ASIAN JOURNAL OF ELECTRICAL AND ELECTRONIC ENGINEERING
Vol. 6 No. 1 2026
E-ISSN: 2785-8189

Hybrid Deep Reinforcement Learning for RIS-Assisted 6G IoT Networks: A Comparative Study of DDPG, TD3, and Adaptive Hybrid Policies

Ezdihar Osman Taj Almowla Mohomad[1*], Khalid Hamid Bilal[2], Zeinab Mahmoud Omer[1], Eltaf Abdalsalam Mohamed[3], and Rania Ali Elkhidir[4]

[1]*University of Bahri Khartoum, Sudan*
[2]*University of Science & Technology, Omdurman*, Sudan
[3]*Blue Nile University, Damazee, Sudan*
[4]*University of Hail, Saudi Arabia*

*Corresponding author: Ezdiharosman22@gmail.com

(Received: 24 February 2026; Accepted: 25 May 2026)

Abstract—Reconfigurable Intelligent Surfaces (RIS) are becoming essential for improving coverage and signal reliability in the upcoming 6G wireless networks. In this research, we present a hybrid deep reinforcement learning (DRL) framework to optimize beamforming in RIS-assisted systems. It employs several policy-selection mechanisms to enhance performance stability and reward consistency under dynamic channel conditions. The extensive experimental evaluation over the last 30 testing episodes used trimmed-mean analysis with 95% confidence intervals.

With an average return of 14.02 ± 0.62, the Hybrid Best-Action method outperformed the conventional DDPG baseline (11.13 ± 0.87) by 25.97%, which is marked by a significant effect size (Cohen's d = 1.42, $p < 0.0001$). Although TD3 achieved a competitive performance (13.34 ± 0.73), the hybrid strategy outperformed it in reward stability and reduced performance variability, which is evidenced by the rolling standard deviation analysis. The results of a one-way ANOVA showed statistically significant differences among all the policies assessed ($F(4,145) = 9.8$, $p < 0.0001$), indicating a considerable overall effect size ($\eta^2 \approx 0.213$).

A post hoc power analysis indicates that the sample size we selected (n = 30 per policy) has high statistical power (>0.99) to detect moderate-to-large performance differences. In RIS-assisted IoT applications, the proposed hybrid framework's reduced reward variability and improved convergence stability are key factors that enhance the reliability of beam alignment and ensure consistent coverage. The findings show that the proposed hybrid DRL method yields statistically significant and practically meaningful improvements over traditional single-policy methods, making it a strong contender for intelligent beamforming optimization in dynamic 6G wireless environments.

***Keywords*:** Reconfigurable Intelligent Surface (RIS), Deep Reinforcement Learning (DRL), DDPG, TD3. Hybrid Learning, Beamforming Optimization,6G Networks, IoT Coverage Enhancement.

1. INTRODUCTION

The continuous development of wireless communication systems has been driven by the rapid growth of data-intensive and latency-sensitive applications [1]. Despite significant advancements in fifth-generation (5G) networks regarding enhanced mobile broadband, ultra-reliable low-latency communications, and extensive machine-type communications, they encounter increasing challenges from emerging use cases such as large-scale Internet of Things (IoT) [2][3][4], autonomous systems, intelligent sensing, and immersive applications. The evolving requirements drive the transition to sixth-generation (6G) wireless networks, which aim to provide

DOI: https://doi.org/10.69955/ajoeee.2026.v6i1.86

extensive coverage, outstanding reliability, and intrinsic intelligence at both the physical and network levels [5-7].

A defining characteristic of 6G is the shift from conventional communication models to wireless systems that are aware of and able to interact with their surroundings [8-11]. In contrast to 5G, which treats the propagation environment as a passive element [12-14], 6G envisions programmable environments that can actively influence electromagnetic wave propagation. Reconfigurable Intelligent Surfaces (RIS) have emerged as a critical enabling technology [15] [16] [17]. By adjusting the phase shifts of various cost-effective reflective elements, RIS can enhance signal coverage, reduce interference, and improve energy efficiency [18] [19] [20]. Achieving these benefits in dense and dynamic IoT deployments is a complex optimization challenge due to the high-dimensional control space, non-convex system behavior, and rapidly changing channel and interference circumstances.

This research aims to develop a comprehensive hybrid deep reinforcement learning framework for RIS-assisted 6G IoT networks, motivated by identified shortcomings. The primary objective is to integrate the complementary benefits of Deep Deterministic Policy Gradient (DDPG) and Twin Delayed DDPG (TD3) into hybrid decision policies, thereby enabling more stable learning, accelerating convergence, and optimizing performance with a focus on coverage in dynamic channel and interference settings. We present numerous simulation results to compare the proposed hybrid architecture with existing solo DRL methods.

2. RELATED WORK

Initial research endeavors predominantly concentrated on modeling and facilitating communications augmented by RIS. A. Taha, M. Alrabeiah, and A. Alkhateeb wrote "Enabling Large Intelligent Surfaces With Compressive Sensing and Deep Learning" to look into ways to estimate channels that employ compressive sensing and deep learning to cut down on the expense of training for large-scale RIS deployments [21]. In their paper "Reconfigurable Intelligent Surfaces for Wireless Communications: Overview of Hardware Designs, Channel Models, and Estimation Techniques" [22], Jian et al. provide a thorough overview of RIS hardware architectures, channel models, and estimation methods. This work sets the stage for RIS-assisted systems. These attempts are significant, but they don't solve the problem of long-term RIS control or making decisions when there is interference.

Later, to simplify the calculations, supervised deep learning methods were used. Yang, Liu, and Zhang used deep neural networks trained offline to predict optimal RIS phase configurations based on channel state information in their paper "A Deep Learning-Based Modelling of Reconfigurable Intelligent Surface-Assisted Wireless Communications for Phase Shift Configuration" [23]. Their method works well in static or slowly changing channels, but it requires labeled datasets and struggles to adapt to rapidly changing environments. This means it can't be used in real-world 6G IoT situations.

Reinforcement learning (RL) and deep reinforcement learning (DRL) algorithms have been used to address the challenges of RIS optimization. Zhou, Wang, and Li present a deep reinforcement learning framework in their study, "Deep Reinforcement Learning-Based Intelligent Reflecting Surface for Cooperative Jamming Model Design," to improve physical-layer security by optimizing RIS phase shifts [24]. ZUA Tariq, E Baccour, Erbad, and M Hamdi also looked into how to make wireless communication networks more resistant to jamming attacks in their study "Reinforcement Learning for Resilient Aerial-IRS-Assisted Wireless Communications Networks in the Presence of Multiple Jammers" [25]. The results demonstrated that DRL can proficiently address continuous control challenges in dynamic settings. Still, they rely on autonomous learning agents that often struggle with training instability, reward unpredictability, and overestimation bias.

In specific contexts, more utilizations of DRL-based RIS have been investigated. In "Reconfigurable Intelligent Surface-Assisted Localization: Technologies, Challenges, and the Road Ahead," MA and Teng et al. [26]. Examined RIS-enabled localization and emphasized the potential of learning-based optimization. Recent studies, such as "Reinforcement Learning-Based Intelligent Reflecting Surface Optimization for Wireless Communications" [27], employ single-agent reinforcement learning frameworks to improve system performance relative to static baselines. Even with these improvements, these methods often lack thorough statistical tests of their convergence and training stability. Recently, many have suggested using hybrid learning methods to improve performance in "Hybrid Reinforcement Learning for STAR-RISs: A Coupled Phase-Shift Model Based Beamformer" [28], J. Chen et al. presented a hybrid reinforcement learning framework designed to enhance

spectral efficiency for STAR-RIS beamforming. Their research primarily emphasizes throughput measurements, neglecting the assessment of learning stability, reward variance, or performance metrics focused on coverage in densely populated IoT contexts.

From our previous discussion, it's clear that current RIS-assisted learning approaches have major problems. Supervised learning methods aren't very flexible, single-agent deep reinforcement learning methods have problems with stability and convergence when dynamic interference is present, and hybrid approaches don't cover as much ground or go as deep in their evaluations. In the ever-changing world of 6G IoT, it is very hard to ensure that learning remains consistent, that reward variability is low, and that coverage optimization is effective.

3. SYSTEM OVERVIEW AND PROBLEM FORMULATION

3.1 System Overview

We consider a downlink RIS-assisted 6G Internet-of-Things (IoT) network consisting of a base station (BS) equipped with Mantennas, a reconfigurable intelligent surface (RIS) comprising Nnearly passive reflecting elements, and multiple single-antenna IoT devices randomly distributed within the coverage area. The RIS is deployed to enhance wireless propagation by intelligently adjusting the phase shifts of its reflecting elements. Due to blockages, severe path loss, and dense device deployment, the direct BS–IoT links may be weak or unavailable. To address this issue, the RIS establishes an indirect BS–RIS–IoT link, enabling controllable signal reflection and coverage enhancement without additional transmit power. Each RIS element applies a programmable phase shift to the incident signal, allowing the wireless environment to be dynamically reconfigured.

3.2 Channel and Signal Model

Let $\mathrm{G} \in \mathbb{C}^{N\times M}$denote the channel matrix between the BS and the RIS, $\mathrm{h}_k \in \mathbb{C}^{N\times 1}$the channel vector between the RIS and the k-th IoT device, and $\mathrm{d}_k \in \mathbb{C}^{M\times 1}$ the direct BS–IoT channel. The RIS reflection matrix was defined as in [33] using Equation (1).

$$\Phi = \mathrm{diag}\left(e^{j\theta_1}, e^{j\theta_2}, \ldots, e^{j\theta_N}\right), \qquad (1)$$

where $\theta_n \in [0,2\pi)$ represents the phase shift applied by the n-th RIS element. The received signal at the k -th IoT device was expressed by Equation (2).

$$y_k = \left(\mathrm{h}_k^H \Phi \mathrm{G} + \mathrm{d}_k^H\right)\mathrm{w}s + n_k, \qquad (2)$$

where w is the BS beamforming vector, s is the transmitted symbol with unit power, and n_k is additive white Gaussian noise with variance σ^2.

3.3 Performance Metric

The signal-to-noise ratio (SNR) at the k-th IoT device is given by

The signal-to-noise ratio (SNR) at the k-th IoT device was calculated using Equation (3).

$$\mathrm{SNR}_k = \frac{|\left(\mathrm{h}_k^H \Phi \mathrm{G} + \mathrm{d}_k^H\right)\mathrm{w}|^2}{\sigma^2}. \qquad (3)$$

Based on the SNR, the achievable rate or coverage-related reward can be derived and used as a performance indicator for RIS optimization.

3.4 Problem Formulation

The objective of the RIS-assisted system is to determine the optimal RIS phase-shift configuration that maximizes the long-term communication performance under dynamic channel conditions. This optimization problem can be formulated as [29]

$$\max_{\Phi} \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t r_t\right], \qquad (4)$$

subject to

$$\theta_n \in [0,2\pi), \forall n = 1, \dots, N \ , \qquad (5),$$

where r_tIs the reward at the time step t, $\gamma \in (0,1)$is the discount factor, and T is the episode length. Due to the high dimensionality, nonlinearity, and time-varying nature of the wireless environment, solving Equation (4) using conventional optimization techniques is computationally prohibitive. Therefore, the problem was naturally modeled as a Markov Decision Process (MDP), hence, suitable for deep reinforcement learning-based solutions.

3.5 Markov Decision Process Formulation

The tuple$(\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P})$ defines the RIS optimization problem as an MDP, where [30]

- State (S): comprises channel-related data, including received SNR, effective channel gains, and prior RIS setups.

The continuous RIS phase-shift adjustments $\theta = [\theta_1, \dots, \theta_N]$.

- Action ($\mathcal{A}$): consists of continuous control variables corresponding to the RIS phase-shift adjustments. At each decision step t, the action is defined as

$$\mathrm{a}_t = \theta_t = [\theta_{1,t}, \theta_{2,t}, \dots, \theta_{N,t}], \qquad (6)$$

where $\theta_{n,t} \in [0,2\pi)$denotes the phase shift applied by the n-th RIS element.

- Reward (R): is intended to represent performance relating to throughput or coverage.
- The stochastic evolution of wireless channels over time is represented by transition (p).
  To learn the best RIS configurations, this formulation enables the implementation of continuous-control DRL algorithms, such as DDPG, TD3, and their hybrids.

4. DRL-BASED JOINT DESIGN UNDER INTERFERENCE DYNAMIC ENVIRONMENTS

4.1 Motivation and Design Rationale

Wireless links in RIS-assisted 6G IoT networks are naturally vulnerable to significant channel fluctuations, dynamic interference, and unanticipated environmental changes due to elevated device density and mobility. These effects make coverage and signal quality much worse, making them look like general jamming-like problems. So, it's very important to develop robust RIS control rules to address these issues. Setting up RIS with traditional optimization methods isn't very flexible when interference conditions change. They also usually assume that settings don't change very often. This encourages the use of deep reinforcement learning (DRL) in situations where you don't know what's going to happen because it lets you change RIS phase shifts and transmission methods in a way that is both adaptive and doesn't require a model.

4.2 Joint DRL-Based RIS Control Framework

The suggested joint design aims to optimize RIS phase shifts to achieve optimal coverage performance in environments where interference varies over time. The DRL agent takes the same control action at every decision point, regardless of the RIS phase shift. It does this by examining the current state of the network, including the quality of the received signal, effective channel conditions, and previous RIS settings. The DRL framework automatically mitigates interference and channel degradation without directly modeling them. It does this by continually adjusting the RIS setup to meet the area's needs. This is not the same as methods for optimizing that don't change.

4.3 Continuous-Control DRL Algorithms

Two sophisticated continuous-control deep reinforcement learning methods are used in this work to address the high-dimensional and persistent aspects of RIS control:

Deep Deterministic Policy Gradient (DDPG): DDPG uses an actor–critic architecture to create deterministic policies in dynamic action spaces. When changes occur quickly, DDPG may struggle with overestimation bias and training instability, but it performs well for control tasks with multiple dimensions. Target policy smoothing, delayed policy updates, and twin critics make the Twin Delayed Deep Deterministic Policy Gradient (TD3) better than DDPG. Training is much more stable and less likely to anticipate an excessively high score because of these characteristics.

The fundamental guidelines for optimizing RIS in scenarios when interference is anticipated are these algorithms.

4.4 Hybrid DRL-Based Joint Design

This paper introduces hybrid decision-making techniques that dynamically combine the benefits of DDPG and TD3 to overcome the shortcomings of independent DRL algorithms. Three kinds of hybrid policies, namely, fixed, best action, and dynamic, were investigated:

- Fixed-α Policy:

A mix of DDPG and TD3 with a fixed- α parameter that makes both the policies do the same thing.

- Best-Action Policy:

A combination of DDPG and TD3 that picks the best course of action based on the highest estimated immediate reward at each point where a decision needs to be made.

- Dynamic-α(t) Policy:

This is a flexible hybrid policy that adjusts the mixing parameter α(t) over time based on learning stability and reward trends. This allows for dynamic policy supremacy. Because they employ multiple DRL strategies, these hybrid systems are less susceptible to channel changes or interference.

5. HYBRID DRL-BASED JOINT DESIGN ALGORITHM

5.1 Overview

This section presents the proposed hybrid deep reinforcement learning (DRL) framework for RIS-assisted 6G IoT networks characterized by variable interference. The framework uses both Deep Deterministic Policy Gradient (DDPG) and Twin Delayed Deep Deterministic Policy Gradient (TD3) to get the best results when optimizing the continuous RIS phase shift.

The agent uses both DDPG and TD3 policies to decide what to do at each decision point, based on the environment's state. A hybrid decision module then uses rules to select policies that are either already set or can be adjusted, thereby determining the final RIS configuration.

5.2 Standalone DRL Policies

1) DDPG-Based RIS Optimization

The DDPG agent has two parts: a critic network that evaluates state-action pairs and an actor network that produces continuous RIS phase-shift actions. DDPG works well for high-dimensional continuous control problems, but its performance may suffer in very dynamic channel and interference situations.

2) TD3-Based RIS Optimization

The TD3 agent builds on the DDPG architecture by using twin critic networks, delayed actor updates, and target policy smoothing. These processes make TD3 stronger in situations where interference is likely by reducing overestimation bias and stabilizing learning.

5.3 Hybrid Decision Policies

Three hybrid decision-making techniques are used to address the problems with standalone DRL policies.

- The Fixed-α Hybrid Policy

:The last RIS action is the weighted sum of the DDPG and TD3 actions

$$\mathrm{a}_t = \alpha \mathrm{a}_t^{\mathrm{DDPG}} + (1-\alpha)\mathrm{a}_t^{\mathrm{TD3}}, \quad (7)$$

.where $\alpha \in [0,1$ is the weight that doesn't change

- The Best Action Hybrid Plan

At every time step, both DDPG and TD3 suggest possible actions. The choice with the highest expected return is :made

$$\mathrm{a}_t = \arg \max_{\mathrm{a} \in \{\mathrm{a}_t^{\mathrm{DDPG}}, \mathrm{a}_t^{\mathrm{TD3}}\}} Q(s_t, \mathrm{a}), \quad (8)$$

.where $Q(\cdot)$shows the rating from the critic

- A policy that combines dynamic-α(t) with other policies

The Dynamic -$\alpha(t)$ policy modifies the mixing parameter according to reward statistics and training :advancement

$$\mathrm{a}_t = \alpha(t)\mathrm{a}_t^{\mathrm{DDPG}} + \big(1-\alpha(t)\big)\mathrm{a}_t^{\mathrm{TD3}}, \quad (9)$$

where $\alpha(t)$ chooses the strategy that works better and is more stable at each step of training.

5.4 Algorithm Description

Algorithm 1 summarizes the proposed hybrid deep reinforcement learning (DRL) optimization framework for RIS-assisted beamforming. The proposed approach combines two continuous-control DRL agents, namely DDPG and TD3, within a unified hybrid decision-making architecture. During each interaction step, both agents independently generate candidate actions based on the observed state of the wireless environment. The hybrid policy module then selects the final RIS configuration action according to the adopted hybrid selection strategy. The generated experiences are stored in replay buffers and used to update the actor–critic networks iteratively throughout the training process.

Figure 1 depicts the general flow of the proposed hybrid DRL framework. The first step in the process is environment initialisation and state observation. In this step, the wireless channel conditions and RIS-related parameters are collected from the communication environment. For RIS phase optimisation, the continuous control actions are separately output by the DDPG and TD3 agents based on the observed state. The generated actions are processed by the Hybrid Decision Module, which leverages the strengths of both DRL policies to select the most suitable RIS configuration strategy. Then, the optimal RIS phase shifts are applied in the wireless environment to enhance signal propagation and communication quality. The environment provides a reward signal and the next state observed after applying the RIS configuration, indicating the communication performance achieved in terms of SINR, coverage quality and system throughput. All collected transition samples are stored in the replay buffer and then used during the policy update stage to iteratively improve learning. This closed-loop interaction is sustained through the training process until convergence is reached.

6. SIMULATION SETUP

6.1 Simulation Environment and Dataset Description

The experimental assessment of the proposed hybrid DRL framework was performed in a simulated RIS-assisted 6G IoT scenario. The dataset used in this work was sourced from Kaggle [34], a platform that offers

publicly available datasets commonly used for benchmarking and reproducible research. The chosen dataset simulates a wireless communication scenario relevant to RIS-enabled 6G IoT networks, including performance metrics such as throughput, latency, energy consumption, and channel-related parameters across diverse channel and interference conditions. To provide an equitable and uniform comparison across all learning policies, data were collected from various simulation episodes with distinct channel realizations and interference levels. The dataset was subsequently processed and saved in NPZ format (R5_NewData_BestHybrid_Results.npz) to enable efficient loading and smooth integration with DRL algorithms, including DDPG, TD3, and the proposed hybrid policies.

The dataset, although simulation-based and derived from a publicly accessible benchmark, encompasses various channel realizations, interference levels, and performance metrics indicative of RIS-assisted IoT contexts. The main aim of this study is to conduct comparative policy evaluation in dynamic settings rather than focus on physical-layer channel modeling. The suggested hybrid framework is architecture-independent and can be easily integrated with physics-based channel models, such as ray-tracing or Deep MIMO datasets, in subsequent implementations.

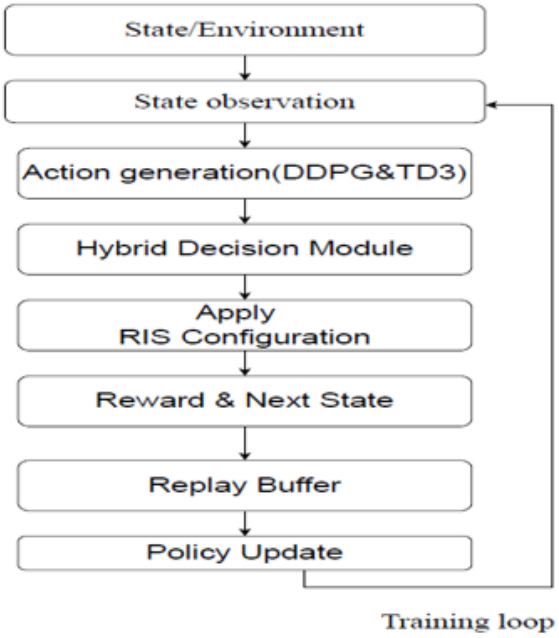

Fig. 1: Workflow of the Proposed Hybrid TD3-DDPG Framework for RIS-Assisted Beamforming Optimization.

6.2 Data Preprocessing and Enhancement

Before training the DRL models, various preprocessing and augmentation techniques were employed to enhance learning stability and mitigate data bias. Initially, partial records and non-numeric entries were discarded to ensure the dataset's uniformity. A feature selection strategy was employed to retain only the metrics pertinent to coverage performance, interference, and transmission efficiency. Third, we standardized all numerical features to a uniform range to prevent high-magnitude variables from dominating and to facilitate faster neural network convergence.

Furthermore, statistical validation was conducted across multiple training iterations to ensure the dataset's capacity to accommodate variations in channel dynamics and interference. The preprocessing approaches ensured that the performance improvements observed during training were primarily attributable to the proposed hybrid learning strategy rather than issues with the input data.

6.3 Feature Description

Table 1 summarizes the key features extracted from the dataset and utilized during the simulation and training process.

Table 1. Dataset Features Used in the Simulation

Feature	Description	Usage in DRL Framework
Throughput (Mbps)	Achievable data rate for IoT devices	Reward maximization
Latency (ms)	End-to-end transmission delay	Penalty term
Energy Consumption (kWh/GB)	Energy efficiency metric	Penalty term
Channel Gain	Composite channel including RIS reflection	State representation
Interference Level	Aggregate interference power	State representation
Noise Power	Thermal noise variance	SNR computation
Episode Return	Cumulative reward per episode	Performance evaluation

6.4 Simulation Advantages

The use of simulation-generated data rather than fixed real-world datasets enables controlled, repeatable experiments, which are essential for evaluating adaptive DRL-based optimization strategies in emerging 6G wireless environments. This setup allows systematic performance analysis under diverse conditions while maintaining experimental consistency.

6.5 Simulation Environment

The performance of the proposed hybrid DRL framework is evaluated using a simulated RIS-assisted 6G IoT environment. The considered network consists of a multi-antenna base station, a reconfigurable intelligent surface with N reflecting elements, and multiple single-antenna IoT devices randomly distributed within the coverage area. Wireless channels follow a block-fading model and vary across episodes to capture dynamic propagation and interference conditions.

All DRL agents are trained under identical environmental settings, channel realizations, and random seeds to ensure a fair and unbiased comparison among different learning policies.

6.6 Training Configuration

Both DDPG and TD3 agents use neural networks with fully connected layers. Target networks help stabilize learning, while experience replay buffers add variety to the training samples. Target policy smoothing, delayed policy updates, and two critics are all used by TD3.

The only thing that makes hybrid policies like Fixed-α, Best-Action, and Dynamic-α(t) different is how they make decisions. All of these policies use the same basic DRL networks. This design ensures that any differences in performance are due solely to the hybrid policy approach, not to changes in the network architecture.

6.7 Protocol for Training and Evaluation

The interaction data (episodes) were split into three parts: training, validation, and testing. This was done to ensure the evaluation was fair. Seventy percent of the episodes were used for policy learning (training), fifteen percent for hyperparameter selection (validation) without changing the agent, and the remaining fifteen percent for final reporting (testing). Also, the assessment was conducted using specific random seeds to reduce variance

and improve reliability. We used trimmed-mean statistics and 95% confidence intervals to assess the final performance across K = 30 testing episodes.

In line with standard statistical power analysis recommendations for identifying moderate-to-large effect sizes in comparative learning studies, we chose K = 30 independent testing episodes to ensure our results are statistically reliable. The post-hoc power analysis we carried out has confirmed that the sample size we used provides statistical power over 0.99, which means it can robustly estimate performance and minimize the chance of a Type-II error.

6.8 Baseline Schemes

The proposed framework is compared against the following baseline schemes:

- No-RIS Configuration: RIS elements are disabled or configured with random phase shifts.
- DDPG-Based Optimization: RIS phase shifts are optimized using standalone DDPG.
- TD3-Based Optimization: RIS phase shifts are optimized using standalone TD3.

These baselines enable systematic evaluation of the performance gains achieved by hybrid decision policies.

6.9 Performance Metrics

We use the following metrics to see how strong a system is and how well it learns:

- The Episode-Based Cumulative Reward shows how well coverage works over time.
- Training Stability: how well things fit together and how much the reward changes.
- The statistical performance is based on how well the rewards are spread out over the episodes.
  These measures, when taken together, show that the agent can learn and adapt to new situations.

7. SIMULATION RESULTS

This section provides a comprehensive performance review of the proposed hybrid deep reinforcement learning (DRL) framework for RIS-enabled 6G IoT networks. The evaluation centers on learning behavior, convergence traits, statistical robustness, and a conclusive performance comparison with standalone DDPG and TD3 benchmarks. To ensure fairness, all schemes are trained and tested under the same simulation settings.

7.1 *Learning Behavior and Convergence Analysis*

Figure 2 shows how the episodic Reward-Trajectories changed over the last 30 evaluation episodes of different reinforcement learning methods. The Hybrid Best-Action-Approach consistently yields the highest rewards over many episodes, demonstrating strong peak performance and greater reward stability. The method seems to work better for handling changes in the channel because it has less oscillation and a more stable path. TD3 and Hybrid Dynamic α(t) both learn quickly, but they differ significantly. However, they don't always get as high of returns as the Hybrid-Best-Action-Method. DDPG doesn't work very well, which quickly degrades rewards. This means the system doesn't adapt quickly to environmental changes, and that beamforming optimization doesn't always work. The Hybrid-Best-Action-Approach reduces randomness, helping the RIS determine the best phase shift. This quickly improves signal coverage and keeps the beam alignment stable. As we move into the 6G-IoT era, factors such as channel fading and mobility could affect performance. To get the best results and reliable coverage, it is important to keep things stable. The episodic behavior indicates that the proposed hybrid action-selection mechanism enhances both optimization efficiency and operational resilience. This makes it a great choice for building advanced wireless networks that use RIS.

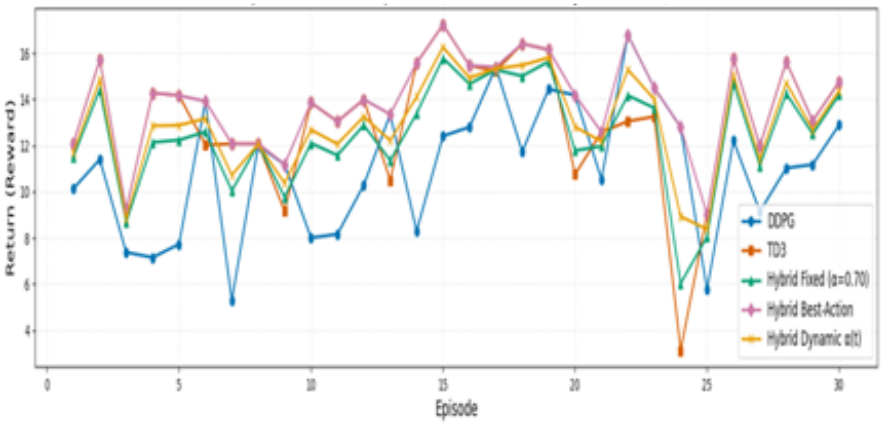

Fig. 2. Episodic reward evolution of evaluated DRL and hybrid policies over the final 30 testing episodes.

7.2 Training Stability Analysis

Figure 3 shows how the rolling standard deviation of returns changes over time. This means that even though the training is the same, the benefits are different in each case. A smaller rolling fluctuation indicates a more stable regulatory environment and a more accurate assessment, especially when channel conditions change.

The Hybrid Best-Action method reduces evaluation variability, making the learning process more stable. TD3 is more susceptible to environmental variations because it is difficult to predict outcomes across sessions. During the first and middle stages of training, DDPG exhibits significant instability. This lengthens the acclimatization period, making it more likely that beamforming changes will fail if not done correctly.

Several studies show that making incentives less random leads to more consistent RIS phase-change choices and more reliable beam alignment. Policy behavior must remain stable in the dynamic 6G IoT environments characterized by channel fading, user mobility, and variable interference. To avoid sudden drops in performance, it's important to ensure that signal coverage is always available. The Hybrid Best-Action approach offers better coverage reliability, improved power distribution, and greater resilience in real RIS-assisted wireless networks, as evidenced by its lower rolling variation.

The stability analysis corroborates the statistical results, showing that the proposed hybrid reinforcement learning framework improves average performance while significantly reducing optimization volatility. This is very important for the practical use of modern smart communication systems to work well.

Fig. 3. Training stability analysis based on rolling standard deviation of episodic returns.

Figure 4 illustrates the distribution of episodic returns for all assessed reinforcement learning policies using a

boxplot. The Hybrid Best-Action strategy demonstrates the highest median return and a reasonably narrow interquartile range, signifying both exceptional central tendency and enhanced performance stability. The distribution is concentrated in the high-reward region, indicating continuous optimization behavior throughout evaluation episodes.

TD3 and Hybrid Dynamic α(t) exhibit similar median performance; however, their broader interquartile ranges indicate marginally greater variability than the Hybrid Best-Action method. Conversely, DDPG has a lower median and higher variance, characterized by significant performance variability and lower-bound outliers, indicating reduced resilience under dynamic channel conditions. The occurrence of sporadic lower outliers in particular policies underscores vulnerability to individual channel realizations, while the Hybrid Best-Action technique ensures more stable high-return results. The boxplot analysis corroborates the statistical findings, demonstrating that the proposed hybrid mechanism yields higher rewards and greater stability in RIS-assisted beamforming optimization contexts.

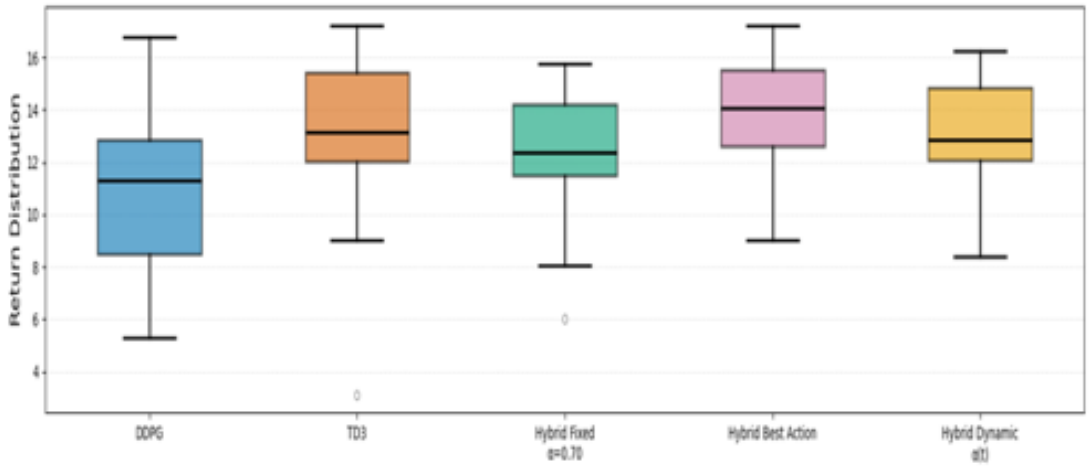

Fig. 4. Distribution of episodic returns for evaluated DRL and hybrid policies (boxplot representation)

7.3 . Statistical Performance and Robustness

Figure 4 shows a box plot of the distribution of returns from each episode. The results show that DDPG has the widest interquartile range and a few low-return outliers. This means that the person doesn't always learn the same way. TD3 limits the range of rewards compared to DDPG, but there is still a lot of difference.

Hybrid policies are better because their return distributions are much narrower. The Best-Action policy differs from the others because it has the highest median return and the smallest spread, indicating greater stability over time. The Dynamic-α(t) policy keeps the distribution small, but it always beats DRL algorithms that don't work together.

7.4 Final Performance Comparison

Figure 5 shows that the Hybrid Best-Action policy has the highest average return among the strategies tested, and its confidence intervals are not very wide. This means that the performance is more consistent. TD3 and Hybrid Dynamic α(t) both perform well, but their average returns are still slightly lower than those of the proposed hybrid approach. DDPG, on the other hand, has the lowest average return and the widest confidence interval. This means that it is less stable and less good at finding the best solution.
The fact that the confidence intervals for Hybrid Best-Action and DDPG don't overlap visually supports the statistically significant improvement indicated by the t-tests and ANOVA. The hybrid method also has tighter error bars, which means that the rewards are more stable from one evaluation episode to the next. This shows that the method still works well even when channel conditions change.
The figure shows that, in general, the proposed Hybrid Best-Action strategy works better and is more stable than standard DRL baselines.

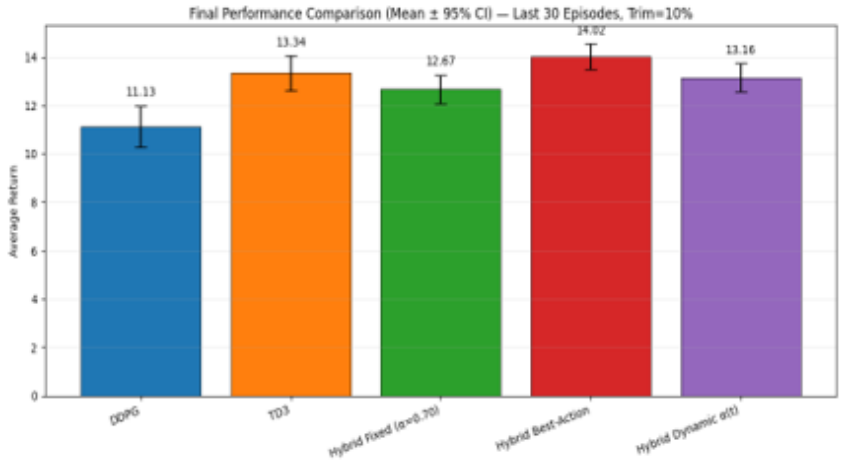

Figure 5. Episode-Based Return Comparison of Standalone and Hybrid DRL Policies in RIS-Assisted 6G IoT Environment.

Table 2. Comparative Performance and Statistical Analysis of Evaluated Policies

Policy	Mean	95% CI	Rel. Improve (%)	Cohen's d	p-value	K
DDPG	**11.13**	**±0.87**	**0%**	**—**	**—**	**30**
TD3	**13.34**	**±0.73**	**+19.86%**	**1.03**	**< 0.001**	**30**
Hybrid Fixed-α	**12.67**	**±0.66**	**+13.84%**	**0.75**	**0.005**	**30**
Hybrid Best-Action	**14.02**	**±0.62**	**+25.97%**	**1.42**	**< 0.0001**	**30**
Hybrid Dynamic-α(t)	**13.16**	**±0.71**	**+18.24%**	**0.95**	**< 0.001**	**30**

According to Table 2, all contemporary DRL and hybrid approaches significantly outperform DDPG. The Hybrid Best-Action technique outperforms the other method both statistically and practically, boasting a remarkable relative improvement of 25.97% and a significant effect size (d=1.42).

7.5 Computational Cost Analysis

To ensure a fair comparison, training time was recorded using Python's time module under the same hardware conditions. With fusion taking place at the action-selection level without adding extra deep network layers, the hybrid policy incurs minimal computational overhead. We assessed the computational efficiency of the suggested DRL algorithms by evaluating the wall-clock training duration under identical experimental settings (Intel i7 CPU, 25,000 timesteps). Table 3 presents an overview of the findings. DDPG required 7.309 minutes, while TD3 required 7.543 minutes, representing a marginal 3.20% increase in training time. The Hybrid (Best-Action) strategy required 14.874 minutes, approximately doubling the computational cost due to the sequential training of both DDPG and TD3 components. Despite the increased cost, the hybrid approach offers improved decision robustness, indicating a trade-off between computational efficiency and performance stability.

Table 3. Comparative Analysis of Training Duration under Identical Experimental Conditions

Algorithm	Training Time (min)	Relative to DDPG
DDPG	7.309	—
TD3	7.543	+3.20%
Hybrid (Best-Action)	14.874	+103.51%

Although the proposed framework was evaluated in a simulation-based environment, it can be extended to practical wireless deployment scenarios using realistic channel generation frameworks such as DeepMIMO and ray-tracing-based propagation models. Integrating the proposed hybrid DRL framework with real-world channel measurements may further improve the robustness and practical applicability of RIS-assisted beamforming optimization in future 6G IoT networks.

8. DISCUSSION

Using the data in Figures 1-5, this section examines how well the hybrid DRL-based RIS optimization framework performs, how stable it is, and how robust it is.

Figure 1 illustrates how the proposed hybrid learning system operates. The system helps TD3 and DDPG agents make decisions by showing them the RIS-assisted wireless environment as a set of choices. The hybrid decision module is responsible for these things. It picks the best control strategy or combines them before setting up RIS. This method promotes adaptive learning and enhances performance in subsequent discoveries without directly replicating interference.

Figures 2 and 4 provide a detailed comparison of episode-based return performance among the independent DDPG, independent TD3, and the proposed hybrid decision policies: Fixed-α, Best-Action, and Dynamic-α(t). The results unequivocally indicate that hybrid strategies consistently yield higher average returns and greater learning stability than DRL-only optimization techniques .The performance trends we discovered correspond with those noted in [31], where the authors examined deep reinforcement learning-based passive beamforming for RIS-assisted wireless networks. The research indicated that TD3 has enhanced stability and superior convergence relative to DDPG, attributable to its twin-critic architecture and reduced overestimation bias. Nonetheless, [31] revealed considerable reward variability under highly dynamic channel and interference settings, highlighting the inadequacies of relying exclusively on a singular learning agent. Figure 2 demonstrates that TD3 achieves higher returns than DDPG, yet it still exhibits variability in performance across episodes .The proposed hybrid decision rules improve TD3-based optimization by dynamically incorporating complementary learning attributes. Figure 2 shows that the boxplot analysis indicates that hybrid strategies display superior median returns and reduced interquartile ranges, implying lower variability and greater robustness. This improvement aligns with the results reported in [32], which established a hybrid reinforcement learning framework to enhance STAR-RIS beamforming. The authors of [32] claim that the convergence rate and stability of hybrid learning processes can be significantly enhanced by addressing the shortcomings of individual DRL algorithms .Unlike [31], which focused on improving stability within a single-agent TD3 framework, and [32], which combined hybrid learning into STAR-RIS systems, the proposed framework advances the field by incorporating hybrid decision-making processes tailored to RIS-assisted 6G IoT environments. Figure 2 demonstrates that the Best-Action and Dynamic-α(t) hybrid policies produce enhanced episode-based returns and display increased stability in learning behavior, particularly in environments prone to interference and marked by fast propagation shifts .The comparison with [31] and [32] illustrates that, unlike previous studies that emphasize specific deficiencies of individual DRL methods, our proposed hybrid framework concurrently improves return maximization, training stability, and resilience. The aforementioned qualities make the proposed method particularly suitable for practical implementation in the upcoming 6G IoT networks.

Figure 5 juxtaposes the learning behaviors of the assessed policies across the training episodes. Although TD3 demonstrates greater stability than DDPG, both approaches exhibit performance variability in dynamic channel and interference settings. The suggested hybrid decision procedures demonstrate accelerated convergence and consistently superior returns in the final training period.

Table 2 presents a quantitative assessment of the convergent performance, utilizing the mean return and the 95% confidence interval for the final K episodes. The Hybrid Best-Action policy achieves the highest mean return with the tightest confidence interval, validating its exceptional stability and reliability. The alignment between the visual patterns in Figure 5 and the statistical outcomes in Table 2 corroborates the efficacy of the suggested hybrid framework.

These findings align with recent studies on DRL-based RIS. Yuan et al. [31] demonstrated that single-agent DRL can effectively optimize RIS in dynamic settings. In contrast, Jian et al. [32] highlighted the pronounced susceptibility of RIS-assisted systems to channel fluctuations and hardware characteristics. The proposed hybrid

framework promotes resilience by integrating complementary learning tendencies, resulting in improved convergence and more stable performance.

Figure 3 depicts the rolling standard deviation of episode returns to assess training stability. The findings demonstrate that DDPG exhibits significant variance, whereas TD3 offers some improvement but remains susceptible to dynamic disturbances, as noted in references [24] and [25]. The proposed hybrid policy exhibits consistently lower variation by integrating the exploration capabilities of DDPG with the precise value estimation of TD3, hence ensuring more stable and predictable learning behavior in RIS-assisted wireless contexts.

In short, the analysis of Figures 1 to 5 shows that the proposed hybrid DRL architecture improves coverage performance, accelerates convergence, enhances system stability and strikes the best balance between exploration and exploitation. The hybrid RIS optimization method works best in 6G IoT settings, which are known for their slow transitions and susceptibility to interference. Sometimes, DRL methods alone won't be enough to fix the problem

The proposed Hybrid Best-Action reinforcement learning framework for optimizing RIS-assisted beamforming performs effectively, as demonstrated by both experimental and statistical evidence. By systematically incorporating a variety of learning behaviors, the proposed methodology improves decision-making resilience in dynamic channel environments, unlike traditional single-policy deep reinforcement learning systems. The key enhancements over DDPG, along with the evidence that it performs as well as TD3, indicate that hybrid action selection can boost stability and performance without adding complexity to the model.

The ANOVA and power analysis demonstrate that the performance improvements we've observed are statistically significant and not attributable to sample variation. The proposed strategy's real-world usefulness is underscored by its large effect size compared with traditional baselines, especially when the goals are to stabilize coverage and optimize constant rewards.

Taking a systemic approach, the immediate advantages of stabilizing rewards are improved beam alignment reliability, more consistent signal coverage, and reduced performance variability in RIS-assisted 6G IoT environments. The proposed hybrid DRL approach is a significant advancement in improving the performance of intelligent and adaptive wireless networks.

We chose DDPG and TD3 because they are effective at solving optimisation problems with continuous actions in wireless environments aided by RIS. The presented work intentionally focuses on deterministic continuous-control DRL algorithms, as the RIS phase optimisation naturally constitutes a continuous decision-making problem. It is important to further enrich the comparative analysis by exploring other DRL variants, such as PPO and SAC, which represent an important direction for future research.

9. CONCLUSION AND FUTURE WORK

This study evaluated the effectiveness of hybrid deep reinforcement learning methods in enhancing beamforming in dynamic wireless environments facilitated by RIS. An extensive experimental evaluation was conducted to analyze conventional Deep Reinforcement Learning baselines (DDPG and TD3) in conjunction with several hybridization techniques, including fixed-weight, dynamic-weight, and best-action selection algorithms. The results show that the Hybrid Best-Action method yields the highest average return and is more stable than the other methods examined. Statistical validation shows that this method works much better than DDPG and has a big effect size, while still being competitive with TD3. The one-way ANOVA test shows that the differences in performance among the policies are not random; they are real. A post hoc power analysis indicates that the sample size is sufficient to detect moderate-to-large improvements in performance.

The Hybrid Best-Action method makes it easier to estimate how well something will work and reduces randomness. This makes beamforming more reliable and ensures that IoT networks using RIS achieve better coverage. Next-generation 6G systems require specialized optimization frameworks that can accommodate fluctuating channel conditions, user mobility, and energy-efficiency demands while maintaining statistical integrity.

This study demonstrates that structured hybrid reinforcement learning techniques, enhanced with statistical analysis, significantly outperform conventional single-policy deep reinforcement learning methods for optimizing smart wireless networks.

Future research directions include improving the proposed hybrid framework for scenarios with many users and multiple RIS, exploring ways to make large-antenna designs more scalable, and incorporating energy-efficiency constraints into the optimization objectives. Federated or distributed learning techniques may enhance the adaptability of decentralized 6G systems.

The proposed system, in contrast to traditional ensemble learning methods, features an adaptive decision-level hybridization mechanism that dynamically selects or amalgamates DRL policies based on reward stability and performance assessment. This organized action-level fusion diminishes reward variance while maintaining convergence speed, providing a balanced exploration-exploitation trade-off designed for interference-prone 6G IoT environments. The work thus transcends policy aggregation by delivering statistically proven stability improvements in continuous-control RIS optimization, Channel models, including ray-tracing and DeepMIMO datasets, in forthcoming implementations.

ACKNOWLEDGMENT

The author sincerely thanks Prof. Khalid Hamid Bilal for his dedicated supervision, technical advice, and valuable critiques, all of which have greatly enhanced the quality and rigour of this work.

REFERENCES

1. S. Shukla, Mohd. F. Hassan, D. C. Tran, R. Akbar, I. V. Paputungan, and M. K. Khan, 'Improving latency in Internet-of-Things and cloud computing for real-time data transmission: a systematic literature review (SLR)', Cluster Comput, vol. 26, no. 5, pp. 2657-2680, Oct. 2023, https://doi.org/10.1007/s10586-021-03279-3
2. 'The Evolution of Mobile Communication: A Comprehensive Survey on 5G Technology', J Sen Net Data Comm, vol. 4, no. 1, pp. 01-11, Mar. 2024, https://doi.org/10.33140/JSNDC.04.01.06
3. N. Lassoued and N. Boujnah, 'A Comprehensive Review of Energy Efficiency in 5G Networks: Past Strategies, Present Advances, and Future Research Directions', Computers, vol. 15, no. 1, Jan. 2026, https://doi.org/10.3390/computers15010050
4. K. Dulaj, A. Alhammadi, I. Shayea, A. A. El-Saleh, and M. Alnakhli, 'Harnessing Machine Learning for Intelligent Networking in 5G Technology and Beyond: Advancements, Applications and Challenges', IEEE Open Journal of Intelligent Transportation Systems, vol. 6, pp. 605-633, 2025, https://doi.org/10.1109/OJITS.2025.3564361
5. Z. Li, J. Wang, S. Zhao, Q. Wang, and Y. Wang, 'Evolving Towards Artificial-Intelligence-Driven Sixth-Generation Mobile Networks: An End-to-End Framework, Key Technologies, and Opportunities', Applied Sciences, vol. 15, no. 6, Mar. 2025, https://doi.org/10.3390/app15062920
6. V. Chamola, M. Shall Peelam, M. Guizani, and D. Niyato, 'Future of Connectivity: A Comprehensive Review of Innovations and Challenges in 7G Smart Networks', IEEE Open Journal of the Communications Society, vol. 6, pp. 3555-3613, 2025, https://doi.org/10.1109/OJCOMS.2025.3560035
7. S. Prasad Tera, R. Chinthaginjala, G. Pau, and T. Hoon Kim, 'Toward 6G: An Overview of the Next Generation of Intelligent Network Connectivity', IEEE Access, vol. 13, pp. 925-961, 2025, https://doi.org/10.1109/ACCESS.2024.3523327
8. S. Alraih et al., 'Revolution or Evolution? Technical Requirements and Considerations towards 6G Mobile Communications', Sensors, vol. 22, no. 3, Jan. 2022, https://doi.org/10.3390/s22030762
9. M. Chafii, L. Bariah, S. Muhaidat, and M. Debbah, 'Twelve Scientific Challenges for 6G: Rethinking the Foundations of Communications Theory', IEEE Communications Surveys & Tutorials, vol. 25, no. 2, pp. 868-904, 2023, https://doi.org/10.1109/COMST.2023.3243918
10. F. Zhu et al., 'Wireless Large AI Model: Shaping the AI-Native Future of 6G and Beyond', Dec. 18, 2025, arXiv: arXiv:2504.14653, https://doi.org/10.48550/arXiv.2504.14653
11. S. Shafaei et al., 'Toward AI in 6G: Concepts, Techniques, and Standards', IEEE Access, vol. 13, pp. 143843-143874, 2025, https://doi.org/10.1109/ACCESS.2025.3595752

12. M. R. Fasihi and B. L. Mark, 'Device-to-Device Communication in 5G/6G: Architectural Foundations and Convergence with Enabling Technologies', Jul. 09, 2025, arXiv: arXiv:2507.06946, https://doi.org/10.48550/arXiv.2507.06946.
13. Z. Aasa, F. Elias, and S. C. Ekpo, 'Hybrid Energy and Spectrum Efficient Wireless Network Design for 5G/6G And Wi-Fi 7/8 Applications', Jan. 08, 2026, Research Square. https://doi.org/10.21203/rs.3.rs-8535275/v1
14. H. Wang et al., 'Navigating the Dual-Use Nature and Security Implications of Reconfigurable Intelligent Surfaces in Next-Generation Wireless Systems', IEEE Communications Surveys & Tutorials, vol. 28, pp. 3346-3387, 2026, https://doi.org/10.1109/COMST.2025.3621610
15. W. M. Othman et al., 'Key Enabling Technologies for 6G: The Role of UAVs, Terahertz Communication, and Intelligent Reconfigurable Surfaces in Shaping the Future of Wireless Networks', Journal of Sensor and Actuator Networks, vol. 14, no. 2, Mar. 2025, https://doi.org/10.3390/jsan14020030
16. X. Gan et al., 'Multi-Functional Programmable Metasurfaces for 6G and Beyond', Dec. 07, 2025, arXiv: arXiv:2512.06693, https://doi.org/10.48550/arXiv.2512.06693.
17. A. Tishchenko et al., 'The Emergence of Multi-Functional and Hybrid Reconfigurable Intelligent Surfaces for Integrated Sensing and Communications - A Survey', IEEE Communications Surveys & Tutorials, vol. 27, no. 5, pp. 2895-2936, Oct. 2025, https://doi.org/10.1109/COMST.2024.3519785
18. M. I. Khalil, K. Wang, J. Lin, and J. Choi, 'Mitigating Phase Errors to Improve Signal Quality in RIS-Assisted Satellite Communications', IEEE Transactions on Vehicular Technology, vol. 74, no. 9, pp. 14388-14403, Sep. 2025, https://doi.org/10.1109/TVT.2025.3566480
19. M. Ejaz, G. Jinsong, M. Asim, K. A. Shakil, and M. A. Wani, 'Joint Phase-Shift and Power Allocation Optimization in RIS-Enhanced Wireless Networks: An Intelligent Framework', IEEE Open Journal of the Communications Society, vol. 6, pp. 7389-7404, 2025, https://doi.org/10.1109/OJCOMS.2025.3602856
20. M. Iqbal, T. Ashraf, M. Zubair, S. M. Jameel, M. Jazib, and J.-Y. Pan, 'A comprehensive survey on reconfigurable intelligent surfaces (RIS) and STAR-RIS for next-generation wireless networks', Discov Appl Sci, vol. 7, no. 11, p. 1253, Oct. 2025, https://doi.org/10.1007/s42452-025-07684-w
21. A. Taha, M. Alrabeiah, and A. Alkhateeb, 'Enabling Large Intelligent Surfaces With Compressive Sensing and Deep Learning', IEEE Access, vol. 9, pp. 44304-44321, 2021, https://doi.org/10.1109/ACCESS.2021.3064073
22. M. Jian et al., 'Reconfigurable intelligent surfaces for wireless communications: Overview of hardware designs, channel models, and estimation techniques', Intelligent and Converged Networks, vol. 3, no. 1, pp. 1-32, Mar. 2022, https://doi.org/10.23919/ICN.2022.0005
23. B. Sheen, J. Yang, X. Feng, and M. M. U. Chowdhury, 'A Deep Learning Based Modeling of Reconfigurable Intelligent Surface Assisted Wireless Communications for Phase Shift Configuration', IEEE Open Journal of the Communications Society, vol. 2, pp. 262-272, 2021, https://doi.org/10.1109/OJCOMS.2021.3050119
24. S. Lu et al., 'Deep Reinforcement Learning-Based Intelligent Reflecting Surface for Cooperative Jamming Model Design', IEEE Access, vol. 11, pp. 98764-98775, 2023, https://doi.org/10.1109/ACCESS.2023.3312546
25. Z. U. A. Tariq, E. Baccour, A. Erbad, and M. Hamdi, 'Reinforcement Learning for Resilient Aerial-IRS Assisted Wireless Communications Networks in the Presence of Multiple Jammers', IEEE Open Journal of the Communications Society, vol. 5, pp. 15-37, 2024, https://doi.org/10.1109/OJCOMS.2023.3334489
26. T. Ma, Y. Xiao, X. Lei, L. Zhang, Y. Niu, and G. K. Karagiannidis, 'Reconfigurable Intelligent Surface-Assisted Localization: Technologies, Challenges, and the Road Ahead', IEEE Open Journal of the Communications Society, vol. 4, pp. 1430-1451, 2023, https://doi.org/10.1109/OJCOMS.2023.3292052

27. J. Wang and S. Chen, 'Deep Reinforcement Learning-Based Secrecy Rate Optimization for Simultaneously Transmitting and Reflecting Reconfigurable Intelligent Surface-Assisted Unmanned Aerial Vehicle-Integrated Sensing and Communication Systems', Sensors, vol. 25, no. 5, Mar. 2025, https://doi.org/10.3390/s25051541
28. J. Chen et al., 'Hybrid Reinforcement Learning for Joint Beamforming in STAR-RIS-Assisted CoMP Systems', IEEE Transactions on Wireless Communications, vol. 24, no. 9, pp. 7955-7969, Sep. 2025, https://doi.org/10.1109/TWC.2025.3563597
29. Y. Zhang et al., 'A Unified Deterministic Channel Model for Multi-Type RIS With Reflective, Transmissive, and Polarization Operations', IEEE Transactions on Vehicular Technology, pp. 1-13, 2025, https://doi.org/10.1109/TVT.2025.3605727
30. Y. Huang et al., 'Sum Rate Maximization in STAR-RIS-UAV-Assisted Networks: A CA-DDPG Approach for Joint Optimization', Dec. 01, 2025, arXiv: arXiv:2512.01202. https://doi.org/10.48550/arXiv.2512.01202.
31. C. Cai, X. Yuan, W. Yan, Z. Huang, Y.-C. Liang, and W. Zhang, 'Hierarchical Passive Beamforming for Reconfigurable Intelligent Surface Aided Communications', IEEE Wireless Communications Letters, vol. 10, no. 9, pp. 1909-1913, Sep. 2021, https://doi.org/10.1109/LWC.2021.3085497
32. R. Zhong, Y. Liu, X. Mu, Y. Chen, X. Wang, and L. Hanzo, 'Hybrid Reinforcement Learning for STAR-RISs: A Coupled Phase-Shift Model Based Beamformer', IEEE Journal on Selected Areas in Communications, vol. 40, no. 9, pp. 2556-2569, Sep. 2022, https://doi.org/10.1109/JSAC.2022.3192053
33. X. Yuan, S. Hu, W. Ni, X. Wang, and A. Jamalipour, 'Deep Reinforcement Learning-Driven Reconfigurable Intelligent Surface-Assisted Radio Surveillance with a Fixed-Wing UAV', IEEE Transactions on Information Forensics and Security, vol. 18, pp. 4546-4560, 2023, https://doi.org/10.1109/TIFS.2023.3297021
34. https://www.kaggle.com/datasets/ziya07/6g-iot-intelligent-management-dataset
35. 'A machine learning approach to assess the climate change impacts on single and dual-axis tracking photovoltaic systems | Scientific Reports'. Accessed: Feb. 10, 2026. [Online]. Available: https://www.nature.com/articles/s41598-025-10831-3