International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Adversarial Reinforcement Learning for Multi-Agent Pursuit - Evasion and Decision Systems

Saai Charan R
Department of Artificial Intelligence and Data Science
Rajalakshmi Institute of Technology Chennai, India

Rakshana P B
Department of Artificial Intelligence and Data Science
Rajalakshmi Institute of Technology Chennai, India

Fouzia Sulthana K
Department of Artificial Intelligence and Data Science
Rajalakshmi Institute of Technology Chennai, India

Sandmanleo E
Department of Artificial Intelligence and Data Science
Rajalakshmi Institute of Technology Chennai, India

Abstract **-** Pursuit-evasion environments create a structured abstraction that can facilitate the study of adversarial interaction, coordination, and survival within a multi-agent environment. Traditional Samurai-Assassin environments use deterministic heuristics and scoring functions, which restrict flexibility within a dynamic environment. This paper introduces a novel adversarial reinforcement learning-based Agent-Based Modelling (ABM) that simultaneously optimizes Assassin capture coordination and survival-based Samurai evasion strategies. The Assassin agents use reinforcement learning within a grid-based environment with obstacles and random initialization. The Samurai agent uses a survival-based strategy that maximizes its probability of survival within the environment.

Interpretable machine learning is also introduced, which includes the use of Gradient Boosting and SHAP-based analysis, to evaluate the performance of the simulated environment, along with the emergent behaviour of the agents within the environment. The paper also includes extensive experimentation that shows the stability of policy convergence, improved coordination efficiency, and survival-based agent behaviour.

This paper also introduces a novel concept that extends the adversarial reinforcement learning-based architecture within a financial decision environment, specifically within a non-stationary adversarial environment that represents the equity markets.

Keywords—*Agent-Based Modeling, Reinforcement Learning, Multi-Agent Systems, Pursuit–Evasion, Adversarial Learning, Explainable AI, Financial Decision Systems*

I. INTRODUCTION

Agent-Based Modeling (ABM) provides a flexible computational paradigm for modeling complex systems of interacting agents. Pursuit-evasion games, in which predator agents seek to apprehend an evader within a given environment, represent basic models of agent-based systems, allowing analysts to study coordination, competition, and adaptation in a variety of contexts.

The traditional Samurai-Assassin model uses deterministic decision rules based on mathematical scoring systems. These systems, although useful in analytical contexts, represent overly constrained agent systems that cannot respond effectively to unseen spatial configurations or adversarial strategies. These systems are inadequate because they do not address the uncertainty present in real-world systems, which often involve adaptive robotics or cybersecurity systems.
Reinforcement Learning (RL) provides a solution to these issues because it allows agents to improve their strategies through interaction with their environment. By incorporating RL into ABM systems, analysts can model emergent coordination and survival strategies. However, it is not enough to simply measure agent performance; it is equally important to understand why agents succeed or fail.
The current study presents a new adversarial RL-based Samurai-Assassin model, in which coordination and survival strategies are balanced. Additionally, this study presents a discussion on the extension of this adversarial model into financial decision-making systems.

II. LITERATURE SURVEY

In classical pursuit-evasion models, deterministic models of behaviour and capture probabilities are used. The models are easy to interpret but do not offer learning. The previous implementation of the Samurai Assassin model used skill measures, influence, and probabilistic scoring.

For navigation, Q-learning and Deep Q-Networks (DQN) have shown promising results. The application of Multi-Agent Reinforcement Learning has shown that cooperative actions such as flanking and funnelling can occur without coordinated protocols.

However, there is a problem with interpretability. To address this, recent models used machine learning interpretability methods such as Gradient Boosting and SHAP values.

Currently, there is a lack of literature that incorporates ABM, Adversarial RL, and interpretability methods using a unified framework. Another contribution is the application of domain transfer, especially in financial domains.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

III. PROPOSED TECHNOLOGY

A. Environment Design

The simulation environment consists of a grid with various obstacles, dead ends, and corridors. The Samurai and Assassin agents are positioned randomly throughout the grid. The simulation ends when the Samurai is caught or escapes

B. Assassin Reinforcement Learning.

Assassin agents utilize the Q-learning algorithm for optimal capture.

State Representation :

- Relative Samurai positions
- Encoded distant positions using the quadrant system
- Wall proximity detection
- Local Assassin cluster metrics
- Cornering density metrics

C. Samurai Survival Strategy

The Samurai agent follows a potential field approach, which consists of Goal attraction to the escape zone Repulsion from Assassin agents Panic mode acceleration when proximity Obstacle avoidance. This approach helps the Samurai navigate the grid realistically, focusing on survival rather than combat with the Assassins.

IV. SYSTEM ARCHITECTURE

The proposed system architecture combines Agent-Based Modeling, Reinforcement Learning, and Data-Driven Analytical components in an integrated Adversarial Simulation Environment. Fundamentally, the environment includes a grid-based space with static obstacles, internal walls, and constrained paths. A Scheduler is used to manage interactions among one Samurai agent and multiple Assassin agents in a turn-based manner. Each episode in the environment starts with random agent positioning, terminating when the Samurai agent successfully escapes to a predefined safe zone or is captured by Assassin agents. Each Assassin agent uses Reinforcement Learning modules with Q-learning algorithms to update their decision-making strategies based on feedback from the environment. Each agent receives the current state representation at each time step, which includes relative positioning, proximity to internal walls, clustering, and spatial features. An Epsilon-Greedy action selection strategy is used, allowing agents to transition smoothly from random exploration to learned strategies. After each action, the environment updates to a new state, computes rewards based on environmental feedback, and uses these rewards to update Q-values iteratively. These updates continue until convergence is achieved. In addition, a Data Logging component is used in parallel with the Reinforcement Learning components. This component records various metrics at each episode level and each step level, including agent capture, survival, clustering, and collision frequency. This logged data is then processed in a machine learning analysis module, where Gradient Boosting is used to assess feature importance. SHAP-based explainability techniques are used in this module. Thus, the proposed architecture not only provides a platform for adaptive agent learning strategies but also offers a platform for understanding agent behavior through machine learning analysis. Each agent has an independent operation and can operate under a common environmental state as a result of how modular interaction layers are implemented in the architecture. This enables decoupled architectures to provide inherent scalability relative to agent population without any material effect on system performance.

The architecture also employs a scheduler which provides synchronized updates when multiple agents perform their actions at the same time, thus ensuring there is no state inconsistency in the system.

The reinforcement learning module is designed to adapt to different complexity of the environment, with the learning rate and exploration parameters being able to adjust depending on the environment's level of complexity. As such, agents will effectively balance exploration versus exploitation at all different training phases. Additionally, the inclusion of logging mechanisms ensures that all transitions and reward signals are recorded for use in post-simulation analysis and for reproducing experimental test results.

V. Results and Analysis

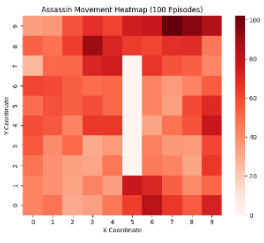
A. Learning Convergence

Assassin agents' convergence is stable with respect to various training seed values. The capture time consistently decreases with the development of coordination behaviors.

B. Survival Capture Balance

Assassin agents' optimal coordination does not compromise the Samurai agents' robustness, suggesting a balance between the two adversarial behaviours.

C. Heatmap Observations



D. Explainable ML Insights

Gradient Boosting indicates clustering density, and spatial topological features are the most significant factors for capture efficiency, whereas SHAP analysis shows instability in internal policies is more significant than the strength of adversaries for failure.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

E. Baseline Comparison

Compared to the greedy approach for the predators, the proposed framework for the predators using the RL approach achieves. Significantly higher win rates Fewer average capture steps Improved coordination indices. In addition, the experimental evaluation analyzes how environmental constraints affect agent behavior. Coordination of Assassin agents is affected by how dense the obstacles are and how complicated the grid layout is. Increased density of obstacles leads to the formation of localized clusters among agents and improves trapping strategies but can result in unnecessary redundancy in movements.

The stability of reinforcement learning was evaluated by running multiple simulations, all had different random start states (seeds). There were consistent behavior patterns in the convergence of all runs, therefore reinforcement learning is a robust approach to modelling. The experiments proved that the proposed model does not depend on the initial conditions and has remain consistent in its performance across a variety of scenarios.

VI. PROPOSED EXTENSION

The adversarial RL framework developed in this research provides a structured abstraction for decision-making in uncertain environments. Although the framework was developed for the pursuit-evasion problem, it is possible to extend it to financial decision environments with stochastic behavior and adversarial agents.

i. *Specifically:* for the equity market

ii. *Possible actions for the agent:* Buy, Sell, or Hold.

iii. *Reward function:* Risk-Adjusted Returns, where volatility and transaction costs are penalized.

iv. *Adversarial components:* market regime shifts, competing agents, and stochastic volatility.

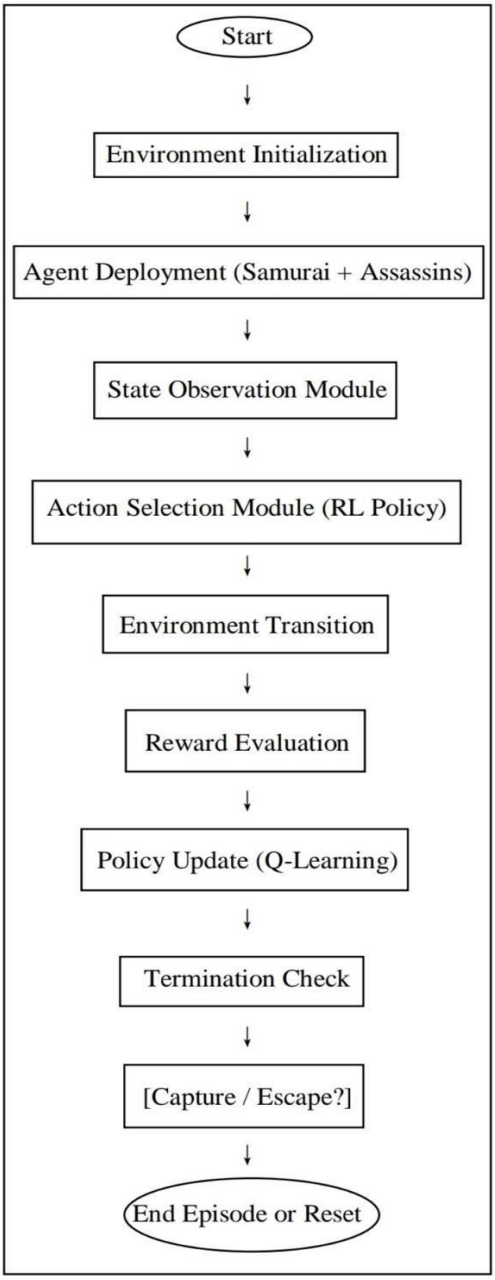
v. *Possible representations for the state variables:* Historical data, technical analysis, macroeconomic variables, and sentiment analysis using retrieval-suggested techniques.

Note: however, that the proposed extension does not assume predictability in the market or the ability to generate profits. The extension to financial environments adds more complexities, considering partial observability and stochastic nature of financial system behavior. Unlike spatial environments, financial systems involve delayed rewards and nonlinear relationships between state variables. Hence, adapting adversarial reinforcement learning to financial environments demands careful design of reward functions and state representations.

Additionally, incorporating sentiment analysis using natural language processing techniques will help improve decision-making capabilities of the agent. The agent will be able to adapt better to changing financial conditions using external data sources such as news articles and financial reports.

VII. FLOWCHART

The flowchart presents a sequential decision-making process for the adversarial (opponent) reinforcement learning (RL) system, which includes initialization of the environment, deployment of an agent, iterative measurements or observations of state, selection of an action via the RL policy, and a response from the environment with the new updated state and a reward for that action. This cycle continues until a pre-defined termination condition is met, which allows for ongoing learning and evolution of agent strategy. This iterative loop continues until a termination condition is met, ensuring continuous learning and adaptation of agent strategies.



International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

VIII. ROBUSTNESS AND GENERALIZATION ANALYSIS

Evaluating robustness is a key requirement within adversarial multi-agent systems, especially within dynamic and partially observable environments. For the suggested model, robustness is achieved through experimentation with varying environment configurations, grid size, obstacles, and Assassin agents. The objective is to evaluate the overall stability of the policy within unseen spatial configurations or adversarial agent distributions. From the experiment, reinforcement learning agents show consistent coordination regardless of the grid size, implying a generalization of policy.

Secondly, robustness is also evaluated within the model through sensitivity analysis, whereby the model is run with varying random initialization parameters. From the experiment, it is evident that the overall policy converges regardless of the random initialization parameters, implying that the policy is not based on a deterministic environment configuration.

IX. REWARD SHAPING AND POLICY STABILITY

Reward shaping is a key requirement within reinforcement learning, especially within emergent policy behavior. For the suggested model, reward shaping is carefully configured with a view to ensuring efficiency, coordination, and overall stability within the model. For this reason, a penalty is assigned for every step within the model, implying that agents must make timely decisions. On the other hand, overcrowding and collision penalties are assigned to discourage chaotic clustering and overall redundancy.

X. CROSS-DOMAIN TRANSFERABILITY

Another significant contribution of the current research lies in demonstrating the transferability of adversarial reinforcement learning architectures beyond the traditional grid-based pursuit-evasion game domain. The structured model of competing players navigating towards objective goals in uncertain situations has significant implications for strategic modeling in financial markets, cybersecurity threat mitigation strategies, and even supply chain optimization strategies. By redefining the action space, reward models, and state spaces, the adversarial reinforcement learning model can be easily extended and applied for decision-making in uncertain regime shifts.

In financial markets, for instance, the objective of the player would shift from spatial capture towards optimization of returns and risk assessment. Similarly, the adversarial model would represent market volatility, competing players, and structural regime shifts in the financial sector. However, it must be noted that such cross-domain modeling is not based on any assumptions of predictability in uncertain situations and is based on measuring adaptability in uncertain situations. This allows it to be applicable in uncertain, partially observable, and dynamic systems. The transferability of this model between different domains reveals its robustness in terms of the policies it learns

XI. COMPUTATIONAL COMPLEXITY AND SCALABILITY

As the size of the multi-agent systems increases, along with the complexity of their environment, another important consideration is the computational efficiency of the model. In the case of the suggested adversarial reinforcement learning model, the scalability of the model can be affected by the dimensionality of the state space, the number of agents, and the grid size. In the case of tabular Q-learning, although the model is effective within a certain grid size, there is a limitation that arises with a substantial increase in environment size, considering that the combinations of actions and states increase exponentially. To overcome this limitation, the model uses a variety of state abstraction methods, such as quadrant encoding, clustering signals, and proximity-based representations, which reduce the dimensionality of the environment without losing important contextual information.

XII. INTERPRETABILITY AND BEHAVIORAL TRANSPARENCY IN ADVERSARIAL SYSTEMS

In the case of adversarial learning, performance metrics cannot solely explain the behavior of the model, especially considering that the reliability of the model is a key consideration, especially when transferring the model to a real-world scenario, such as finance or decision-making systems. To overcome this limitation, the suggested model uses machine learning-based explainability methods that assess feature importance and behavior influence.

XIII. LIMITATIONS AND FUTURE WORKS

However, while the proposed adversarial reinforcement learning framework has shown promise in achieving high performance, there are several limitations which need to be addressed in the future. To begin with, while the proposed framework has shown promise in achieving high performance, the current implementation of the proposed approach uses tabular Q-learning, which may not be efficient in handling large state spaces.

In the future, the proposed approach can be extended to use Deep Reinforcement Learning techniques, which can handle the scalability challenges associated with the proposed approach. Some of the techniques which can be extended are Deep Q-Networks (DQN) and Proximal Policy Optimization (PPO).

Further, the current environment has been designed to handle discrete state movement, which may not accurately represent the real world. In the future, the proposed approach can be extended to handle continuous state movement, which will make the environment more realistic.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

In addition, the proposed approach can be extended to incorporate the use of communication protocols, which will improve the efficiency of the agents in the environment. Further, the proposed extension in the financial environment has been conceptual in nature, which will need to be implemented in the future using simulated environments.

XIV. EXPERIMENTAL SETUP AND EVALUATION METRICS

The experimental setup for the proposed adversarial reinforcement learning model is designed to test the performance of the agent for different scenarios. The simulation is performed in a grid environment with different configurations for obstacles, agent densities, and initial values. Each experiment is performed for multiple episodes to ensure statistical consistency and robustness for the results obtained.

Different evaluation parameters are considered to measure the performance and effectiveness of the model. The parameters include the average capture time, success rate for Assassin agents, survival time for the Samurai agent, and coordination efficiency for the agents. Moreover, convergence rates are also considered to analyze the time required for reinforcement learning agents to converge on a stable policy.

To further validate the performance and effectiveness of the model, experiments are performed for different configurations with different random seed values for initialization. The performance is validated to ensure that the results are not biased for a particular configuration and that the model is generalizable for different scenarios. Additionally, experiments are performed to compare the results with baseline models that use a heuristic approach, highlighting the performance improvements obtained through reinforcement learning.

XV. IMPLEMENTATION DETAILS AND REPRODUCIBILITY

The proposed system is implemented in Python, and it follows a modular approach to distinguish between environment dynamics, agent policies, and analysis components. The grid-based environment is implemented using the Mesa framework, allowing for efficient scheduling of agents. The reinforcement learning approach is based on tabular Q-learning with a greedy exploration strategy, where ε decreases over time to switch from exploration to exploitation.

Various hyperparameters are set based on empirical analysis to ensure convergence of the proposed approach. Multiple experiments are run with different random seeds to test the consistency of the proposed approach. All experiments are conducted in a standard computing environment without using any special hardware, thus proving the efficiency of the proposed approach.

For reproducibility, logs are maintained for all simulations for different episodes. These logs can be further utilized for post-hoc analysis using machine learning techniques and SHAP-based analysis. The proposed implementation can be extended to deep reinforcement learning techniques and can be applied to real-world data with ease due to its modular approach. Logging is incorporated to monitor state changes, reward values, and interactions among the agents at each step. The simulation environment is based on grid-based architecture with tunable parameters. All experiments are performed on a regular computing machine, thus ensuring the efficiency of the solution.

XVI. DISCUSSION AND PRACTICAL IMPLICATIONS

The results obtained through the proposed adversarial reinforcement learning approach offer significant insights into the behavior of multiple agents functioning within a competitive environment. The formation of coordinated behaviors among the Assassin agents signifies the capabilities of reinforcement learning to allow for decentralized decision-making without the need for communication protocols.

From a practical perspective, this approach can be effectively used for applications such as autonomous surveillance systems, cybersecurity defense systems, and robot coordination, among others, where multiple agents are required to function within an uncertain and competitive environment. The adaptability of the approach to dynamic environments also signifies its potential for use in real-time applications.

Moreover, the incorporation of interpretability also ensures that the decision-making process is transparent, which is significant for applications within uncertain and competitive environments. Through an understanding of the factors that affect the behavior of multiple agents, significant improvements can be made to ensure the reliability and fairness of such systems.

XVII. PERFORMANCE ANALYSIS AND OBSERVATIONS

The performance of the proposed adversarial reinforcement learning framework can be further analyzed based on the interaction dynamics between agents over multiple simulation runs. It can be clearly seen that the Assassin agents have improved coordination over multiple learning cycles, thereby reducing the overall capture time and increasing the overall spatial efficiency.

On the other hand, the survival-based strategy for the Samurai agent adapts to different situations based on the proximity and clustering of Assassin agents. This clearly indicates the overall effectiveness of the survival-based strategy for the agent in highly constrained environments.

Moreover, it can be clearly seen that exploration and

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

exploitation have a significant impact on the overall efficiency of the learning process. In the initial stages, exploration is crucial for discovering different strategies, while in subsequent stages, it is more efficient for refining optimal policies.

XVIII. CONCLUSION

This research proposed a reinforcement learning-based Agent-Based Modeling framework for balancing coordinated pursuit and survival-oriented evasion in adversarial environments. The proposed framework leverages explainable machine learning analysis and moves beyond the usual performance metrics to provide deeper insight into the behavior of the system.

The proposed extension for financial decision environments leverages the adversarial reinforcement learning framework for decision-making in uncertain environments, providing a potential direction for future research in this area.

Overall, the proposed framework demonstrates how adversarial reinforcement learning can foster structured coordination and adaptive survival within complex, dynamic environments. The integration of interpretability mechanisms further strengthens the analytical credibility of the system by linking performance outcomes to measurable behavioral features. By validating robustness across varying environmental conditions, the study establishes both stability and scalability of the learned policies. The cross-domain extension into financial decision environments highlights the broader applicability of adversarial learning architectures beyond spatial simulations. Collectively, these contributions position the framework as a flexible and research-oriented foundation for modeling strategic decision-making under uncertainty. Moreover, the proposed system holds a high potential for practical application in real-world academic settings, where manual work can be minimized without compromising compliance. Its flexibility allows it to be applicable for various academic needs. that it

XIX. FUTURE SCOPE

The suggested model can be enhanced through approaches like DQN and PPO that make use of deep reinforcement learning. Moreover, future work can include the implementation of the framework in continuous state spaces to simulate even more complex scenarios.

In addition, the inclusion of communication channels between agents can increase efficiency when working within a multi-agent system. The model can also be used in practical applications including but not limited to finance, cybersecurity, and robotics.

REFERENCES

[1] T. Vincent Gnanaraj and J. Chenni Kumaran, “Samurai and assassins strategy using agent-based modelling and simulation for finding success probability and efficiency in complex scenarios,” *Discover Computing*, vol. 28, p. 193, 2025.

[2] D. Masad and J. Kazil, “Mesa: An agent-based modelling framework in Python,” in *Proceedings of the 14th Python in Science Conference*, 2015, pp. 53–60.

[3] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[4] C. J. C. H. Watkins and P. Dayan, “Q-learning,” *Machine Learning*, vol. 8, no. 3–4, pp. 279–292, 1992.

[5] C. M. Macal and M. J. North, “Tutorial on agent-based modelling and simulation,” *Journal of Simulation*, vol. 4, no. 3, pp. 151–162, 2010.

[6] V. Mnih *et al*., “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, pp. 529– 533, 2015.

[7] M. Wooldridge, *An Introduction to Multi-Agent Systems*, 2nd ed. Chichester, U.K.: John Wiley & Sons, 2009.

[8] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal Policy Optimization Algorithms,” *arXiv preprint arXiv:1707.06347*, 2017.

[9] M. L. Littman, “Markov games as a framework for multi- agent reinforcement learning,” in *Proceedings of the 11th International Conference on Machine Learning*, 1994, pp. 157–163.

[10] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, “Multi-agent actor-critic for mixed cooperative- competitive environments,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[11] I. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and harnessing adversarial examples,” in *International Conference on Learning Representations (ICLR)*, 2015.

[12] Z. Zhang, S. Zohren, and S. Roberts, “Deep learning for trading,” *Journal of Financial Data Science*, vol. 2, no. 2, pp. 8–20, 2020.

[13] A. Paszke *et al*., “PyTorch: An imperative style, high- performance deep learning library,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[14] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.