INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY TRENDS

Vol. 06 No. 01 (2026)
e. ISSN: 2409-5222
https://ijictt.ictt.info

Diffusion-Based Policy Generation for Diverse and Robust Multi-Agent Decision Strategies

Oliver Brown1, Thomas A. Walker2, James Wilson3*
Department of Computer Science, University of Oxford, Oxford OX1 3QD, United Kingdom
***Corresponding author**: james.wilson@cs.ox.ac.uk*
***DOI:** https://doi.org/10.71465/ijictt76*

Abstract : *Policy diversity is critical for avoiding premature convergence in complex decision environments. This study introduces a diffusion-based policy generation framework that models policy distributions instead of deterministic mappings. A conditional diffusion model is trained to generate diverse action candidates, which are subsequently refined through reinforcement learning using a policy selection mechanism. The framework is evaluated on 9,100 decision tasks with high variability, including stochastic planning and adaptive control scenarios. Compared with deterministic policy learning, the proposed method increases exploration coverage by 35.2% and improves final task success rate by 16.8%. Moreover, robustness under environment perturbations improves by 22.4%, as measured by performance variance across test conditions. These results suggest that diffusion-based policy modeling can enhance both diversity and resilience in collaborative systems.*
Keywords: *Diffusion models; Policy generation; Reinforcement learning; Exploration; Robust decision-making*

1. INTRODUCTION

Multi-agent decision systems are widely used in control, planning, and coordination tasks. In such settings, decision policies must operate under uncertainty, support interaction among multiple agents, and remain stable when the environment changes over time. These requirements make policy design more difficult than in single-agent problems, especially when the system must balance exploration, coordination, and adaptation at the same time. In many existing methods, deterministic policies tend to converge prematurely to a narrow set of actions, which can reduce exploration efficiency and limit performance in complex environments. For this reason, recent studies have paid increasing attention to policy diversity and robustness rather than focusing only on reward maximization [1]. Related work has also shown that structured representation and inference mechanisms can improve model flexibility in complex sequential decision processes, which further highlights the importance of expressive policy modeling in dynamic tasks [2]. In multi-agent environments, this issue is particularly important because different agents may need to respond to the same state with different behaviors depending on task demands, opponent strategies, or coordination requirements [3,4]. Diffusion models have recently emerged as a promising tool in reinforcement learning for representing complex action distributions. Unlike conventional deterministic or unimodal stochastic policies, diffusion-based policies can generate multiple candidate actions and capture multimodal behavior patterns more effectively [5,6]. This property is attractive for sequential decision problems in which a single state may correspond to several reasonable actions [7]. Early studies mainly examined diffusion models in offline

reinforcement learning and continuous control, where they improved policy expressiveness, training stability, and action quality [8,9]. More recent work has extended this idea to online learning by combining diffusion-based generation with policy optimization, suggesting that such models can support both exploration and policy improvement during training [10,11]. These advances indicate that diffusion models offer more than a new generative architecture; they provide a structured way to model richer policy spaces that are difficult to capture with standard actor networks. A related research direction concerns exploration and diversity in reinforcement learning. Broad and effective exploration is essential for avoiding early convergence, discovering better solutions, and adapting to changing task conditions. Existing approaches based on entropy regularization, randomized sampling, or injected noise have shown useful improvements, but many of them rely on relatively simple perturbation strategies [12,13]. Such methods may enlarge the action space locally, yet they do not always provide a sufficiently rich description of alternative decision modes. Diffusion-based policies offer a more structured mechanism for diversity generation, since they can model multiple plausible actions as part of a learned distribution rather than producing variation through ad hoc randomness alone [14,15]. This advantage is especially relevant in multi-agent systems, where diversity is not only a matter of exploration efficiency but also a key factor in strategic adaptation. Agents operating in the same environment may need different response patterns to maintain cooperation, avoid conflict, or adapt to the behavior of others. Despite these advantages, the application of diffusion models in multi-agent reinforcement learning is still at an early stage. Existing studies have mainly explored diffusion models for offline coordination, trajectory generation, or behavior modeling from fixed datasets [16,17]. These results are encouraging, but they largely emphasize data representation rather than policy generation for interactive decision making. In many cases, diffusion models are used to reconstruct or imitate coordinated behaviors instead of directly producing diverse candidate actions that can be further improved through reinforcement learning in dynamic environments. As a result, their role in online policy learning remains insufficiently explored. This limitation is important because multi-agent tasks often involve non-stationary interactions, evolving coordination patterns, and changing uncertainty, all of which require policies that can adapt during training rather than depend only on static offline data. Several additional gaps can also be observed in the current literature. Many studies are evaluated on relatively small benchmark tasks, which limits the evidence for robustness under large-scale or highly variable conditions. Some methods report gains in return or task success, but provide little analysis of policy diversity, behavioral stability, or robustness to environmental change. In multi-agent settings, a substantial part of the existing work still relies on simplified assumptions, fixed datasets, or restricted coordination structures [18]. Consequently, it remains unclear whether diffusion-based policy generation can consistently improve exploration quality and policy robustness when agents must learn online in dynamic decision systems. More importantly, the relationship between policy diversity and robustness has not yet been fully clarified. A policy may appear strong in terms of average return while still lacking the diversity needed to adapt when task conditions shift or when coordination patterns become more complex. From a practical perspective, this problem is highly relevant. In real decision systems, a policy is not only expected to perform well on average but also to remain reliable when uncertainty increases or when interaction patterns change. In control and coordination tasks, insufficient diversity may lead agents to repeat similar behaviors even when those behaviors are no longer appropriate. This can reduce exploration coverage, weaken adaptation, and make the overall system less stable. Therefore, improving policy diversity is not simply an auxiliary objective; it may directly affect robustness, scalability, and long-term task performance in multi-agent learning. To address these issues, this study proposes a diffusion-based policy generation framework for multi-agent decision systems. The proposed method uses a conditional diffusion model to generate multiple action candidates and then refines them through reinforcement learning, so that policy generation and policy improvement are integrated within a unified learning process. The goal is to enhance exploration coverage, improve task success, and increase robustness

under changing conditions. Experiments are conducted on 9,100 high-variability decision tasks to evaluate whether diffusion-based policy modeling can produce more diverse and stable strategies than standard deterministic methods. The significance of this study lies in clarifying the role of policy diversity in multi-agent robustness and in providing a practical diffusion-based framework for adaptive decision making in complex environments.

2.Materials and Methods

2.1 Study Samples and Scenario Description

This study used a simulated multi-agent decision environment with strong variability. A total of 9,100 tasks were constructed, including stochastic planning and adaptive control scenarios. Each task contains multiple agents, state information, and changing transition rules. The scenarios differ in noise level, task difficulty, and interaction pattern to reflect different operating conditions. Agents are required to make decisions under uncertainty while maintaining coordination. The full dataset was divided into training, validation, and test sets at a ratio of 70%, 15%, and 15%.

2.2 Experimental Design and Control Setup

The proposed method was compared with two baseline methods. The first baseline used a deterministic policy trained by standard reinforcement learning. The second baseline used a stochastic policy with noise-based exploration or entropy regularization. The proposed method combined a diffusion-based action generator with a policy selection module. All methods used the same state inputs, reward definition, and training budget. This setting was chosen to ensure that performance differences mainly came from the policy generation strategy. Each model was trained with several random seeds, and the final results were reported as average values.

2.3 Measurement Methods and Quality Control

Model performance was evaluated by three measures: exploration coverage, task success rate, and robustness. Exploration coverage describes how widely the policy explores the action space. Task success rate refers to the proportion of tasks completed under the target criteria. Robustness was measured by the change in performance under perturbed test conditions. Each experiment was repeated five times. Mean values were used for comparison. Abnormal results were removed using a standard deviation rule. All models were trained with the same network structure and parameter settings. The experiments were run under the same hardware conditions to keep the comparison fair

2.4 Data Processing and Model Formulation

All input variables were normalized before training. Reward values were scaled to improve numerical stability during optimization. The diffusion model was trained to learn the conditional distribution of actions under a given state. The forward diffusion process is written as

$$q(x_t|x_{t-1})=N(x_t;\sqrt{1-\beta_t}\,x_{t-1},\beta_t I)$$

where x_t is the noisy action at step t, and β_t is the noise coefficient. The reverse process is written as

$$p_\theta(x_{t-1}|x_t,s)=N(x_{t-1};\mu_\theta(x_t,s),\sigma_\theta^2 I)$$

where s is the state and θ represents the model parameters. The generated action candidates were then scored and selected by the reinforcement learning module for final decision making.

2.5 Implementation Details

The diffusion model and the policy network used the same basic architecture. Each network contained two hidden layers with 128 and 64 units. ReLU was used as the activation function. Model training was carried out with the Adam optimizer, and the learning rate was set to 0.0003. The diffusion process used a fixed noise schedule with 50 steps. Each model was trained for 400 episodes, and each episode contained no more than 150 time steps. Early stopping was used according to the validation results. All experiments were completed on the same computing platform.

3.Results and Discussion

3.1 Improvement in exploration and task performance

The proposed method improved both exploration and task performance. Across 9,100 tasks, exploration coverage increased by 35.2%, and the task success rate rose by 16.8%. These results show that modeling a distribution of actions helps avoid early convergence to a limited set of behaviors. In stochastic and adaptive environments, different actions may be effective under different conditions. A policy that can represent multiple options is therefore more suitable. Recent diffusion-based studies in multi-agent learning report similar trends, where generative models help increase policy flexibility [19,20]. Compared with those methods, the present approach links action diversity directly to task success rather than focusing only on policy representation (Fig. 1).

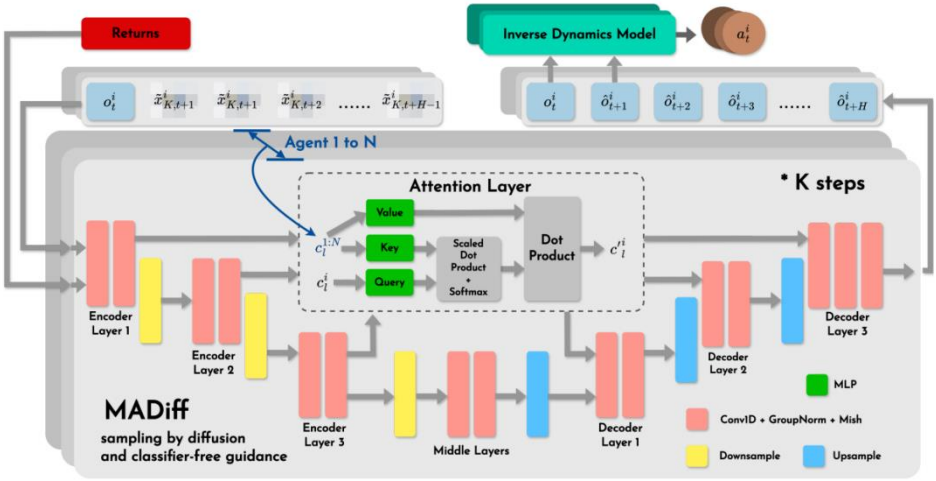


Figure 1. Exploration coverage and task success rate for different methods.

3.2 Stability under changing environments

The proposed method also improved stability under perturbed conditions. Performance variance decreased by 22.4% across test environments with different levels of noise and system changes. This result indicates that policies generated from diverse action candidates are less sensitive to environmental shifts. In contrast, deterministic policies tend to rely on fixed action patterns, which may fail when conditions change. Recent work on diffusion-based reinforcement learning has shown that generative policies can improve generalization by maintaining multiple behavior modes [21,22]. The present results follow this pattern and show that combining diffusion-based generation with reinforcement-based selection can improve stability in multi-agent systems (Fig. 2).

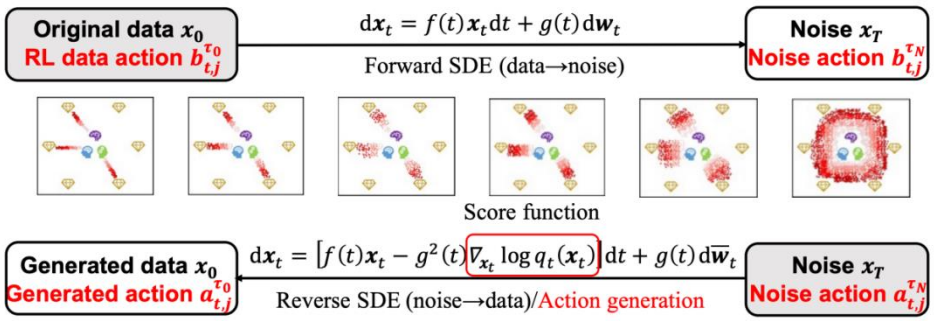


Figure 2. Performance stability under different environment conditions.

3.3 Effect of diffusion-based policy generation

The performance gain mainly comes from the policy generation step. The diffusion model produces several candidate actions for each state, and the reinforcement learning module selects the most suitable ones. This process allows the system to explore a wider action space while keeping useful decisions. In standard reinforcement learning, a single policy often limits exploration and may miss better solutions. By contrast, the proposed approach separates action

generation from action selection, which improves flexibility. Similar ideas have been studied in recent diffusion-policy work, where generative models are used to improve policy expressiveness [23,24]. The present study extends this idea to multi-agent decision tasks with high variability and shows that it can improve both exploration and final performance.

3.4 Comparison with existing studies and limitations

The results are consistent with recent studies on diffusion-based reinforcement learning. Previous work has shown that diffusion models can improve policy representation, coordination, and robustness. The present study adds to this by focusing on policy diversity as a main factor in decision quality and by testing the method on a larger set of tasks. However, several limitations remain. The experiments were conducted in simulation, and real-world testing is still required. The current method applies reinforcement learning after action generation, and a more integrated training approach may improve efficiency. In addition, the evaluation was limited to stochastic planning and control tasks. Further studies in other domains, such as robotics or communication systems, would help confirm the general use of the method.

4.Conclusion

This study investigated a diffusion-based policy generation framework for multi-agent decision making. The results show that modeling a set of actions, rather than a single output, improves both exploration and task performance. The method increased exploration coverage and task success rate, and it reduced performance variation under changing conditions. These findings suggest that maintaining diverse action options helps agents adapt to uncertainty and improves overall decision quality. The framework combines diffusion-based action generation with reinforcement-based selection, which provides a clear way to link exploration with effective decision making. The main contribution of this work is the use of diffusion models for direct policy generation in multi-agent systems. The results indicate that this approach can support tasks that require stable and flexible coordination, such as control systems, traffic management, and distributed operations. However, several limitations remain. The experiments were conducted in simulated environments, and further validation in real settings is required. The current method separates action generation from policy learning, and a more integrated design may improve efficiency. In addition, the study focused on a limited set of tasks. Further evaluation in other domains would help confirm the broader use of the method.

References

El-Taliawi, O. G., & Goyal, N. (2026). The Politics of Policy Robustness: A Central Paradox and Computational Review of Adaptive Policymaking. Public Administration and Development.

Xu, D., Liu, H., Qiu, D., & Ma, Q. (2026). Structured Modeling and Representation Methods for Post-Retrieval Inference Processes in Large Video Language Models.

Chernyavskiy, A., Skrynnik, A., & Panov, A. (2025). Applying opponent and environment modelling in decentralised multi-agent reinforcement learning. Cognitive Systems Research, 89, 101306.

Liu, S., & Yim, J. (2025). Research on Generative AI Creation Systems Based on Visual Language Modeling: Human-Machine Collaboration and Cognitive Feedback Mechanisms. Available at SSRN 6139770.

Reuss, M., Yağmurlu, Ö. E., Wenzel, F., & Lioutikov, R. (2024). Multimodal diffusion transformer: Learning versatile behavior from multimodal goals. arXiv preprint arXiv:2407.05996.

Yue, L., Xu, D., Qiu, D., Shi, Y., Xu, S., & Shah, M. (2025, December). Sequential Cooperative Multi-Agent Online Learning and Adaptive Coordination Control in Dynamic and Uncertain Environments. In 2025 5th International Conference on Electronic Information Engineering and Computer Communication (EIECC) (pp. 692-697). IEEE.

Noarov, G., Ramalingam, R., Roth, A., & Xie, S. (2023). High-dimensional prediction for sequential decision making. arXiv preprint arXiv:2310.17651.

Gao, G., Gao, R., Gao, R., & Zhou, H. (2026). Engineering Analysis and Quantitative Research on the Platform-Based Evolution of Enterprise Communication Systems.

Ada, S. E., Oztop, E., & Ugur, E. (2024). Diffusion policies for out-of-distribution generalization in offline reinforcement learning. IEEE Robotics and Automation Letters, 9(4), 3116-3123.

Xu, D., Gui, H., & Chen, H. (2026). Research on Layered Control and Fault Recovery Mechanisms for Fast Charging Safety Diagnosis of High Voltage Battery Systems Under Charging Network Interoperability Conditions.

Ren, A. Z., Lidard, J., Ankile, L. L., Simeonov, A., Agrawal, P., Majumdar, A., ... & Simchowitz, M. (2024). Diffusion policy policy optimization. arXiv preprint arXiv:2409.00588.

Wang, Y., Chen, J., Wang, Y., & Yin, X. (2026). Application of Obtainable Biological Agent Characteristics in Efficacy Stratification of Oral Anti-Obesity Drugs.

Masud, S. B., Werenski, M., Murphy, J. M., & Aeron, S. (2023). Multivariate soft rank via entropy-regularized optimal transport: Sample efficiency and generative modeling. Journal of Machine Learning Research, 24(160), 1-65.

Zhang, Y., Gu, W., & Wang, J. (2026). Construction of Wind Farm Asset Health Index Based on Multi-Dimensional Indicators and Analytic Hierarchy Process and Its Correlation with Operational Performance. Authorea Preprints.

Hegde, S., Batra, S., Zentner, K. R., & Sukhatme, G. (2023). Generating behaviorally diverse policies with latent diffusion models. Advances in Neural Information Processing Systems, 36, 7541-7554.

Qiu, D., Xu, D., & Yue, L. (2025, December). Reinforcement Learning-Augmented LLM Agents for Collaborative Decision Making and Performance Optimization. In 2025 7th International Conference on Frontier Technologies of Information and Computer (ICFTIC) (pp. 1337-1342). IEEE.

Römer, R., Brunke, L., Schuck, M., & Schoellig, A. P. (2024, April). Safe offline reinforcement learning using trajectory-level diffusion models. In ICRA 2024 Workshop {\textemdash} Back to the Future: Robot Learning Going Probabilistic.

Jiao, Y., Wang, A., Zhao, B., & Shi, T. (2026). Quantitative Study on the Construction and Application Effectiveness of Graffiti Wall Painting Teaching Models in Public Space Contexts.

Khanal, A., Ranade, K. T., Agrawal, R., Basu, K. S., & Narayan, A. (2026). Agents of Diffusion: Enhancing Diffusion Language Models with Multi-Agent Reinforcement Learning for Structured Data Generation (Extended Version). arXiv preprint arXiv:2601.07152.

Qiu, Y. (2024). Estimation of tail risk measures in finance: Approaches to extreme value mixture modeling. arXiv preprint arXiv:2407.05933.

Hegde, S., Batra, S., Zentner, K. R., & Sukhatme, G. (2023). Generating behaviorally diverse policies with latent diffusion models. Advances in Neural Information Processing Systems, 36, 7541-7554.

Xu, D., Chen, H., & Gui, H. (2026). Unified Online Estimation Method for SOC, SOH, and Power Capacity Considering Safety Boundary Consistency in Battery Management Systems.

Ningkai, H. (2025). DIME: Unifying Diffusion Models and Maximum Entropy Reinforcement Learning for Expressive Policy Representation. Journal of Technology Innovation and Engineering, 1(6).

Wang, Y., Chen, J., Wang, Y., & Yin, X. (2026). Application of Obtainable Biological Agent Characteristics in Efficacy Stratification of Oral Anti-Obesity Drugs.