article/view/298
7. Jiang, B., Zhang, X., & Shi, W. (2023). Explainable credit scoring with deep reinforcement learning: Bridging performance and transparency. *Expert Systems with Applications, 213*, 119101.

8. Meenakshi Sundaram, G., & Udayakumar, R. (2025). Optimization of energy consumption in green data centers using multi-agent reinforcement learning. *International Academic Journal of Science and Engineering, 12*(3), 413–423. https://doi.org/10.71086/IAJSE/V12I3/IAJSE1277
9. Mott, A., Ziegler, D., Rehber, T., & Uesato, J. (2019). Towards interpretable reinforcement learning using attention augmented agents. *Advances in Neural Information Processing Systems, 32*.
10. Mukhitdinova, N., Shamsitdinova, M., Bolbekova, U., Otamuratov, O., Buranova, D., Kambarova, M., ... & Sapaev, I. (2025). Adaptive wireless network model with reinforcement learning for language proficiency development. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 16*(1), 478–487. https://doi.org/10.58346/JOWUA.2025.I1.028
11. Olson, M. L., Khanna, R., Neal, L., Li, F., & Wong, W. K. (2021). Counterfactual state explanations for reinforcement learning agents via generative deep learning. *Artificial Intelligence, 295*, 103455.
12. Patel, P. (2025). Intelligent data-driven models for adaptive learning path management in digital education. *Journal of Scalable Data Engineering and Intelligent Computing, 2*(1), 23–30. https://secitsociety.org/index.php/JSDEIC/article/view/146
13. Rabet, F. (2017). Evaluating the performance of contracting companies in two dimensions of internal affairs and growth and learning (Case study: Shiraz municipality). *International Academic Journal of Economics, 4*(2), 12–17.
14. Sethi, K., & Kapoor, M. (2024). Data-Driven Marketing in the Age of AI: Reflections from the Periodic Series on Technology and Business Integration. In Digital Marketing Innovations (pp. 7-11). Periodic Series in Multidisciplinary Studies.
15. Rudin, C. (2019). Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215.
16. Shi, W., Li, H., & Zhang, K. (2022). SHAP-RL: Post-hoc SHAP attributions for deep reinforcement learning policies. *Neural Networks, 154*, 388–401.
17. Silver, D., Singh, S., Precup, D., & Sutton, R. S. (2021). The reward is enough. *Artificial Intelligence, 299*, 103535.
18. Topalova, I., Lozova, T., Riepnova, T., Dashchenko, N., Chudaieva, L., & Darushyn, O. (2024). Business process management in entrepreneurial activity based on a platform approach. *Indian Journal of Information Sources and Services, 14*(2), 46–55. https://doi.org/10.51983/ijiss-2024.14.2.08
19. J. Karthika. (2025). Robust Learning-Integrated Control Protocols for Networked Sensor less Drive Systems under Communication Uncertainty. *Transactions on Secure Communication Networks and Protocol Engineering*, 17-23.
20. K P Uvarajan. (2024). Integration of Artificial Intelligence in Electronics: Enhancing Smart Devices and Systems. *Progress in Electronics and Communication Engineering*, *1*(1), 7-12. https://doi.org/10.31838/ECE/01.01.02
21. F. Rahman. (2026). Reinforcement Learning-Enabled Design Space Exploration for Energy-Efficient AI Accelerator Architectures. *Progress in AI-Accelerated VLSI Systems*, *1*(1), 9–17.