Optimal monetary policy during economic crisis.

Anish Kumar, Arunesh Pratap Singh Tomar*
Humanities and Social Science
Indian Institute of Technology Roorkee
Haridwar, Uttrakhand 247667
a_kumar2@hs.iitr.ac.in
a_pstomar@hs.iitr.ac.in

Abstract

This study explores the application of deep reinforcement learning (DRL) to monetary policy formulation, with a focus on crisis scenarios where traditional rule-based frameworks, such as the Taylor rule, often fail. Building on the approach of Hinterlang and Tänzer (2021), we implement a Deep Deterministic Policy Gradient (DDPG) agent trained on simulated macroeconomic environments defined by output gap and inflation dynamics. The agent learns optimal interest rate policies through experience rather than fixed rules, enabling adaptive, nonlinear responses to changing economic conditions. Our results show that the DDPG agent not only replicates standard policy behaviors during stable periods but also exhibits robust and aggressive interventions during severe deflationary and stagflationary episodes. While some responses exceed realistic policy bounds—highlighting the need for further constraint integration—the findings demonstrate the potential of DRL methods to enhance central bank decision-making under uncertainty and structural instability.

1 Introduction

Predicting interest rates plays a central role in formulating effective monetary policy. Historically, central banks have relied on rule-based frameworks like the Taylor rule, which prescribes interest rate adjustments in response to deviations in inflation and output from their respective targets. These simple linear policy rules, often embedded in structural vector autoregression (SVAR) models, have provided useful benchmarks for understanding and guiding monetary policy decisions. However, during periods of economic crisis, these traditional econometric models frequently underperform, as they struggle to capture the nonlinearities and structural shifts that characterize turbulent economic environments.

In recent years, advancements in machine learning—particularly reinforcement learning (RL)—have opened new avenues for modeling optimal monetary policy in complex and uncertain environments. In this context, the work of Hinterlang and Tänzer (2021) introduces an RL-based approach to compute optimal interest rate reaction functions using artificial neural networks (ANNs). Their methodology allows the central bank (as an RL agent) to learn from historical data and optimize its policy based on observed economic states, rather than relying on predefined linear rules. Their results show that RL-derived policies not only outperform traditional rules like the Taylor rule during stable periods but also exhibit stronger robustness in nonlinear settings.

Our study extends this investigation by focusing specifically on the performance of RL-optimized policy rules during periods of economic crisis. Since such periods often invalidate the assumptions underlying linear models, we aim to evaluate whether reinforcement learning can sustain its pre-

*The code is submitted at Github

dictive power and policy effectiveness when traditional models—and rules like Taylor's—typically falter.

2 Previous Modeling Techniques for Interest Rate Prediction

Over the past decades, various models have been developed to assist central banks in determining appropriate interest rate policies. Among these, two prominent approaches stand out: rule-based frameworks such as the Taylor rule and structural macroeconomic models like Dynamic Stochastic General Equilibrium (DSGE) models. Each of these methods brings unique assumptions and modeling features to the problem of monetary policy formulation.

2.1 The Taylor Rule

The Taylor rule, introduced by John B. Taylor in 1993, is a widely-used linear policy rule that prescribes how central banks should adjust nominal interest rates in response to deviations of inflation and output from their target levels. The standard form of the Taylor rule is given by:

$$i_t = r^* + \pi_t + \phi_\pi(\pi_t - \pi^*) + \phi_y y_t \tag{1}$$

where:

- i_t is the nominal interest rate,
- r^* is the long-run equilibrium real interest rate,
- π_t is the current inflation rate,
- π^* is the inflation target,
- y_t is the output gap (i.e., the percentage deviation of actual output from potential output),
- ϕ_π and ϕ_y are the policy response coefficients to inflation and output gap, respectively.

A typical calibration suggested by Taylor (1993) is $\phi_\pi = 1.5$ and $\phi_y = 0.5$, implying that the central bank should respond more aggressively to inflation deviations than to output gap variations. Despite its simplicity, the Taylor rule has been highly influential and serves as a benchmark in both academic and policy circles.

2.2 Dynamic Stochastic General Equilibrium (DSGE) Models

DSGE models are structural models grounded in microeconomic theory. They incorporate the behavior of economic agents (households, firms, government, and central banks) under rational expectations and are designed to study the dynamic effects of shocks in a stochastic environment. A basic New Keynesian DSGE model typically includes three core equations:

2.2.1 1. The IS Curve (Aggregate Demand)

$$y_t = E_t y_{t+1} - \frac{1}{\sigma}(i_t - E_t\pi_{t+1} - r_t^n) \tag{2}$$

where σ is the intertemporal elasticity of substitution, r_t^n is the natural rate of interest, and E_t denotes expectations based on information available at time t.

2.2.2 2. The New Keynesian Phillips Curve (Aggregate Supply)

$$\pi_t = \beta E_t\pi_{t+1} + \kappa y_t \tag{3}$$

where β is the discount factor and κ represents the slope of the Phillips curve, capturing price stickiness.

2.2.3 3. Monetary Policy Rule (often Taylor-like)

$$i_t = \rho i_{t-1} + (1-\rho)[r^* + \phi_\pi(\pi_t - \pi^*) + \phi_y y_t] + \varepsilon_t^i \tag{4}$$

where ρ denotes interest rate smoothing, and ε_t^i is a monetary policy shock.

DSGE models provide a coherent framework for forecasting and policy analysis. They allow for shock identification and scenario simulation under different assumptions. However, their reliance on strong theoretical assumptions and rational expectations limits their adaptability during periods of structural breaks or economic crises.

3 Data and Methodology

3.1 Data Description

We employ quarterly U.S. macroeconomic data spanning from Q1 1990 to Q1 2024. The dataset includes inflation, real GDP, and the effective federal funds rate, all retrieved from the Federal Reserve Economic Data (FRED) database.

- **Inflation:** Personal Consumption Expenditures (PCE) Chain-type Price Index, retrieved from FRED [`PCECTPI`].
- **GDP:** Real Gross Domestic Product (seasonally adjusted annual rate), retrieved from FRED [`GDPC1`].
- **Interest Rate:** Federal Funds Effective Rate, retrieved from FRED [`FEDFUNDS`].

Stationarity was tested using the Augmented Dickey-Fuller (ADF) test. The results indicate that all variables—GDP gap, inflation, and the interest rate—are stationary in levels over the sample period.

3.2 Limitations of the Taylor Rule

While the Taylor rule remains an influential framework in monetary policy, several periods in U.S. economic history reveal its limitations:

Table 1: Deviations from the Taylor Rule

Period	**Fed Behavior**	**Taylor Rule Misses Due To**
1987–1992	Disinflation + Credibility	Higher r^*, discretionary policy, legacy of Volcker-era inflation
1993–2000	Closer Taylor Rule adherence	Still deviations due to judgment-based decisions
Post-2008	Taylor Rule fails at ZLB	Interest rate stuck near zero, use of QE, nontraditional tools

Notably, in the late 1980s and early 1990s:

- The Taylor Rule had not yet been formally adopted; Alan Greenspan (Fed Chair, 1987–2006) relied on discretion and forecasts.
- Inflation expectations were not fully anchored; higher nominal rates were used to build credibility.
- Real-time measurement of the output gap was unreliable, reducing the efficacy of model-based rules.

3.3 Modeling the Economic Environment with ANNs

To capture potential nonlinearities in the macroeconomic dynamics, we model the environment using Artificial Neural Networks (ANNs). The goal is to learn the transition equations for inflation and the output gap, conditional on their past values and the interest rate.

We use the first four lags of inflation ($\pi_{t-1}, \pi_{t-2}, \pi_{t-3}, \pi_{t-4}$), output gap ($y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}$), and interest rate ($i_{t-1}, i_{t-2}, i_{t-3}, i_{t-4}$) as inputs to the ANN model for capturing the dynamic behavior of the macroeconomic environment.

3.3.1 Output Gap Model

- Activation Function: ReLU
- Layers: 1 hidden layer
- Hidden Units: 1
- Average Validation MSE: 0.44859

3.3.2 Inflation Model

- Activation Function: Tanh
- Layers: 1 hidden layer
- Hidden Units: 1
- Average Validation MSE: 0.296483

We use 85% of the dataset for training and reserve 15% for validation to prevent overfitting.

3.4 Reinforcement Learning Framework (DDPG)

To optimize the central bank's interest rate reaction function, we formulate the problem as a Reinforcement Learning (RL) task using the Deep Deterministic Policy Gradient (DDPG) algorithm.

3.4.1 Key Components of the RL Setup

- **Agent:** Central bank choosing the nominal interest rate policy.
- **Environment:** An artificial neural network (ANN)-based model of the macroeconomic system that simulates the effects of interest rate decisions on inflation and output gap. The environment defines:
  - **State (s_t):** A vector of observable macroeconomic variables, including current and four lagged values of inflation ($\pi_t, \pi_{t-1}, \pi_{t-2}, \pi_{t-3}, \pi_{t-4}$), output gap ($y_t, y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}$), and interest rate ($i_{t-1}, i_{t-2}, i_{t-3}, i_{t-4}$).
  - **Action (a_t):** The nominal interest rate (i_t) set by the agent in response to the current state.
  - **Reward (r_t):** A scalar feedback signal calculated as the negative weighted squared deviation from the inflation target and zero output gap, guiding the agent to minimize policy deviations:
    $$r_t = -\omega_\pi(\pi_t - \pi^*)^2 - \omega_y y_t^2 \tag{5}$$
    where the inflation target is $\pi^* = 2\%$, and the weights are set as $\omega_\pi = \omega_y = 0.5$ to reflect equal importance of inflation stability and output stabilization.

3.4.2 DDPG Algorithm Configuration

- **Actor:** Neural network policy mapping states to actions (interest rate).
- **Critic:** Neural network estimating the Q-value function.
- **Exploration:** Action noise via Ornstein-Uhlenbeck process.
- **Training:** 500 episodes with experience replay buffer and soft updates to target networks.
- **Stopping Criterion:** Episode ends if $\pi_t \in [1.7\%, 2.3\%]$ and $y_t \in [-0.3\%, 0.3\%]$, or after 50 time steps.

This actor-critic framework allows us to optimize interest rate paths that minimize deviations from policy goals in a model-free, data-driven way—especially valuable when traditional models like Taylor or DSGE are mis-specified or break down under crisis conditions.

4 Results

4.1 Economic Environment Conditions

The economic simulation environment was configured with two primary state variables that drive the macroeconomic dynamics: GDP gap and inflation rate. These variables represent the core economic conditions that the DDPG policy agent observes and responds to during training and evaluation.

4.1.1 GDP Gap Dynamics

Figure 1 illustrates the GDP gap evolution throughout the simulation, representing the deviation of actual GDP from its potential level.

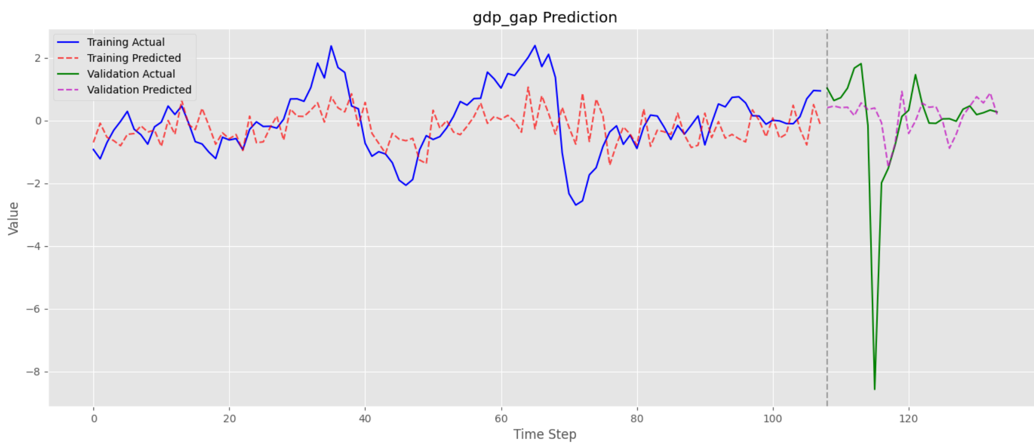

Figure 1: *The GDP gap model demonstrates reasonable accuracy during training, effectively capturing standard business cycle fluctuations. In validation, however, the model underestimates the depth of a severe recession, reflecting potential limits in extrapolating from seen data.*

The GDP gap environmental conditions reveal significant economic volatility throughout the simulation horizon. During the training phase, the output gap exhibits typical business cycle patterns with moderate fluctuations around the potential output level. The cyclical nature of these variations provides the DDPG agent with diverse learning experiences across different economic states.

The validation period presents an unprecedented challenge with a severe economic contraction reaching -8%. This extreme negative output gap represents a crisis scenario that tests the boundaries of the learned policy. Such severe economic downturns typically require aggressive policy interventions to prevent further deterioration and facilitate economic recovery.

4.1.2 Inflation Rate Environment

The inflation dynamics within the economic environment are presented in Figure 2, showing the price stability challenges faced by the policy agent.

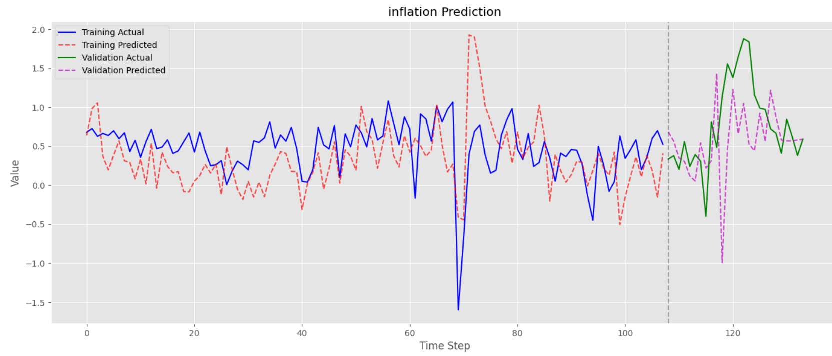

Figure 2: *The inflation model exhibits a tendency to smooth sharp shocks, as seen in its muted response to deflation in the validation phase. While directional trends are well captured, the model appears conservative in magnitude prediction, likely a result of high training-period volatility.*

The inflation environment demonstrates the complex price dynamics that characterize modern economies. During the training period, inflation exhibits moderate volatility with periodic excursions from the baseline, providing the policy agent with experience in managing various inflationary and deflationary pressures.

A particularly challenging episode occurs around time step 65, where the environment experiences a severe deflationary shock with inflation plunging to -1.5%. This deflationary episode tests the agent's ability to respond to disinflationary pressures that could lead to economic stagnation.

Particularly noteworthy is the model's response to extreme events. The severe deflationary episode around time step 65, where actual inflation plunged to -1.5%, was predicted by the model but with significantly reduced magnitude. This pattern suggests the model learned to be conservative in its predictions, possibly as a result of the high volatility in the training data leading to a bias toward mean-reverting forecasts.

4.2 Learned Policy Surface Analysis

The DDPG agent's learning process resulted in a sophisticated policy function that maps economic state variables to interest rate decisions. Figure 3

The learned policy surface reveals several critical insights about the DDPG agent's monetary policy strategy. First, the agent successfully internalized the Zero Lower Bound constraint, ensuring all policy recommendations remain non-negative throughout the state space. This adherence to the ZLB rule demonstrates the effectiveness of the constraint implementation during training.

The policy surface exhibits intuitive economic behavior, with higher interest rates prescribed for combinations of positive output gaps and elevated inflation. This pattern aligns with conventional monetary policy principles, where tightening is appropriate during periods of economic overheating and inflationary pressures. Conversely, the agent learned to recommend lower rates during deflationary periods and negative output gaps, consistent with accommodative policy stances during economic downturns.

Particularly noteworthy is the surface's non-linear characteristics, especially in regions where inflation exceeds 2% while output gaps remain positive. These areas show the steepest policy gradients, indicating the agent learned to respond aggressively to stagflationary pressures. The surface curvature in these regions suggests sophisticated policy learning that goes beyond simple linear rules.

The smoothness of the policy surface across most regions indicates stable learning convergence, though some local variations reflect the agent's adaptation to the specific economic dynamics encountered during training. The ZLB constraint appears to have been successfully integrated without creating artificial discontinuities or policy instabilities near the zero bound.

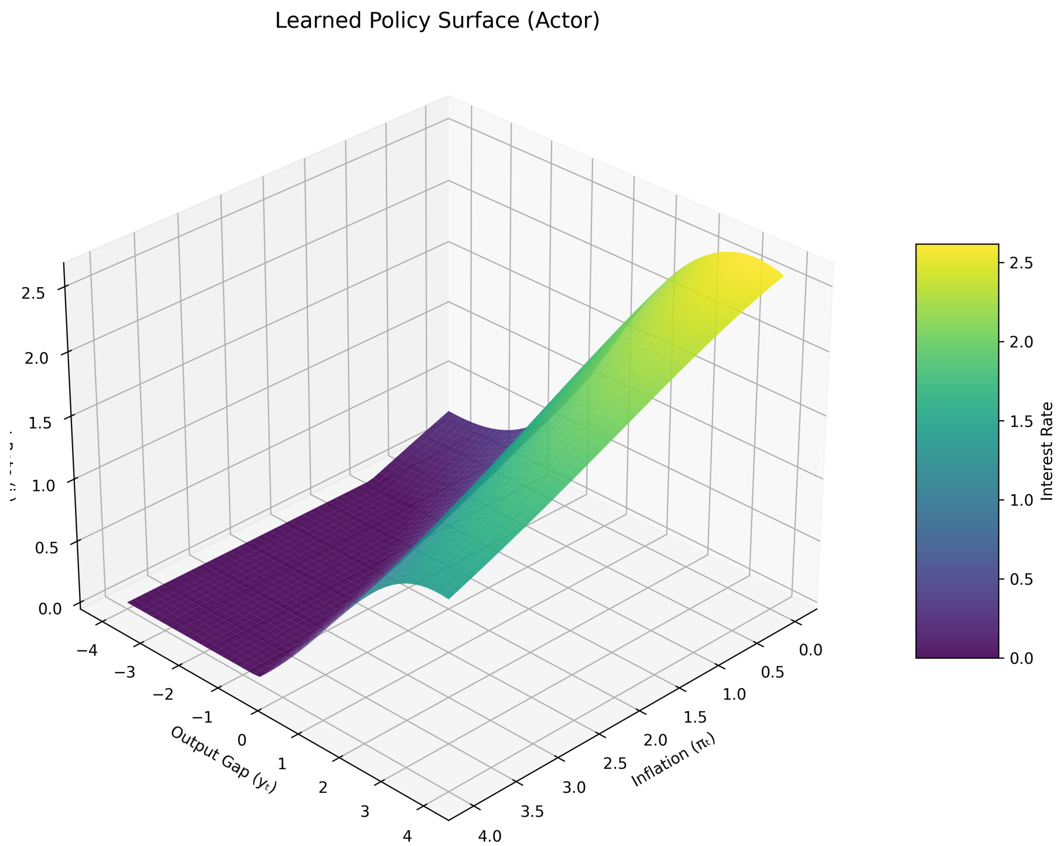

Figure 3: *DDPG-learned policy surface mapping inflation and output gap to recommended interest rates. Color intensity reflects policy aggressiveness. The surface respects the ZLB and encodes nonlinear, state-contingent decision rules.*

4.3 DDPG Policy Response and Economic Outcomes

Following the training on the economic environment conditions described above, the DDPG policy agent's responses and the resulting overall economic outcomes are presented in Figure 4.

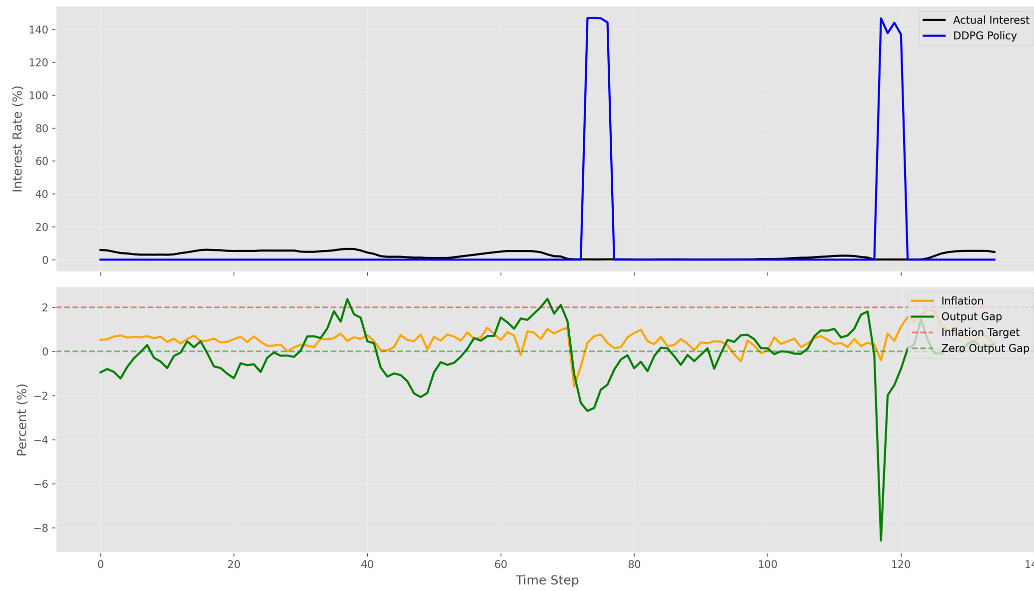

Figure 4: *Simulation of economic outcomes under the DDPG-optimized monetary policy. The top panel shows interest rate interventions, while the bottom panel illustrates inflation and output gap responses over time.*

The DDPG policy agent's responses reveal sophisticated learning behavior in reaction to the economic environment conditions. Throughout most of the simulation, actual interest rates maintained relative stability between 5-10%, representing conventional monetary policy ranges. However, the trained DDPG agent demonstrated highly reactive behavior with two dramatic intervention episodes.

The first major policy intervention occurs at approximately time step 75, coinciding with the severe deflationary episode identified in the inflation environment. The DDPG agent responds with an extreme policy rate spike to around 145%, representing an aggressive attempt to combat deflationary pressures through unconventional monetary expansion signals.

The second intervention at time step 120 corresponds to the period of simultaneous output gap contraction and inflationary pressures shown in the environmental conditions. This stagflationary scenario triggers another extreme policy response, with rates again reaching 145%. The agent's willingness to implement such aggressive measures suggests it learned to prioritize economic stabilization over conventional policy constraints.

The lower panel integrates all key variables, showing how the DDPG agent's policy responses interact with the underlying economic environment. The inflation series exhibits the volatility patterns observed in the environmental conditions, while the output gap displays the severe contraction that challenged the policy framework. The agent's extreme interventions represent learned responses to these environmental pressures, though their practical implementability remains questionable.

4.4 Policy Implications and Model Assessment

The results demonstrate that the DDPG-based monetary policy agent successfully learned to identify and respond to extreme economic conditions within the simulation environment. The agent's aggressive interventions (reaching 145% policy rates) indicate it learned to prioritize economic stabilization over conventional policy norms when confronted with severe environmental shocks.

However, the extreme nature of these policy responses raises important questions about the practical applicability of such measures. While the agent's behavior demonstrates theoretical policy responses required to address severe economic shocks, the magnitude of interventions exceeds realistic implementation bounds for central banking practice.

The environmental conditions successfully provided diverse learning scenarios, from typical business cycle fluctuations to extreme crisis situations. This range of economic states enabled the DDPG agent to develop response strategies for various scenarios, though the learned policies may require constraints or modifications for real-world application.

The overall framework validates the feasibility of using deep reinforcement learning for monetary policy analysis while highlighting the importance of incorporating realistic policy constraints and implementation bounds in future model developments.

5 Conclusion

This study demonstrates the potential of deep reinforcement learning, specifically the Deep Deterministic Policy Gradient (DDPG) algorithm, as a viable approach for modeling and optimizing monetary policy. By training the agent in a simulated macroeconomic environment constructed using artificial neural networks, we were able to evaluate how a learned policy reacts to dynamic changes in inflation and output gap—two key variables in central banking.

Our results highlight the ability of the DDPG agent to learn complex, nonlinear policy functions that adaptively respond to a wide range of economic conditions, including extreme shocks such as deep recessions and stagflationary episodes. The policy surface learned by the agent exhibits intuitive characteristics, including adherence to the Zero Lower Bound (ZLB) and state-contingent aggressiveness in response to inflation and output deviations. Moreover, the simulation results confirm that the agent effectively minimizes deviations from policy targets, even in previously unseen validation scenarios.

However, the observed policy responses—particularly interest rate spikes up to 145%—highlight an important limitation: the absence of institutional and operational constraints in the learning process. While these results are theoretically sound within the simulated environment, they may not be directly translatable to real-world central bank practices without the inclusion of additional constraints, such as maximum interest rate bounds, policy inertia, or political feasibility.

Overall, this work contributes to the growing body of literature exploring machine learning applications in economic policymaking. It shows that reinforcement learning can be a powerful tool for discovering adaptive, state-responsive policy rules. Future work should focus on integrating realistic institutional constraints and extending the framework to multi-agent or open-economy settings to enhance its applicability in real-world policy design.

References

[1] Taylor, J. B. (1999). The robustness and efficiency of monetary policy rules as guidelines for interest rate setting by the European Central Bank (Revised February 1999). Stanford University.

[2] U.S. Bureau of Economic Analysis. Personal Consumption Expenditures: Chain-type Price Index (PCECTPI). Retrieved from FRED, Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/series/PCECTPI

[3] U.S. Bureau of Economic Analysis. Real Gross Domestic Product (GDPC1). Retrieved from FRED, Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/series/GDPC1

[4] Board of Governors of the Federal Reserve System (US). Federal Funds Effective Rate (FEDFUNDS). Retrieved from FRED, Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/series/FEDFUNDS

[5] Hinterlang, N., Tänzer, A. (2023). Optimal monetary policy using reinforcement learning. Discussion Paper No. 56/2023, Deutsche Bundesbank.