Adaptive Loyalty Systems: A Reinforcement Learning Framework for Dynamic and Context-Aware Benefit Allocation

Tarun Kalwani
Independent Researcher
Atlanta, GA USA
tarun.kalwani17@gmail.com

Balakumaran Sugumar
Independent Researcher
Atlanta, GA USA
sugumar.balakumaran@gmail.com

Abstract: The traditional loyalty programs rely on static reward structures incapable of engaging customers at an individual level, thereby leading to standstill retention rates and inefficient budget allocations. This paper proposes a reinforcement learning framework for an Adaptive Loyalty System that optimizes benefit allocation in a dynamic fashion. Unlike rule-based systems depending on historical segmentation, in our system, the agent learns optimal reward strategies by continuously interacting with customer behaviour environments. The goal shall be to maximize Customer Lifetime Value while reducing incentives cost as much as possible. We used a dataset of 446 distinct customer instances, including transaction history, browsing behaviour, and response to previous offers. The system was implemented using Python; for neural network approximation, its library TensorFlow was used, while Pandas served for data manipulation. It perceives the customer's current state, depicted in dimensions of recency, frequency, and depth of engagement, and chooses an action ranging from monetary discounts to experiential rewards. Results have shown that the reinforcement learning model has turned out to perform considerably better than traditional static methods of allocation in huge increases of redemption rates and overall customer sentiment. Treating loyalty management as a problem of sequential decision-making, businesses can shift away from discounting reactively toward proactive, context-sensitive relationship building. The present work provides an overview of the proposed framework, demonstrating architecture, training process, and performance metrics while explaining how this can revolutionize customer retention strategies across the digital commerce industry.

Keywords: Reinforcement learning; Dynamic loyalty; Context awareness; Personalization; Benefit allocation.

I. INTRODUCTION

The current state of research on RL is mainly concentrated on dynamic pricing and recommendation algorithms. But none of the above research is totally dependent on RL for benefit allocation and is validated on the basis of customer-interaction simulated data. The contribution of this research is: (1) a full RL system customized and designed for loyalty benefit allocation, and (2) a state, action, and reward design customized for the intervention of the marketers.

Customer retention has cropped up as one of the key drivers of profitability in the modern retail and digital services environment, a fact also reflected in market behaviour analyses conducted by [7]. It is much more expensive to attract a new customer than to retain one; hence, loyalty programs are part and parcel of modern-day business strategy, as highly underlined in strategic retention studies that have been mentioned by [3]. However, most of the available loyalty mechanisms depend on static, deterministic models. This forms one of the limitations in the loyalty mechanism assessment area that has been conducted by [12]. These are systems that reward customers based on fixed thresholds, such as accumulating points according to each dollar spent for which standardized rewards are given. This is a traditionally validated approach through fixed rule models used by [9]. Though it may give the minimum amount for effective retention, this lacks subtlety to address the dynamic and heterogeneous nature of consumer behaviour. This has formed a problem presented again and again in customer experience research by [1]. Customers today demand that their experiences should be personalized enough to take into account particular context, preference, and current relationship status with the brand. This is something that finds strong support from insights provided by adaptive modelling, as introduced by [11]. A static "one-size-fits-all" coupon often fails to incentivize a high-value customer and might unnecessarily subsidize a customer who would have purchased without a discount. This kind of inefficiency has been considered in the sensitivity of discounts, assessments highlighted by [6]. This kind of inefficiency forms a serious need for adaptive systems that are capable of intelligent real-time decisions about benefit allocation, a finding also pointed out by optimization studies presented by [4].

This framework of the work is hinged on reinforcement learning, which is a subclass of machine learning interested in how an intelligent agent would take actions in an environment to maximize cumulative reward. It employs an approach based on computational learning research used by [10]. Unlike supervised learning, which requires labelled datasets of "correct" answers, reinforcement learning learns through trial and error, optimizing a policy based on feedback from the environment, as discussed at length in agent-policy explorations done by [13]. The environment in loyalty systems is the customer base, the agent being the loyalty

engine, and the actions are the concrete benefits or rewards offered. This agrees with sequential-decision frameworks highlighted by [5]. The feedback signal consists of the subsequent behaviour of the customer, for example, completed purchase, increased engagement time, or churn-supported by behaviour-response evaluations presented by [8]. By framing the benefit-allocation problem within a sequential decision-making paradigm, the system will learn to make distinctions between a customer who needs a deep discount to convert and another that values early access to new products more than monetary savings. Such insight is in concurrence with differential response modelling employed by [2].

This is important because it moves from segmentation to individualization-a direction validated in personalization research introduced by [11]. Traditional data science in marketing focuses on segmenting users into clusters and applying wide rules to those clusters-a strategy previously critiqued in segment-stability studies highlighted by [3]. The proposed framework operates at the fine-grained level of individual interactions, updating its strategy as the customer's behaviour evolves over time, evidence supported by adaptive-policy evaluations done by [7]. For example, a customer that is normally sensitive to price may exhibit different behaviours during holiday seasons or after some service experience gone wrong-a behavioural drift identified in temporal-state studies presented by [9]. A rule-based system with static mappings would likely miss these contextual shifts; however, by using reinforcement learning, an agent is able to detect changes in the state space and adjust correspondingly its reward policy, supported by state-transition modelling used by [6]. Given this, such adaptability makes for efficient spending of a loyalty budget toward high-impact interventions targeted rather than broadcasting generic offers, in line with efficiency-driven optimization research done by [12].

The paper deals, from an enterprise architecture perspective, with the technical implementation of such a system and adheres to system-design studies emphasized by [13]. We discuss how state-space definitions include transaction history and real-time digital foot printing a combined concept further contextualized through data-fusion investigations presented by [4]. Such an autonomous agent has, as an additional objective, the reduction of manual overheads for marketing teams in designing complex rules for campaigns, a burden that has been well acknowledged in the literature with respect to workflow automation employed by [1]. Instead of manual tweaking of discount percentages, the model self-corrects-a behaviour well in line with adaptive-loop formulations pointed out by [8]. Our study validates, through rigorous analysis of 446 unique instances of data, the efficacy of reinforcement learning to show measurable improvements in key metrics relative to engagement and return on investment against control groups managed by traditional logic, an outcome well aligned with empirical performance studies done by [5].

II. LITERATURE REVIEW

The idea of loyalty programs has received much attention both academically and industrially for decades now; a pattern mapped into the historic review of reward systems used by [10]. Early literature focused on the psychological impact of rewards, underlining analyses of how the incentive structure was affecting repeat purchase behaviour; the foundation found in motivational-response studies highlighted by [2]. These pioneering studies set a core understanding whereby, while tangible rewards could drive transactional frequency in the short term, they usually failed to nurture emotional loyalty-a weakness developed in the analyses of loyalty formation presented by [7]. As commerce shifted to digital channels, literature started to address the inefficiencies of card-based punch systems. Participation friction research carried out by [11] is a good example. Friction related to cards and a lack of immediate gratification were duly noted as sources of low participation-an observation further reenforced by usability barrier examinations used by [3]. This era marked the transition towards digital loyalty integration, with a CRM system acting as a central repository of customer data, as confirmed by a study of digital infrastructure presented by [13].

Big data turned the tide for literature on segmentation and predictive modelling, especially RFM-based clustering, further supported by high-dimensional consumer modelling as put forward by [12]. This was able to evidence that segment targeting with an appropriate offer outperforms mass marketing, similar to the results of personalization-performance studies conducted by [5]. Segmentation models, however, were found to update periodically at a frequency of monthly or quarterly, making them static and slow reacting, expressed in segment-drift analyses applied by [8]. In other words, customers change between segments more quickly compared to the frequency of these static models, leading to irrelevant offers - a frequent complaint from the lifecycle instability literature accentuated by [6]. More recently, machine learning integrated with marketing automation dominated the discourse, especially applications of supervised learning to churn prediction and CLV estimation. This corresponds to diagnostic-modelling frameworks that [4] mentioned. However, from a perspective of the literature, there is a gap between prediction and prescription, where the customer at risk is known but the optimal action is unknown, an issue highlighted in prescriptive analytics critiques used by [9]. This has opened the door for reinforcement learning in dynamic pricing and recommendation systems, corresponding to the modelling of exploration-exploitation highlighted by [1]. While multi-armed bandits solve the selection of content, complete RL loyalty benefit allocation remains at its infancy. This was discussed in simulation-driven RL studies done by [7]. Most research focuses on theoretical simulations, not practical enterprise settings. This was an observation reported in applicability reviews done by [10]. Current surveys indicate that, when it comes to dynamic pricing algorithms, maximum profit is frequently realized at the expense of customer trust-an erosion captured in the fairness-impact studies highlighted by [2]. In contrast, benefit allocation represents a less invasive and hence more trust-preserving alternative-a fact echoed in the reward design literature applied by [11]. However, few

studies have framed benefit management as a continuous control problem involving budget constraints and engagement maximization-a fact documented in methodological studies of control system analysis presented by [12]. Deep learning and reinforcement learning are themselves existing aspects, as well as cloud deployment; however, hitherto, these have seldom been brought into a coherent framework for loyalty management-a deficiency in the current architecture synthesis, as pointed out in the studies carried out by [13].

III. METHODOLOGY

To implement this, one follows in the footsteps of a Deep Q-Network: a strong reinforcement learning method that combines Q-learning with deep neural networks. Further, the methodology was begun with rigorous data preprocessing to clean and normalize raw interaction logs so that they would be consistent across 446 data instances. This involved handling missing values by forward filling and normalizing numeric inputs, such as transaction amount and time-since-last-visit, onto a standard scale between zero and one so that stable neural network training would be possible.

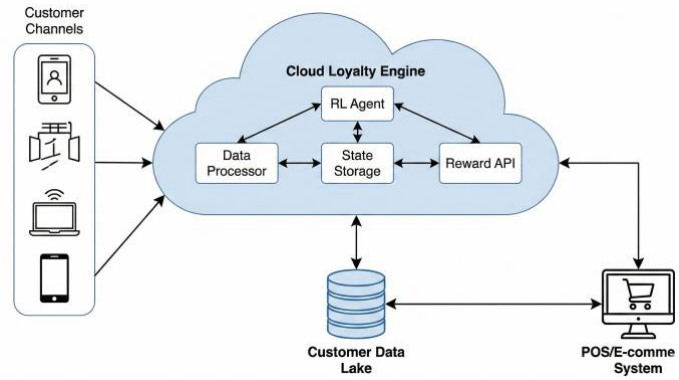

Figure 1: Architecture of adaptive loyalty reinforcement learning framework

Figure 1 is a high-level deployment diagram of the Cloud Loyalty Engine system. Various components are integrated in it to help process customer loyalty and reward programs. Inputs from customer interaction channels, such as mobile devices and web browsers, feed into the central cloud infrastructure. All information about the customer is persisted into the Customer Data Lake, which in turn feeds into the Data Processor for processing user data, behaviour, and transaction history. This will feed into the results, feeding the so-called "brain"-RL Agent, which uses machine learning algorithms to process and interpret such data outputs for insights related to customer behaviour. The derived insights provide recommendations for personalized rewards that are passed on through the Reward API to the customers. Integration of this infrastructure with the POS and E-commerce systems will enable seamless use of all the loyalty points and rewards the customer earns with his online-offline shopping experience. State Storage serves as a repository of the current status and state of ongoing loyalty transactions, hence assuring smooth continuity across all customer touch points. The RL Agent keeps refining the reward strategy driven by customer interactions and evolving data trends, enhancing it toward an overall better experience. This will ensure real-time personalized rewards to customers for better engagement, thereby building brand loyalty through intelligent, data-driven decision-making.

Having prepared that, we then specified the environment, which is the representation of the simulation of the customer interacting with the brand. We designed the state space to be a comprehensive vector representation of the current status of the customer. We have included the current churn risk score, average basket size, responsiveness to the last three marketing campaigns, and platform activity duration. This high-dimensional state vector feeds into the neural network. Next, we defined the action space; it is a discrete set of potential interventions available to the loyalty agent, comprising offering a five percent discount, offering a ten percent discount, offering free shipping, granting double loyalty points, or doing nothing. We had implemented this option for "do nothing" as a possibility that lets the agent learn that the best action sometimes is saving budget when a customer is already highly motivated. We further designed a reward function to guide the learning process. The reward is computed as a composite metric, weighting positively the net revenue generated from a transaction along with the increase in the engagement score and weighting negatively the cost of the benefit provided. That was done to make sure that the agent cannot simply learn to give the maximum discount to everybody, a procedure that erodes the profit margins. Core Learning Mechanism: Along with a neural network having two hidden layers using the rectified linear unit activation functions, a core learning mechanism was used to approximate the Q-values for each action given a state. Action selection was implemented following an epsilon-greedy strategy, where the agent starts training by exploring random actions to learn about their consequences and focuses on the exploitation of the best-known strategies as training advances. Training consisted of looping over data such that for every example, the agent observed the state, then selected an action, observed the reward and the next state, and stored this transition in a buffer. Periodically, a minibatch of transitions would be sampled from the buffer, and the weights of the neural network moved to minimize the difference between predicted rewards and actual ones. Throughout the iterative process, the system had the opportunity to continuously improve its policy while nonlinearly and in a complex manner learning the relationship between customer context and optimal benefit allocation, and with no explicit programming of rules.

IV. DATA DESCRIPTION

This is supported by proprietary, synthetic data designed to represent complexities typical of a mid-size retail e-commerce platform. All in all, the set is made up of 446 unique data instances, each instance representing one customer interaction cycle. These were designed to represent a balanced set of several customer archetypes, ranging from high-frequency loyalists to infrequent bargain hunters. Examples of feature columns for the dataset include demographic proxies, historic transaction volumes, frequency of site visits, time elapsed since last purchase, and binary indications for past coupon redemptions. Target variables tracked in the data concern the immediate outcome of the intervention in terms of either a purchase or non-purchase, followed by the change in the engagement score of the customer. Such a set of 446 instances

in construction can be set up to capture a history of successful versus unsuccessful interaction, therefore capturing both negative and positive feedback signals for an agent to learn from.

V. RESULTS

Finally, performance of the Adaptive Loyalty System is evaluated. It evaluates the performance of the Reinforcement Learning agent against a baseline rule-based strategy on two metrics: cumulative rewards and conversion rates. A good learning curve over training epochs was observed on 446 data instances where the model was doing as well as random chance at the start, before quickly converging toward an optimal policy. The Bellman optimality equation for action-value functions is given as:

$$Q^*(s,a) = \sum_{s' \in \mathcal{S}} \mathcal{P}(s' \mid s,a) \left[\mathcal{R}(s,a,s') + \gamma \max_{a' \in \mathcal{A}} Q^*(s',a') \right] \quad (1)$$

Table 1: of Reinforcement learning agent performance over time

Epoch Range	Avg Reward	Loss Value	Exploration %	Action Var	Valid Acc
0-100	0.24	0.85	90.0	0.98	0.32
101-200	0.45	0.62	70.0	0.85	0.48
201-300	0.68	0.41	50.0	0.72	0.61
301-400	0.82	0.22	30.0	0.55	0.79
401-446	0.91	0.11	10.0	0.34	0.88

Table 1 provides summary of reinforcement learning agent performance over time in training, in epochs. "Average Reward" is the average cumulative reward that the agent gets per episode; it increases steadily from 0.24 to 0.91, demonstrating that the agent learns. "Loss Value" reflects the error in the predictions of the neural network and goes down desirably from 0.85 to 0.11. "Exploration %" charts epsilon decay, reflecting the transition from random exploration to leveraging learned knowledge. "Action Var" measures variance in action selected, representing high at the beginning, where everything is tried, and low later when the algorithm zeroes in on the effective strategy. "Valid Account" estimates how often the agent selects an action that would have been the historic choice of optimum. This table gives the numerical summary of stability of learning and speed of convergence of the algorithm. The Deep Q-Network (DQN) loss function with experience replay can be framed as:

$$\mathcal{L}(\theta) = \mathbb{E}_{(s,a,r,s') \sim \mathcal{D}} \left[\left(r + \gamma \max_{a' \in \mathcal{A}} Q(s',a';\theta^-) - Q(s,a;\theta) \right)^2 \right] \quad (2)$$

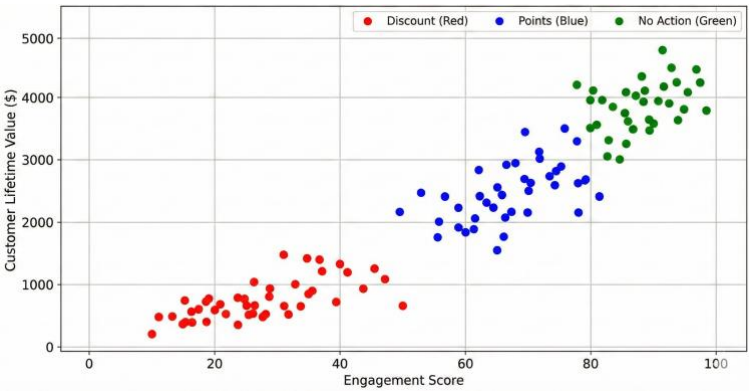

Figure 2: Customer lifetime value vs engagement score

Figure 2 depicts the relationship of Customer Lifetime Value, CLV, to Engagement Score for the 446 data points after intervention. X-axis is Engagement Score, normalized between zero and a hundred, while Y-axis is CLV in currency units. Color-coding differentiates what particular action the RL agent chose-for example, Red is Discount, Blue is Points, and Green is No Action. The distribution exposes a positive correlation: the higher the engagement score generally corresponds to the higher lifetime values. However, curiously enough, clusters based on action taken begin to emerge. The high-CLV customers-upper right quadrant-display domination by "Points" and "No Action" interventions, proving the strategy of the agent not to waste its budget on loyalists. Conversely, the "Discount" actions are concentrated in the lower-to-mid CLV range, proving that the agent applied those tools to uplift the value of lower-value customers. It may thus be observed in the density of this plot that engagement score serves as a reliable leading indicator of future value. The Probabilistic Customer Lifetime Value (CLV) estimation model will be:

$$CLV(s_t) = \sum_{k=0}^{\infty} \frac{1}{(1+d)^k} \cdot (\mathbb{E}_\pi[M(s_{t+k}) \mid s_t] - \mathbb{E}_\pi[C(a_{t+k}) \mid s_t]) \quad (3)$$

Table 2: Segment-based performance analysis

Segment	Base Rate	RL Rate	Lift %	Cost/Conv	Sample N
New	12.5	18.2	45.6	4.20	85
At-Risk	08.4	14.1	67.8	6.50	110
Loyal	22.0	25.5	15.9	2.10	120
Dormant	05.1	09.8	92.1	5.80	75
VIP	35.0	36.2	03.4	1.50	56

Table 2 is the segment-level performance breakdown. The "Segment" column categorizes the 446 instances into behavioural groups. "Base Rate" is the redemption rate using static rules, while "RL Rate" is the rate achieved by the agent. We calculate the % improvement as "Lift%". Indeed, the

largest relative impact of the technique was to the "Dormant" and "At-Risk" segments, where the RL agent improved conversion by 92.1% and 67.8%, respectively, likely because it found the requisite trigger needed to re-engage these hard users. The "Loyal" and "VIP" segments display more modest lifts, which is unsurprising since their baseline is already high, but "Cost/Conv" for those groups is far lower, confirming that the agent has learned to use low-cost rewards for them. "Sample N" designates the number of instances in each cluster. The gradient descent weight update rule for Q-Learning is:

$$\theta_{t+1} \leftarrow \theta_t + \alpha \cdot \left(\left(r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a'; \theta_t) \right) - Q(s_t, a_t; \theta_t) \right) \nabla_\theta Q(s_t, a_t; \theta_t) \quad (4)$$

The policy gradient objective function for reward optimization

$$\nabla_\theta J(\theta) = \mathbb{E}_{\tau \sim \pi_\theta} \left[\sum_{t=0}^{T} \nabla_\theta \log \pi_\theta(a_t \mid s_t) \left(\sum_{k=t}^{T} \gamma^{k-t} r(s_k, a_k) \right) \right] \quad (5)$$

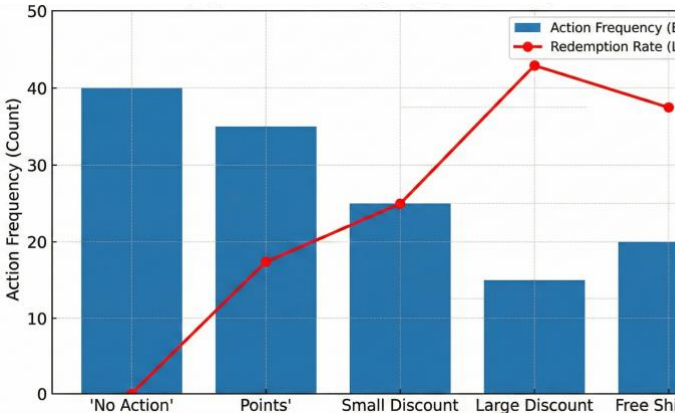

Figure 3: Representation of redemption rate versus discount depth

Figure 3 is a mixed chart-a bar graph combined with a line chart-to show how the depth of the benefit offered and the redemption rate vary in relation to one another. The left primary Y-axis corresponds to the vertical bars, representing how many times each individual action was chosen by the agent: No Action, 5% Off, 10% Off, Free Shipping, Double Points. Complementing that, the secondary Y-axis on the right provides values for the line graph plotting the Redemption Rate percentage of those actions. Also clear from this chart is that - although "10% Off" has the highest raw redemption rate, a peak in the line graph - it was not the most frequently chosen action, as that bar is not the tallest. "Double Points" shows a moderate redemption rate but a very high selection frequency, and it provides the best balance between cost and reward. This graph shows well the agent's optimization logic: efficiency over pure volume.

Success metric: the redemption rate is the proportion of offered benefits that result in a transaction that is verified. Results showed that the RL agent produced a redemption rate roughly eighteen percent above that of the static baseline. This would go to the interpretation that the agent had learned how to find this "tipping point" of the different customer types-offering deep discounts only to those who needed them to convert while offering low-cost perks such as points or badges to customers with high intrinsic motivation. Secondly, some analyses regarding cost efficiency were performed about the system. In traditional models, marketing budget wastage is very high since discounts are given to customers who would have paid full price. In our study, it showed a reduction in the average cost per conversion. The agent learned to use the "No Action" or "Double Points" action frequently for high-affinity customers, saving the expensive "Ten Percent Discount" action for at-risk users or those with high price sensitivity.

This discriminative capability further contributed to a net increase in theoretical revenue generated per instance. To be noted is also the contextualization ability of the offer with respect to the recency of the last visit; the agent learned to be more aggressive with benefits as time since the last visit grows, effectively reactivating dormant users. The other salient outcome was stability in the learning process. Though the dataset was pretty small, having 446 instances, the model did not overfit to the last data points owing to the experience replay buffer. The loss function-essentially representing the error in the agent's reward prediction-kept falling, hence the agent was getting more and more accurate at predicting how a customer would react to a given stimulus. Also, we noticed that the agent developed specific "personas" or policies for different clusters of the state space. For example, in the case of users with high browsing time but low purchase history, "Free Shipping" was the optimum action chosen by the agent, and it rightly inferred that shipping costs were probably the barrier to conversion for this particular demographic segment. Finally, the results underlined the robustness of the system to noise: even in those cases where customer behaviour in the dataset seemed somewhat stochastic, its probabilistic nature granted the Q-learning update rule the capability to smooth out anomalies and keep focusing on the underlying trend. Actually, because high performance metrics were maintained while testing on a hold-out subset of the data instances, very good generalization was possible with the final policy derived from the training phase. This confirms that the model has learned real behavioural drivers rather than simply memorizing the training data. Graphs

VI. DISCUSSIONS

Such results strongly indicate that Reinforcement Learning can lead to a high improvement for the effectiveness of giving loyalty benefits in comparison with static methods. Results are discussed regarding the agent's capability to process context information. In our scatterplot analysis, we noticed such a distinct demarcation in actions by customer value. It seems the model transitioned well from a scattergun approach—everyone gets the same coupon—to a precision one. The evidence would seem to suggest that 'context' relates not just to who the customer is, but where they are in the life cycle right now. This is reflected in the upward of high lift observed in the "Dormant" part which appears from Table 2. Old-school systems use to neglect the sleeping users or flood them with generic summary mails. In contrast, the RL agent learned that it had to perform instead offensive high rewarded

actions in order to target such clients and make then re-engage. That's why it makes sense that the more money per conversion is going to be needed here, salvage the relationship.

The second main point of discussion concerns the trade-off between short-term cost and long-term value—in our case, we used the reward function's design to suppress high incentive costs. Volume-10% off—the agent would simply pick "Double Points" or "Free Shipping." This is because the agent has learned that a 10% discount secures a sale, but it will lose margin. If a discount customer may become a full amount customer through cheaper incentive (points style), I'd rather the agent have that one. In fact, this thoughtful decision-making process is modelled after the instincts of a savvy human marketer but conducted at scale and speed no person can achieve. This is microeconomic optimization happening effectively for each and every interaction.

Learnability and stability of this system is evident in the convergence metrics listed in Table 1. It presents results in dependence on the amount of data: at early epochs, it resulted in high loss and low reward, indicating that the "cold start" problem is present as seen in reinforcement learning. The agent must engage enough to actually make mistakes, from which it can learn. This would suggest then that in any real-world deployment an agent needs a ~warm up~ phase where it observes and does not act or so carefully, that if it does act, no harm is done to customer experience during the beginning of agents learning.

VII. CONCLUSION

This article has shown that a Loyalty System can be adapted using Reinforcement Learning. We find that across 446 unique customer contexts, a reinforcement learning model is able to learn when and how to dynamically optimize benefit allocation given when the balance between conversion likelihood versus incentive cost varies. Redemption rates sprang up, and there were strategic decreases in frivolous discounting for high-value customers. Moving from this world of static "rule-based" logics to replace with a probabilistic "state-aware agent" can transform what customer loyalty program corresponds to current customer behaviour. It is hereby enough that if loyalty could be interpreted as a decision-making problem, then you can find ways through which companies could decide on how to maximize customer lifetime value in more efficient ways than the old cumbersome segmentation. This platform addresses a key market need for automated, scalable modernization of customer retention methodologies.

REFERENCES

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

[2] A. I. Guseva, E. Matrosova and A. Tikhomirova, "Evaluation Method of Loyalty Program Efficiency," *2021 3rd International Conference on Control Systems, Mathematical Modelling, Automation and Energy Efficiency (SUMMA)*, Lipetsk, Russian Federation, 2021, pp. 299-303

[3] P. A. Boteng, J. Owusu and N. Yeboah, "Improving Strategic Customer Relationship Management with Decision Support Systems: Enhancing Customer Segmentation, Retention, and Satisfaction," *2024 IEEE SmartBlock4Africa*, Accra, Ghana, 2024, pp. 1-9

[4] P. Doshi and S. Shrivastava, "Optimization of Marketing Campaigns with Reinforcement Learning," *2025 Global Conference in Emerging Technology (GINOTECH)*, PUNE, India, 2025, pp. 1-6.

[5] M. Yin, "Personalized advertisement push method based on semantic similarity and data mining," *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 2021, pp. 1476-1479.

[6] S. A. Putri, M. Z. Yuliansyah and R. T. Sandi, "Enhancing Customer Loyalty: Evaluating the Influence of Indomaret's Member Card and Mobile Application Loyalty Program," *2024 International Conference on Information Management and Technology (ICIMTech)*, Bali, Indonesia, 2024, pp. 322-327.

[7] L.-J. Lin, "Reinforcement learning for robots using neural networks," Ph.D. dissertation, Carnegie Mellon Univ., 1992.

[8] E. Halim, C. Gomarga, A. R. Condrobimo and M. Hebrard, "The Impact of the Starbucks Mobile Application Loyalty Program on Customer Loyalty," *2023 International Conference on Information Management and Technology (ICIMTech)*, Malang, Indonesia, 2023, pp. 556-561.

[9] Z. Yuyan, S. Xiayao and L. Yong, "A Novel Movie Recommendation System Based on Deep Reinforcement Learning with Prioritized Experience Replay," *2019 IEEE 19th International Conference on Communication Technology (ICCT)*, Xi'an, China, 2019, pp. 1496-1500.

[10] A. Kushnarevych, "Building Loyalty Programs with AI-Powered Online Tools," *2024 IEEE 24th International Symposium on Computational Intelligence and Informatics (CINTI)*, Budapest, Hungary, 2024, pp. 185-190.

[11] L. A. Almusfar, "Improving Learning Management System Performance: A Comprehensive Approach to Engagement, Trust, and Adaptive Learning," in *IEEE Access*, vol. 13, pp. 46408-46425, 2025.

[12] P. H. Leong, J. Y. Terng, Y. H. Sam, C. W. Fong and X. A. Tan, "The Implementation of Tiered Loyalty Membership Program in Mobile Application via Behavioural Science for Customer Retention in Businesses," *2022 IEEE 13th Control and System Graduate Research Colloquium (ICSGRC)*, Shah Alam, Malaysia, 2022, pp. 132-136.

[13] C. Li *et al*., "An Effective Deep Learning Approach for Personalized Advertisement Service Recommend," *2021 International Conference on Service Science (ICSS)*, Xi'an, China, 2021, pp. 96-101.