Dean&Francis
ISSN 2959-409X

Exploring the Role of Dopamine in Curiosity-Driven Learning Using Reinforcement Learning Models

Meixin Zhou

Abstract:

Curiosity-driven learning is a core feature of biological intelligence, manifested in the active exploration of environments by organisms even in the absence of external rewards. This study integrates neurobiological findings with reinforcement learning theory to construct a unified computational framework that reconceptualizes dopamine as an encoder of "total prediction error." In this framework, behavioral selection maximizes total value, which comprises both external rewards and intrinsic informational value. Intrinsic rewards are quantified based on novelty or prediction error, while the weight assigned to exploration is dynamically modulated by the prefrontal cortex. Dopamine neurons encode total prediction error, thereby extending the conventional role of dopamine from a reward prediction error encoder to a more general encoder of violated expectations. The hippocampus generates intrinsic reward signals, the prefrontal cortex regulates the exploration weight, the ventral tegmental area integrates these signals to compute total prediction error, and the striatum translates this error into action selection. This model bridges the explanatory gap between neuroscience and computational theory, offering a unified perspective on the intrinsic motivation underlying curiosity and exploratory behavior. In practical terms, it also provides new insights into the mechanisms of psychiatric disorders.

Keywords: Curiosity-Driven Learning; Dopamine; Reinforcement Learning; Intrinsic Motivation

1. Introduction

Curiosity-driven learning is a fundamental characteristic of biological intelligence: infants repeatedly manipulate objects without external rewards, animals actively explore novel environments, and humans immerse themselves in play and creative endeavors. Why do organisms engage in active learning when

ISSN 2959-409X

explicit external rewards are absent? Classical reinforcement learning theory, which relies on external reward signals to drive behavior, struggles to account for learning in reward-free contexts. Concurrently, seminal findings in neuroscience have traditionally anchored the dopamine system to the encoding of reward prediction error. However, subsequent research has revealed that dopamine neurons also respond robustly to novel stimuli. This raises a critical question: Does dopamine encode "reward" specifically, or a more general concept of "violated expectation"? And how can these two types of responses be unified within a single computational framework? This study integrates neurobiological discoveries with reinforcement learning theory to develop a unified computational framework that reinterprets dopamine as an encoder of "total prediction error," which integrates both external rewards and intrinsic novelty. This framework aims to elucidate the core role of dopamine in curiosity-driven learning. It bridges the explanatory gap between neuroscience and computational theory, providing a unified perspective for understanding the intrinsic motivation behind curiosity and exploration. Furthermore, it offers new avenues for exploring the mechanisms underlying psychiatric disorders.

2. Relevant Theory and Technical Foundations

2.1 Neurobiological Basis: The Dual Function of the Dopamine System

Seminal work by Schultz and colleagues, using classical conditioning paradigms, established the foundation for understanding dopamine function [1]. They demonstrated that the phasic firing patterns of dopamine neurons precisely encode the discrepancy between predicted and actual rewards: firing increases when an unexpected reward occurs, decreases when a predicted reward is omitted, and shifts to the time of a conditioned stimulus when a reward is fully predicted. This firing pattern closely mirrors the reward prediction error in temporal difference learning, solidifying dopamine's theoretical status as a "reward learning signal." However, subsequent research has revealed that dopamine's responsiveness extends far beyond external rewards [2]. The hippocampal-VTA loop model proposed by Lisman and Grace suggests that when the hippocampus detects environmental novelty—a mismatch between current input and memory-based expectation—it can indirectly activate VTA dopamine neurons via the subiculum and the shell of the nucleus accumbens, eliciting dopamine release even in the complete absence of external reward. Electrophysiological studies confirm that dopamine neurons respond vigorously to initially presented stimuli, with this response habituating upon repeated presentation. These two lines of evidence collectively demonstrate the dual response characteristics of the dopamine system. Whether these two types of responses operate in parallel or are functionally integrated remains an open question in neuroscience, and this theoretical gap constitutes the starting point of the present study.

2.2 Computational Basis: A Paradigm Shift from External Reward to Intrinsic Motivation

Traditional reinforcement learning, framed within Markov decision processes, computes reward prediction error via temporal difference learning to drive agents towards maximizing cumulative external reward. While highly successful in explaining goal-directed behavior, its fundamental limitation lies in its passive dependence on external rewards; the agent learns only upon receiving a reward, failing to account for active exploratory behavior in reward-free contexts [3]. The exploration-exploitation dilemma is a core challenge in reinforcement learning. Conventional solutions treat exploration merely as a "strategy" to achieve maximal reward, rather than a "motivation" stemming from intrinsic drives, thus failing to address why organisms are inherently inclined to explore [4]. Intrinsically motivated reinforcement learning treats information itself as a reward signal with intrinsic value, leading to several quantitative modeling approaches: novelty-based models define intrinsic reward as a decreasing function of state visitation frequency, driving agents to explore under-sampled areas; prediction error-based models assign high intrinsic reward to transitions that yield large errors in a forward dynamics model, causing agents to be attracted by surprising events; information gain-based models define intrinsic reward as the reduction in uncertainty of model parameters, quantifying "learning progress" and enabling agents to learn continuously without external rewards, providing theoretical tools for addressing the sparse reward problem.

2.3 Research Review

Some studies have attempted to map intrinsically motivated reinforcement learning models onto neurotransmitter systems, for instance, analogizing a prediction error-based curiosity module to dopamine function, or linking novelty count models to the pattern separation functions of the hippocampus. These efforts have preliminarily established correspondences between computational models and neural mechanisms [5]. However, existing integrative research suffers from three significant shortcomings: First, limited

scope: most studies still confine dopamine to an "extrinsic reward" framework, treating novelty responses as an exceptional phenomenon rather than systematically integrating its dual functionality. Second, mechanistic ambiguity: there is a lack of computational explanation for how novelty is transformed into a neural signal, and the integration mechanism between intrinsic and extrinsic rewards remains unclear. Third, absence of dynamic explanation: current models fail to capture how the dopamine system dynamically balances exploration and exploitation over time. Addressing these theoretical gaps, this study proposes a unified framework that reconceptualizes dopamine as an encoder of total prediction error. It posits that phasic dopamine activity encodes a total prediction error signal integrating both extrinsic reward and intrinsic information value, with the exploration weight dynamically modulated by the prefrontal cortex. By mapping the computational processes onto the hippocampus, prefrontal cortex, VTA, and striatum, this framework provides a neurobiological implementation basis for abstract computations, aiming to bridge the explanatory gap between neurobiological findings and computational theory.

3. A Computational Model of Dopamine Encoding Total Prediction Error and Its Neural Implementation

3.1 Formalizing Total Value Maximization and Recasting Dopamine Function

Classical reinforcement learning (RL) frameworks have achieved considerable explanatory success by positing that behavior is organized around the maximization of cumulative extrinsic rewards [6]. However, the inherent limitations of this formulation become apparent when accounting for the intrinsically motivated exploratory behaviors ubiquitously observed in biological agents—consider, for instance, an infant's persistent manipulation of a novel object, an animal's spontaneous investigation of an unfamiliar environment, or the immersive human engagement in play and creative endeavors. These phenomena suggest that the utility function governing behavior likely incorporates a component beyond immediate extrinsic payoff [7]. Within this expanded view, information itself possesses intrinsic biological value, as its acquisition reduces environmental uncertainty, refines predictive models of future rewards, and thereby confers an indirect fitness advantage. We therefore propose that behavioral selection operates to maximize a composite construct: total value, which integrates conventional extrinsic reward with an intrinsic information value term. In the formal language of RL, this is instantiated by a redefined total reward function. At each discrete time step t, the agent, in state s_t, executes an action a_t, transitions to a new state s_{t+1}, and receives a scalar reward signal comprising two components:

$$R_{total}(s,a,s') = R_{ext}(s,a,s') + \eta \cdot R_i(s,a,s')$$

where $R_{ext}(s,a,s')$ represents the extrinsic reward—the traditional survival-oriented signal furnished by the environment. $R_i(s,a,s')$ denotes the intrinsic reward, a mathematical formalization of curiosity quantifying the information gained from the transition. The parameter η serves as a dynamic weighting factor, modulating the contribution of intrinsic value to the total. When $\eta = 0$, the agent reverts to a conventional, reward-maximizing entity; when $\eta > 0$, it simultaneously pursues information. This formulation refocuses the inquiry onto two core elements: the precise quantification of $R_i(s,a,s')$ and the mechanism governing the dynamic adjustment of η.

Several candidate operationalizations for intrinsic reward, grounded in both neurobiology and computational theory, merit consideration. A parsimonious approach links intrinsic value directly to novelty, assigning higher worth to states that have been infrequently encountered. This can be expressed as a decreasing function of state visitation frequency:

$$R_i(s,a,s') = \frac{\beta}{\sqrt{N(s')+1}}$$

where $N(s')$ is the visitation count for state s' and β is a scaling constant. This formulation incentivizes exploration of under-sampled regions of the state space. Biologically, this may be implemented by hippocampal circuits that estimate familiarity or novelty based on representational distinctiveness.

An alternative, and computationally distinct, formulation posits intrinsic reward as a function of prediction error. This captures the intuition that surprising events—those deviating from an agent's internal model—are inherently attractive because they signal potential for learning. If the agent maintains a forward model $f_\theta(s,a)$ predicting the next state, transitions yielding high prediction error are deemed information-rich and thus rewarding:

$$R_i(s,a,s') = \alpha \cdot \left\| f_\theta(s,a) - s' \right\|^2$$

with α as a scaling factor. This computation aligns with mismatch detection mechanisms attributed to hippocampal substructures, such as the CA1 region, which compares actual sensory input with predictions derived from CA3.

ISSN 2959-409X

Such a process resonates strongly with the novelty-detection circuitry described in the Lisman-Grace model.

From an information-theoretic standpoint, intrinsic reward can be more fundamentally defined as the reduction in uncertainty about the agent's model parameters—the information gain. Given a prior distribution over model parameters $p(\theta)$ and a posterior $p(\theta \mid s,a,s')$ after observing a transition, the intrinsic reward is:

$$R_i(s,a,s') = D_{KL}\left[p(\theta \mid s,a,s') \middle\| p(\theta) \right]$$

This quantity directly operationalizes "learning progress." Although computationally more intensive, it aligns with the normative Bayesian brain hypothesis and suggests that dopaminergic signals might encode such dynamic changes in model certainty.

Regarding the dynamic weighting factor η, we conceptualize it as a function integrating interoceptive and exteroceptive state information. When an organism is in a state of energetic deficit, the salience of extrinsic rewards is amplified, warranting a reduction in η to prioritize exploitation of known resources. Conversely, during periods of environmental volatility or following recent episodes of high information gain, η should increase to favor further exploration. Formally:

$$\eta = g(hunger, uncertainty, novelty\_trend, \ldots)$$

where inputs may include homeostatic signals (e.g., ghrelin levels reflecting energy need), estimates of environmental change rate, and metrics of recent exploratory success. Neurally, the prefrontal cortex (PFC) is well-positioned to integrate these diverse signals from regions like the hypothalamus and hippocampus, implementing dynamic control over η by modulating VTA sensitivity to novelty inputs.

3.2 Dopamine as an Encoder of Total Prediction Error

With information value incorporated into the utility function, the role of dopamine—classically viewed as encoding reward prediction error (RPE)—requires fundamental reinterpretation. The canonical RPE hypothesis posits that dopamine release, denoted δ, signals the discrepancy between received and expected reward:

$$\delta = R - V(s)$$

where R is exclusively extrinsic reward. However, this formulation cannot account for the well-documented phasic responses of dopamine neurons to novel, non-rewarding stimuli, responses that systematically habituate with stimulus repetition. To reconcile these observations, we propose that dopamine neurons encode a more generalized signal: the total prediction error (TPE), defined as the discrepancy between the *actual* total value obtained (extrinsic plus intrinsic) and the *expected* total value.

Consider an agent encountering a novel stimulus. Even in the absence of extrinsic reward, the intrinsic reward component $R_i > 0$ generates a positive δ_{total}, triggering a phasic dopamine burst. As the stimulus is repeated, its novelty, and consequently R_i, decays to zero. Dopamine responses then revert to reflecting only extrinsic RPEs. This parsimonious hypothesis unifies the seemingly disparate responses of dopamine to both reward and novelty, and intrinsically explains the habituation profile of novelty responses.

Formally, the agent learns a total value function , representing the expected cumulative discounted total reward from state *s* under its policy:

$$V_{total}(s) = E_{\pi}\left[\sum_{k=0}^{\infty} \gamma^k R_{total}(s_{t+k}, a_{t+k}, s_{t+1}) \middle| s_t = s \right]$$

with discount factor $\gamma \in [0,1)$. This value function encapsulates the overall "goodness" of a state, integrating both extrinsic and intrinsic prospects.

Employing temporal-difference learning, the total prediction error is:

$$\delta_{total} = R_{total}(s,a,s') + \gamma V_{total}(s') - V_{total}(s)$$

This signal measures how much the obtained total reward, plus the discounted value of the subsequent state, deviates from the current estimate of the state's value. A positive δ_{total} indicates a better-than-expected outcome, prompting an upward revision of the state-action value, and vice-versa. Our core proposal is that the phasic firing rate of midbrain dopamine neurons is proportional to δ_{total}. This recasts dopamine from a specialized RPE encoder to a generalized encoder of expectation violation, whose activity reflects surprise relative to a composite prediction encompassing both informational and rewarding outcomes.

3.3 Neural Substrates for Generating and Utilizing Total Prediction Error

The hippocampus constitutes a primary locus for computing intrinsic reward. The Lisman-Grace model provides a foundational circuit: hippocampal CA1, by comparing current sensory input with predictions originating from CA3, detects "mismatch" indicative of novelty. Enhanced CA1 output following mismatch detection can, via the subiculum and the shell of the nucleus accumbens, indirectly activate VTA dopamine neurons [8]. This pathway furnishes a biological substrate for generating intrinsic reward signals. Furthermore, sparse coding mechanisms in the dentate gyrus may subserve state novelty estimation via pattern separation, while the mismatch detection in

CA1 naturally instantiates a form of prediction error computation.

The novelty signals generated by the hippocampus are subject to dynamic regulation by the prefrontal cortex. The PFC integrates a broad array of information: interoceptive signals from the hypothalamus reflecting homeostatic state, statistical regularities of the environment from hippocampal and cortical afferents, and recent exploration history from basal ganglia loops. Through its projections to the VTA, the PFC modulates the sensitivity of dopamine neurons to hippocampal novelty inputs. Excitatory PFC outputs can effectively increase the weighting factor η, promoting exploration, while inhibitory influences, potentially via VTA GABAergic interneurons, can decrease η, favoring exploitation. This circuitry provides the basis for dynamically balancing exploration and exploitation as a function of integrated internal and external demands.

Novelty signals, generated by the hippocampus and gated by the PFC, ultimately converge with information about extrinsic reward prediction errors within dopamine neurons of the VTA and substantia nigra pars compacta (SNc). These neurons receive dual inputs: extrinsic RPE signals from sources such as the pedunculopontine tegmental nucleus, and novelty signals modulated by the PFC. Through dendritic integration mechanisms, these inputs are combined and converted into a phasic firing rate output proportional to δ_{total}, effectively realizing the neural encoding of total prediction error.

Finally, the striatum, as a principal projection target of midbrain dopamine neurons, translates the TPE signal into behavioral adaptation. The dorsomedial striatum is implicated in storing state-action values based on total value, guiding goal-directed behavior. The dorsolateral striatum supports the formation of habits via more automatic stimulus-response mappings. Phasic dopamine release, by modulating synaptic plasticity at corticostriatal synapses onto medium spiny neurons, serves to update these value representations based on δ_{total}. Through this mechanism, intrinsic motivation, triggered by novelty and operationalized as positive TPE, is converted into concrete exploratory actions. As an environment becomes familiar and intrinsic reward decays, the behavioral policy gradually transitions from exploration to more habitual, efficient exploitation.

4. Mechanistic Insights and Translational Implications

4.1 Phasic Transitions in the Exploration-Exploitation Balance

The proposed dopaminergic mechanism encoding total prediction error (TPE) exhibits characteristic temporal dynamics that subserve the adaptive balance between exploration and exploitation. Upon initial exposure to a novel environment, the hippocampus detects a pronounced mismatch between sensory input and existing predictions, thereby generating a robust intrinsic reward signal (R_i). In the absence of prior knowledge, this intrinsic component dominates the total value computation, corresponding to an elevated exploratory weighting factor (η). This configuration yields a substantial positive TPE (δ_{total}), triggering vigorous phasic bursting in VTA dopamine neurons. The resulting behavioral repertoire encompasses active exploration—manifesting as attentional orienting, approach tendencies, and diversified action selection. At the neural level, this phase corresponds to the well-documented initial dopaminergic surge in response to novelty.

As sampling proceeds, the organism's internal model of the environment is progressively refined. Concomitant reductions in forward model prediction errors, coupled with accumulating state visitation frequencies (N(s)), lead to the gradual attenuation of R_i as a function of increasing familiarity. Consequently, the novelty-driven component of δ_{total} diminishes, and dopaminergic responses to the now-familiar stimulus exhibit systematic habituation. Once the environment no longer yields significant informational gains, the marginal utility of continued exploratory investment declines, rendering further sampling suboptimal.

Following comprehensive environmental learning, R_i approaches zero, and the total reward function effectively collapses to $R_{total} \approx R_{ext}$. Behavioral control accordingly shifts from intrinsic motivation to extrinsic reward pursuit. If the environment affords stable reinforcement contingencies, the organism transitions to habitual exploitation—executing well-learned actions with minimal cognitive overhead. At the neural level, phasic dopamine activity reverts to encoding exclusively extrinsic reward prediction errors, while behavioral control migrates from the goal-directed dorsomedial striatal circuit to the habit-based dorsolateral striatal system.

Importantly, should a familiar environment undergo significant perturbation, the hippocampus detects renewed mismatch between current input and established predictions, thereby reactivating R_i. This novelty reactivation mechanism reinitiates the exploratory cycle: dopaminergic phasic responses are reinstated, and behavior flexibly switches back from habitual exploitation to active exploration.

4.2 Empirically Testable Predictions and Ex-

ISSN 2959-409X

perimental Paradigms

The proposed framework generates distinct, falsifiable predictions across circuit, cellular, and behavioral levels of analysis.

At the circuit level, the hippocampal projection to VTA via the subiculum-nucleus accumbens pathway constitutes a critical substrate for novelty-driven exploration. Selective disruption of this pathway—via targeted lesions or chemogenetic inactivation—should therefore abolish novelty-induced exploratory behavior while sparing extrinsically motivated responding. This prediction can be tested using a paradigm combining novel object recognition with operant conditioning: animals with hippocampal-VTA pathway lesions are predicted to exhibit significantly reduced exploration of novel objects compared to controls, yet demonstrate intact operant responding for food reward.

At the cellular level, phasic dopaminergic responses to a novel stimulus should decay exponentially with repeated exposure, with a time constant matching the behavioral habituation rate. This can be empirically examined using in vivo electrophysiological recordings from VTA dopamine neurons during a novelty habituation paradigm. The predicted tight coupling between the decay kinetics of neural responses and exploratory behavior can be quantified. Furthermore, dopaminergic response magnitude should scale parametrically with either state novelty or the amplitude of prediction error, depending on the specific operationalization of R_i.

At the systems level, prefrontal cortical (PFC) integration of multimodal information for dynamic η regulation predicts that PFC lesions should produce characteristic dysregulation patterns. Two principal phenotypes are anticipated: (1) persistently elevated η, resulting in maladaptive over-exploration of familiar environments and inefficient resource exploitation; or (2) tonically suppressed η, yielding blunted novelty responsiveness and diminished exploratory drive. The specific manifestation may depend on lesion localization, with orbitofrontal lesions potentially impacting reward value integration and medial PFC lesions disrupting uncertainty assessment.

4.3 A Computational Perspective on Neuropsychiatric Dysfunction

This framework offers a unified computational lens for interpreting dopaminergic dysfunction across diverse neuropsychiatric conditions.

Schizophrenia—particularly its negative symptom constellation encompassing anhedonia, social withdrawal, and cognitive rigidity—aligns with deficits in intrinsic reward processing. The Lisman-Grace model implicates hippocampal pathology, notably CA1 hyperexcitability, in disrupting novelty detection mechanisms [9]. Within our framework, such disruption translates to impaired R_i generation: affected individuals fail to derive normal informational value from novelty, consequently lacking curiosity-driven exploratory motivation [10]. Cognitive rigidity can be understood as a consequence of diminished intrinsically motivated exploration, resulting in behavioral perseveration on a limited repertoire of strategies and inflexible adaptation to environmental change.

Major depressive disorder is characterized by profound anhedonia and diminished interest in novel stimuli, potentially reflecting aberrant regulation of the exploratory weighting factor η. Under physiological conditions, PFC circuits dynamically calibrate η based on integrated homeostatic signals: upregulation during periods of energetic sufficiency to encourage exploration, downregulation during depletion to prioritize exploitation of known resources. The homeostatic disturbances common in depression—including appetite and sleep dysregulation—may disrupt this PFC-mediated modulation, resulting in chronically suppressed η and failure to translate potential informational value into effective exploratory motivation. Anhedonia may additionally reflect impaired updating of the total value function due to deficient translation of prediction errors into value modifications.

Anxiety disorders feature prominent hypersensitivity to uncertainty, which may correspond to pathologically elevated η or excessive amplification of novelty signals. When R_i gain is inappropriately high, even minor environmental uncertainties are assigned exaggerated informational value, trapping the organism in a persistent exploratory state and precluding normal transition to exploitation. Phenomena including psychomotor agitation, hypervigilance, and compulsive checking behaviors can be conceptualized as manifestations of this exploration-exploitation imbalance, wherein the individual remains fixated in "exploration mode," unable to achieve habituation-based safety learning. The apparent paradox between avoidance behavior and over-exploration in anxiety disorders may reflect a unified mechanism: when the intrinsic reward of exploration is overshadowed by negative outcome expectancies, avoidance emerges as a maladaptive strategy for coping with intolerable uncertainty.

Substance use disorders represent an extreme case of hijacked intrinsic motivation circuitry, wherein exogenous compounds directly stimulate dopamine release, bypassing the normative δ_{total} computation and generating supraphysiological signals. This pharmacological intervention produces dual pathological consequences: (1) drug-asso-

ciated cues acquire pathologically inflated value, driving maladaptive seeking behavior; and (2) naturally occurring novelty signals are relatively attenuated, rendering affected individuals hyposensitive to intrinsic informational value in the normal environment. From a computational standpoint, addiction fundamentally distorts TPE computation, co-opting a signal that ordinarily mediates adaptive exploration into a driver of pathological drug craving, thereby narrowing the behavioral repertoire from flexible exploration to compulsive drug-seeking.

5. Conclusion

Curiosity-driven learning constitutes a defining attribute of biological intelligence, manifesting in the propensity of organisms to actively explore their surroundings even in the absence of extrinsic reinforcement. This investigation synthesizes neurobiological evidence with reinforcement learning theory to advance a unified computational architecture that reconceptualizes dopaminergic signaling as encoding "total prediction error." Within this framework, behavioral selection operates to maximize a composite value function integrating both extrinsic rewards and intrinsic informational utility. The latter—intrinsic reward—may be operationalized through novelty metrics or prediction error signals, with the exploration-exploitation balance dynamically modulated by prefrontal cortical circuits integrating homeostatic demands and environmental uncertainty estimates. Phasic activity of midbrain dopamine neurons is proposed to encode total prediction error, thereby extending the canonical interpretation of dopamine from a specialized reward prediction error signal to a more generalized encoder of expectation violation. At the neural implementation level, the hippocampus generates intrinsic reward signals, the prefrontal cortex dynamically calibrates exploratory weighting, ventral tegmental area dopamine neurons integrate these streams to compute total prediction error, and the striatum translates this composite error signal into adaptive behavioral selection. Our framework elucidates the temporal dynamics characterizing exploration-exploitation transitions, generates empirically testable predictions across multiple levels of analysis, and offers a coherent computational account for diverse neuropsychiatric conditions including schizophrenia, major depressive disorder, anxiety spectrum conditions, and substance use disorders.

Future investigations employing optogenetic manipulations, in vivo electrophysiological recordings, and targeted behavioral assays can empirically validate the core predictions emerging from this framework. Extending the model to encompass more complex cognitive domains—such as social learning and creative cognition—represents a promising avenue for theoretical development. Elucidating the fine-grained temporal dynamics of dopamine release and its functional differentiation across distinct projection pathways will refine our understanding of its computational roles. Finally, developing meta-learning algorithms capable of dynamically adjusting exploratory weighting may advance the engineering of human-like autonomous learning agents.

References

[1] Schultz, W. (2019). Recent advances in understanding the role of phasic dopamine activity. F1000Research, 8, F1000-Faculty.

[2] Spanagel, R., & Weiss, F. (1999). The dopamine hypothesis of reward: past and current status. Trends in neurosciences, 22(11), 521-527.

[3] Part, A. (2000). CURRICULUM VITAE (CVA). Computer Science, 79, 164.

[4] LIU Huimin. (2024). Exploration situation and development strategy of Shengli Oilfield during“ 14th Five-Year Plan in 2021-2025 ”[J]. Petroleum Geology and Recovery Efficiency, 31(4):1~12

[5] Chirimuuta, M. (2014). Minimal models and canonical neural computations: The distinctness of computational explanation in neuroscience. Synthese, 191(2), 127-153.

[6] Barto, A. G. (2012). Intrinsic motivation and reinforcement learning. In Intrinsically motivated learning in natural and artificial systems (pp. 17-47). Berlin, Heidelberg: Springer Berlin Heidelberg.

[7] Zhu, Y., & Zhu, S. C. (2026). Utility. In Computer Vision: Cognitive Models for Visual Commonsense (pp. 213-242). Cham: Springer Nature Switzerland.

[8] Perez, S. M., & Lodge, D. J. (2018). Convergent inputs from the hippocampus and thalamus to the nucleus accumbens regulate dopamine neuron activity. Journal of Neuroscience, 38(50), 10607-10618.

[9] Targa Dias Anastacio, H., Matosin, N., & Ooi, L. (2022). Neuronal hyperexcitability in Alzheimer’s disease: what are the drivers behind this aberrant phenotype?. Translational psychiatry, 12(1), 257.

[10] Salaka, R. J., Nair, K. P., Annamalai, K., Srikumar, B. N., Kutty, B. M., & Shankaranarayana Rao, B. S. (2021). Enriched environment ameliorates chronic temporal lobe epilepsy-induced behavioral hyperexcitability and restores synaptic plasticity in CA3–CA1 synapses in male Wistar rats. Journal of Neuroscience Research, 99(6), 1646-1665.