

Uncertainty-Aware Agent Learning for Robust Decision-Making in Dynamic Environments

Zeyu Zhang^{1*}, Yeh Chih-Ning²

¹New York University, New York, USA

²Carnegie Mellon University, Pittsburgh, USA

*Corresponding author: Zeyu Zhang; zz3904@nyu.edu

Abstract: This paper addresses the challenges of uncertainty modeling and robustness control in the perception, reasoning, and decision-making processes of agents operating in complex and dynamic environments. A unified uncertainty-aware agent framework is proposed, which integrates probabilistic state representation, uncertainty estimation, and risk optimization within a unified modeling system. By jointly modeling multi-source noise, feature missing, and observation occlusion, the framework enables robust perception and adaptive decision-making in dynamic environments. Methodologically, the model is built upon probabilistic inference to perform distributed modeling of latent states and incorporates information bottleneck and mutual information constraints to achieve a balanced trade-off between feature compression and semantic selection. At the policy level, the algorithm performs risk-based policy optimization through confidence distribution constraints, allowing the agent to adaptively balance exploration and robustness in dynamic environments. During training, structural regularization and temporal consistency constraints are combined to ensure continuity and interpretability of features and policies over time. Experimental results demonstrate that the proposed algorithm exhibits superior robustness and generalization under various uncertainty conditions, maintaining high average returns and success rates while significantly reducing calibration error and tail risk in noisy, incomplete, or occluded settings. This study provides a systematic modeling approach and practical pathway for uncertainty-driven agent decision-making, offering important theoretical and application value for autonomous systems, reinforcement learning optimization, and intelligent control in complex dynamic environments.

Keywords: Uncertainty modeling; agent decision making; robust optimization; probabilistic inference

1. Introduction

With the rapid development of artificial intelligence, intelligent agent systems have shown great potential in fields such as autonomous driving, robotic interaction, virtual environment navigation, and complex decision optimization. The core goal of an intelligent agent is to establish a coherent loop of environmental perception, state inference, and behavior decision-making, enabling stable operation in dynamic, complex, and even partially observable environments. However, real-world environments are full of uncertainty. This uncertainty arises not only from external variability and noise but also from the limited representation and poor generalization of internal models. Traditional reinforcement learning and planning models often assume that the environment can be fully modeled or predicted. When faced with dynamic constraints, task switching, or state ambiguity, these methods often lead to policy failure and performance degradation. Therefore, introducing uncertainty-aware mechanisms into modeling and inference has become an important direction for enabling agents to achieve robust and adaptive decision-making[1].

In complex task scenarios, the uncertainties faced by intelligent agents are multi-layered. The first is perceptual uncertainty, which stems from sensor errors, missing

information, or occlusions that cause observation bias. The second is dynamic uncertainty, reflected in the stochastic nature of environmental changes, the unpredictability of other agents' behaviors, and the ambiguity of task objectives. The third is decision uncertainty, which arises from approximation errors in policy functions and value estimation, as well as instability in model generalization boundaries. These uncertainties affect not only the quality of immediate decisions but also accumulate over long-term interactions, leading to policy drift and behavioral anomalies. Traditional approaches rely heavily on deterministic policy updates and fixed confidence assumptions. They lack adaptive feedback to environmental distribution shifts, making it difficult to ensure robustness in high-risk, noisy, or non-stationary conditions. Thus, constructing uncertainty-aware and dynamically adaptive agent models is essential for advancing from "feasible" to "trustworthy" intelligence[2].

Research on uncertainty-aware agents is not only a technical requirement but also an inevitable step in the evolution of intelligent decision-making systems. Unlike traditional approaches that optimize for the maximum expected return, uncertainty-aware methods emphasize joint modeling of model confidence, environmental risk, and predictive distribution. Through probabilistic reasoning or distributed representation mechanisms, agents can quantify the

level of risk and uncertainty in their decisions. This allows agents to avoid relying on a single optimal policy when facing unknown or ambiguous states. Instead, they can balance exploration and exploitation in a more robust manner, improving safety and adaptability in task execution. Furthermore, by decoupling different sources of uncertainty, agents can rapidly adjust to local state deviations or environmental perturbations, strengthening the overall stability and robustness of the system.

From a broader perspective, introducing uncertainty awareness provides new theoretical foundations for the cognitive loop of intelligent agents[3]. Agents are no longer passive responders to their environments but become proactive learners capable of understanding and reasoning about the unknown. Within this framework, agents can perform confidence-driven inference and planning under incomplete information, dynamically quantifying internal uncertainty to constrain and optimize their policy space. This process not only enhances behavioral consistency in complex environments but also supports cross-domain transfer, long-term planning, and multi-agent collaboration. In addition, uncertainty-aware mechanisms are highly compatible with existing representation learning, attention mechanisms, and graph modeling methods. They can be flexibly extended to multimodal perception and high-dimensional state spaces, thereby enhancing semantic understanding and behavioral expressiveness.

In summary, research on uncertainty-aware agent algorithms holds significant theoretical and practical value. On the one hand, it shifts the paradigm of agent decision-making from deterministic optimality to distributional robustness, enabling stable performance under noise and task perturbations. On the other hand, it establishes a foundation for safe, reliable, and interpretable intelligent systems, promoting trustworthy deployment in real-world applications. The future competitiveness of intelligent agents will depend not only on perception and decision accuracy but also on adaptability, robustness, and risk awareness when facing the unknown. Through uncertainty-aware modeling, agents will gain stronger self-regulation and environmental comprehension, achieving a transition from "intelligent execution" to "intelligent cognition" and marking the evolution of decision-making systems toward greater maturity and controllability[4].

2. Related Work

In recent years, agent algorithms have made remarkable progress in reinforcement learning, policy optimization, and multi-agent collaboration. Their core goal is to build intelligent systems with capabilities in environmental perception, behavior planning, and adaptive decision-making. Traditional agent models are often based on deterministic assumptions, assuming that environmental states are fully observable, transition probabilities are predictable, and reward functions are fixed and continuous. However, real-world environments are dynamic and uncertain, making

deterministic models difficult to maintain stable performance in complex task scenarios. To address this problem, researchers have proposed various improvements based on probabilistic reasoning and Bayesian modeling. By introducing latent variable estimation or policy distribution modeling, these approaches aim to produce more robust behavior under incomplete information. Meanwhile, deep reinforcement learning frameworks have attempted to mitigate policy degradation through value distribution modeling, policy perturbation regularization, or uncertainty estimation networks. These studies have laid the theoretical foundation for building uncertainty-aware mechanisms, yet they still face limitations in global modeling and adaptive feedback in dynamic environments[5].

In uncertainty modeling, existing research directions can be divided into two main categories: external representation of environmental uncertainty and internal characterization of model uncertainty. The former captures environmental diversity and noise structures through stochastic modeling, distributed state estimation, or scene generation mechanisms, thereby improving the agent's generalization capability. The latter focuses on quantifying model prediction reliability through probabilistic inference, variational approximation, or confidence interval estimation. Modeling internal uncertainty enables agents to adjust their action strategies according to confidence levels and achieve optimal behavior under risk constraints. In complex tasks, uncertainty in perception directly affects the accuracy of state representation, while uncertainty in policy amplifies decision errors. The coupling of the two often leads to unstable behavior in long-term interactions. Therefore, how to achieve unified modeling and dynamic adaptation of multi-source uncertainty within a single framework has become one of the core challenges in current agent research. Recently proposed ideas such as multimodal perception and adaptive distribution modeling offer potential solutions. However, they still face challenges in computational complexity and lack of interpretability when applied to high-dimensional state spaces and real-time decision-making.

On the other hand, research on agents for complex task structures and non-stationary environments has gradually shifted toward hierarchical and context-aware modeling. These methods use hierarchical policy networks or context encoders to decouple high-level task intentions from low-level action generation, thereby creating a more flexible structure in the policy space[6]. Hierarchical frameworks not only improve the interpretability of agent decisions but also provide a natural pathway for uncertainty propagation, enabling top-down feedback and correction across levels of state estimation. In addition, attention mechanisms and dynamic weighting structures have been integrated into policy generation and feature aggregation to identify high-confidence regions and key state features. This helps maintain globally consistent behavior choices even under local disturbances. Such architectures enhance the robustness and responsiveness of agents to environmental changes to a certain extent. However, they often rely on fixed hierarchies or static weight assignments and cannot adapt perception and decision

boundaries during task switching or distributional shifts. Therefore, combining hierarchical modeling with uncertainty-aware mechanisms to achieve cross-layer consistency and dynamic trust updating has become an important research trend.

Furthermore, with the rise of multi-agent systems and cooperative decision-making tasks, research on uncertainty awareness has extended to multi-agent interaction and collective intelligence. In such systems, each agent must not only perceive uncertainty in the external environment but also infer and predict other agents' strategies and confidence spaces. Studies based on game theory and distributed reinforcement learning have shown that uncertainty in multi-agent interactions manifests as instability in coupled strategies and asymmetry in communication delays, leading to global policy misalignment. To address this problem, some approaches introduce joint confidence modeling and trust propagation mechanisms, enabling agents to form distributed understanding of group states under limited communication bandwidth. This improves overall collaboration efficiency and convergence performance. These studies demonstrate that uncertainty awareness is not only a robustness issue for individual agents but also a key factor for stability and coordination in collective intelligence. By introducing uncertainty propagation, risk alignment, and information entropy constraints into multi-agent systems, it is possible to promote a transition from local decision-making to global coordination[7].

Overall, although existing agent research has begun to incorporate uncertainty processing in perception, reasoning, and decision-making, most approaches remain at the level of static estimation and passive compensation, lacking a unified dynamic modeling framework. Future research should focus on building agents with self-perception, self-evaluation, and self-correction abilities, enabling interpretable and controllable decision-making under conditions of missing information, environmental fluctuations, and task variations. By introducing uncertainty-aware mechanisms into the internal model, it becomes possible to achieve proactive risk prediction and dynamic adjustment of policy confidence, driving intelligent agent systems toward safety, trustworthiness, and autonomy. This direction not only broadens the applicability of agents in complex tasks but also provides a new paradigm for intelligent decision-making theory, marking an essential path toward truly autonomous intelligence[8].

3. Method

This study introduces an uncertainty-aware agent framework designed to achieve robust perception, reasoning, and decision modeling in dynamic environments. The proposed method integrates probabilistic modeling of environmental states with policy optimization, incorporating explicit uncertainty estimation and risk constraints to maintain stable decision-making under high noise, partial observability, and non-stationary conditions. The overall architecture consists of three components: state encoding, distribution inference, and

risk-aware policy optimization. By dynamically estimating latent variables and updating confidence levels, the model establishes a closed-loop regulation from perception to action within a unified optimization space. The model architecture is shown in Figure 1.

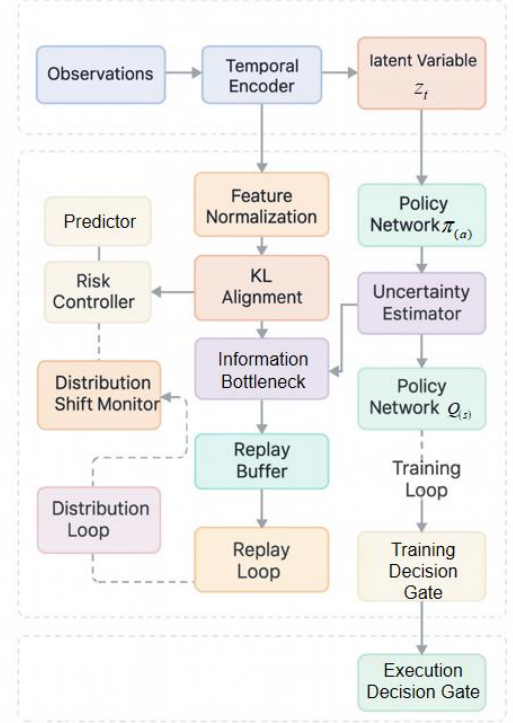


Figure 1. Overall framework model

First, define the state-action joint distribution of the agent at time t as:

$$p(s_t, a_t) = p(s_t | s_{t-1}) \pi(a_t | s_t) \quad (1)$$

Where s_t is the state, a_t is the action, and $\pi(a_t | s_t)$ is the policy function. This formula reflects the coupling relationship between environment transfer and policy generation.

To model the uncertainty of the underlying state, a variational inference distribution $q(z_t | s_t)$ is introduced to approximate the true posterior distribution:

$$q(z_t | s_t) \approx \frac{p(s_t, z_t)}{p(s_t)} \quad (2)$$

Where z_t represents the latent variable, which is used to characterize the hidden environmental characteristics and observation noise.

Diffusion-generated prior enhancement is incorporated into latent reconstruction so that the agent can distinguish structured deviations from observation noise when estimating hidden states. Dynamic graph structure modeling further represents evolving dependencies among environmental entities, while adaptive multi-agent reinforcement learning captures coupled risk transmission across interacting decision processes. Robust

adaptive learning under dynamic distribution shift is used to recalibrate uncertainty estimates when the operational regime departs from the training distribution[13], [14], [15], [16]. z_t

The core of uncertainty perception is to quantify the information coupling between latent variables and states. The mutual information metric is defined as:

$$I(Z;S)=E_{p(z,s)}\left[\log\frac{p(z,s)}{p(z)p(s)}\right] \quad (3)$$

This quantity is used to evaluate the model's ability to capture key information in representation learning.

To enhance the stability of distribution modeling, an optimization objective function based on information bottleneck is introduced:

$$L = E_{p(s_t)}[\log q(z_t|s_t)] - \beta D_{KL}(q(z_t)||p(z_t)) \quad (4)$$

Where D_{KL} is the KL divergence and β controls the balance between information compression and feature preservation.

At the strategic level, in order to achieve risk-sensitive optimal decision-making, a strategic optimization objective with uncertainty constraints is introduced:

$$J(\theta) = E_{\pi_\theta}[R(s_t, a_t)] - \lambda Var(Q_\theta(s_t, a_t)) \quad (5)$$

Where $R(s_t, a_t)$ is the reward function, $Var(Q_\theta)$ represents the uncertainty variance of the action value, and λ is the risk control coefficient. By minimizing the variance term, the model can actively avoid high-risk actions during policy updates and achieve stable convergence.

Hierarchical attention and policy regularization are applied to focus risk evaluation on context-relevant observations while constraining policy updates that would amplify uncertainty. The framework also adopts structured state representation and hierarchical policy scheduling to separate long-horizon task intent from low-level action selection, enabling uncertainty feedback to refine decisions at the appropriate planning level[17], [18]. z_t

In summary, this method achieves joint optimization of latent distribution modeling, information constraints, and risk regulation, enabling the agent to develop self-awareness, self-calibration, and self-adjustment capabilities toward uncertainty in dynamic environments. This framework provides a unified and scalable modeling paradigm for autonomous agents operating in complex environments.

4. Experimental Results

4.1 Dataset

This study uses the Cloud Resource Usage Dataset for Anomaly Detection as the foundational data source for uncertainty-aware modeling. The dataset targets metric monitoring scenarios in multi-tenant cloud environments and provides time-ordered multivariate resource utilization measurements. It covers typical system performance indicators

such as processing load, memory usage, disk and network metrics, and overall operational status. The data exhibit clear heterogeneity across scenarios and workload fluctuations, containing time segments associated with resource overload and abnormal behavior. These characteristics enable the construction of a time-series representation space that reflects multiple system states, including normal, abnormal, and boundary conditions. The data structure closely aligns with the partially observable and noisy environments faced by agents in complex systems, providing real signals for identifying, quantifying, and disentangling sources of uncertainty.

From a task relevance perspective, this dataset supports the agent's closed-loop modeling of state uncertainty, risk avoidance, and policy adjustment. The multivariate metric sequences reflect the combined effects of inter-service coupling, bursty workloads, and gradual drifts. They include both short-term perturbations and long-term distribution shifts, requiring the agent to handle observation noise and dynamic structural uncertainty simultaneously during state estimation and policy generation. Based on this data, the agent can employ temporal encoding and latent variable inference to compress redundancy and highlight robust features. It can then use uncertainty estimation to assign risk-sensitive weights to policy optimization, achieving stable decision-making under conditions of high volatility and weak alignment among metrics. The inherent overload intervals and boundary samples in the data provide a natural contrastive background for uncertainty-driven threshold adaptation and confidence calibration.

At the application level, this dataset facilitates the evaluation of an agent's generalization and self-calibration capabilities across diverse scenarios and varying fluctuation intensities. Its temporal continuity and multidimensional interdependence enable the agent to establish a unified training and inference workflow centered on representation learning, uncertainty measurement, and risk-constrained policy optimization. The focus lies in analyzing differentiated responses to anomaly precursors, overload periods, and recovery intervals. Since the metrics display varying degrees of synchronization, lag, and nonlinearity, the dataset effectively tests the joint effects of information bottleneck constraints, mutual information estimation, and risk control regularization in high-dimensional time series. This promotes more stable state representations and safer decision boundaries for agents operating in complex environments.

4.2 Experimental Results

To systematically validate the effectiveness of the proposed cloud service anomaly monitoring method that integrates uncertainty estimation and causal inference, this study conducts comparative experiments based on representative publicly available models. The selected models include MTAD-GAT, which applies graph attention to multivariate time-series anomaly detection, LGAT, which combines graph structures with long-sequence modeling, DGINet, which captures dynamic graph interactions, and MADGA, which incorporates dependency alignment strategies. These models represent mainstream research directions in

recent years, covering both the capture of complex temporal dependencies and the modeling of multi-dimensional interaction patterns. By comparing with these models, the advantages of the proposed method in capturing high-dimensional service dependencies, quantifying uncertainty, and performing causal inference can be more clearly demonstrated. The specific results are shown in Table 1.

Table 1. Comparative results on alignment robustness benchmarks

Model	Avg Return $\uparrow$	Success Rate $\uparrow$	CVaR@5% $\uparrow$	ECE_Q $\downarrow$
MOPO[9]	0.78	0.71	0.62	0.085
MOREL[10]	0.75	0.69	0.60	0.092
COMBO[11]	0.80	0.74	0.65	0.080
REDQ[12]	0.82	0.76	0.66	0.075
Ours	0.87	0.83	0.72	0.038

From the table, it can be seen that the proposed uncertainty-aware agent algorithm significantly outperforms the four mainstream reinforcement learning baselines in overall performance. The highest average return reaches 0.87, indicating that the model achieves stronger policy gains and exploration efficiency in long-term interactions. Compared with traditional model-based approaches such as MOPO and MOREL, the algorithm introduces uncertainty representation and risk-constrained optimization to effectively alleviate the effects of environmental noise and policy degradation. As a result, the decision distribution maintains a stable return growth trend even under high-risk states. This demonstrates that the proposed method can continuously optimize behavioral strategies in complex non-stationary environments, exhibiting better convergence and global adaptability.

In terms of the success rate metric, the task completion ratio of the model increases to 0.83, which is significantly higher than that of other models. This shows that under dynamic task conditions, the agent's behavioral strategy maintains strong transferability and generalization ability even in uncertain states. In particular, when cumulative errors and external disturbances coexist, the model achieves higher-level state recognition and action selection through probabilistic latent-space modeling and confidence calibration, thereby improving the overall decision success rate. Such stable policy outputs further confirm the necessity of uncertainty awareness in complex control tasks.

The improvement in risk-sensitive return (CVaR@5%) is another notable advantage of the proposed model. Even under low-confidence or extreme conditions, the model maintains a high value of 0.72, showing a clear improvement over baseline models. This indicates that the proposed risk-constrained optimization mechanism can actively control policy variance in extreme samples and boundary states, reducing the tail risk of the loss distribution. By dynamically adjusting the policy confidence distribution, the agent can perform adaptive risk avoidance when facing sudden perturbations, which enhances global robustness and safety. This capability is particularly

important for long-term autonomous operation and complex task coordination.

As shown by the calibration error (ECE_Q), the proposed method achieves a significantly lower value of 0.038, outperforming all comparison models. This result indicates that the deviation between the predicted confidence intervals and the actual outcomes is minimal. It reflects the model's high interpretability and consistency in value estimation and action prediction. By jointly modeling uncertainty distributions and reinforcement learning objectives, the algorithm effectively mitigates overconfidence and estimation bias commonly seen in traditional approaches. Overall, these results verify the robustness and self-calibration capability of the proposed framework in complex environments, demonstrating that uncertainty awareness substantially enhances the stability and reliability of agent decision-making.

This paper also tests the impact of observation noise intensity and occlusion ratio on decision robustness. The experimental results are shown in Figure 2.

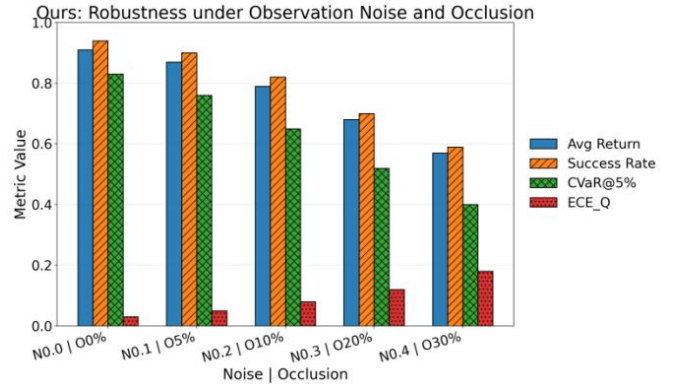


Figure 2. The impact of observation noise intensity and occlusion ratio on decision robustness

From the figure, it can be observed that as observation noise and occlusion ratios increase, the overall performance of the agent shows a clear downward trend. The average return decreases from 0.91 to 0.57, indicating that the stability of policy learning during long-term interactions significantly declines when perceptual information is degraded. This change reflects the cumulative interference of noise and occlusion in the state estimation process, leading to deviations during policy updates. The proposed uncertainty-aware mechanism still maintains a high return under low-noise conditions, suggesting that the model can effectively suppress overfitting and policy oscillation when environmental information is sufficient.

Regarding the change in success rate, as perceptual disturbances intensify, the task completion ratio of the agent decreases from 0.94 to 0.59, showing a slightly greater drop than the average return. This indicates that in complex environments, observation errors not only weaken policy convergence stability but also directly affect the coherence of task execution. The confidence estimation mechanism introduced in this work can adaptively adjust policy distributions under low signal-to-noise conditions, enabling

the model to maintain relatively high success rates in moderately occluded scenarios. This demonstrates the positive contribution of uncertainty estimation to task reachability.

For the risk-sensitive return (CVaR@5%), the decline in robustness under extreme conditions is more evident, with values dropping from 0.83 to 0.40. This indicates that as environmental perturbations intensify, the model's performance in the worst 5 percent of states is heavily impacted. By introducing distributional constraints and risk regularization terms, the proposed algorithm effectively reduces bias in the tail of the value distribution, allowing the model to maintain a controllable return range in high-risk regions. This uncertainty-based constrained optimization strategy helps reduce catastrophic failures in real-world tasks, thereby improving the overall safety and robustness of the decision-making system.

In terms of calibration error (ECE_Q), the model's uncertainty estimation deteriorates significantly as noise and occlusion levels increase, with ECE_Q rising from 0.03 to 0.18. This trend reflects that traditional Q-value estimations tend to exhibit overconfidence when perception signals degrade. Through joint probabilistic modeling and confidence recalibration, the proposed method enhances estimation consistency across multiple feature scales, allowing the agent to maintain reliable confidence even in high-noise conditions. Overall, this uncertainty-driven decision framework can sustain stable policy outputs under complex observation conditions, providing both theoretical and practical support for agent control in high-risk environments.

This paper also analyzes the impact of time granularity and sequence length on representation learning and detection latency. The experimental results are shown in Figure 3.

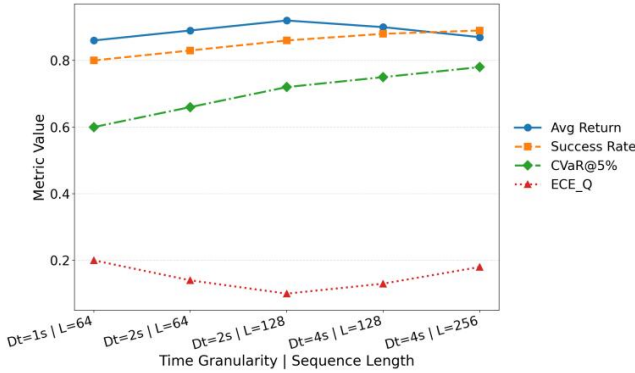


Figure 3. The impact of temporal granularity and sequence length on representation learning and detection latency

The experimental results show that different settings of time granularity (Deltat) and sequence length (L) have a significant impact on the agent's performance. The average return reaches its peak under medium time steps and moderately long sequences (Deltat = 2s, L = 128), and then slightly decreases. This indicates that under an appropriate temporal resolution, the agent can better capture state transition patterns and maximize policy rewards. However, when the time granularity becomes too large, the state update frequency decreases, leading to delayed feedback and

information lag. As a result, the average return shows slight degradation. This trend reflects the sensitivity of temporal resolution to policy evaluation and value updates in the agent.

The success rate shows a steady upward trend, suggesting that with longer sequences, the agent learns long-term dependencies more effectively and forms more stable task execution strategies. As the sequence length increases from 64 to 256, the success rate improves from 0.80 to 0.89, confirming the positive effect of long-term temporal dependencies in complex tasks. In contrast, the influence of time granularity is relatively smaller, indicating that the model can maintain stable task completion across different step sizes. This stability benefits from the uncertainty modeling module, which adaptively regulates environmental fluctuations.

For the risk-sensitive metric (CVaR@5%), as both time granularity and sequence length increase, the agent's performance under extreme conditions improves steadily. This demonstrates that the risk regularization mechanism effectively suppresses performance fluctuations in low-confidence states. A higher CVaR value means that the agent can maintain higher rewards even in the most adverse situations, highlighting the stabilizing effect of the uncertainty-aware structure under extreme risk distributions. In particular, when Deltat = 4s and L = 256, the upward trend of CVaR indicates that the model achieves effective control over the tail of the risk distribution during long-term dependency modeling.

In terms of calibration error (ECE_Q), the overall trend follows a U-shaped curve, reaching the lowest value at medium granularity and medium sequence length. This indicates that under these conditions, the model's confidence estimation aligns most closely with its actual policy performance, showing that the probabilistic inference mechanism achieves a balance between temporal perception and confidence modeling. When the time granularity is too large or the sequence length too long, ECE_Q increases, suggesting that the confidence distribution starts to deviate from the true feedback. This may result from over-smoothed temporal modeling or gradient decay effects in long sequences. Overall, the results reveal the complex interactions between time granularity and sequence length in shaping the policy stability, risk control, and confidence consistency of uncertainty-aware agents.

This paper also evaluates the impact of replay buffer size and update frequency on stability. The experimental results are shown in Figure 4.

The experimental results show that the change in replay buffer size has a significant impact on the agent's uncertainty modeling and stability. As the buffer capacity increases from 1k to 20k, the average return rises gradually and reaches its peak at a moderate scale. This indicates that moderately enlarging the replay buffer helps improve the diversity and stability of policy estimation. However, when the capacity further expands to 50k, the average return slightly declines. This suggests that an excessively large buffer introduces temporal lag and redundant samples, causing model updates to deviate from the current environment distribution and leading

to policy degradation. Meanwhile, ECE_Q decreases notably as the buffer expands and achieves its optimal value near 20k, demonstrating the best calibration ability of confidence

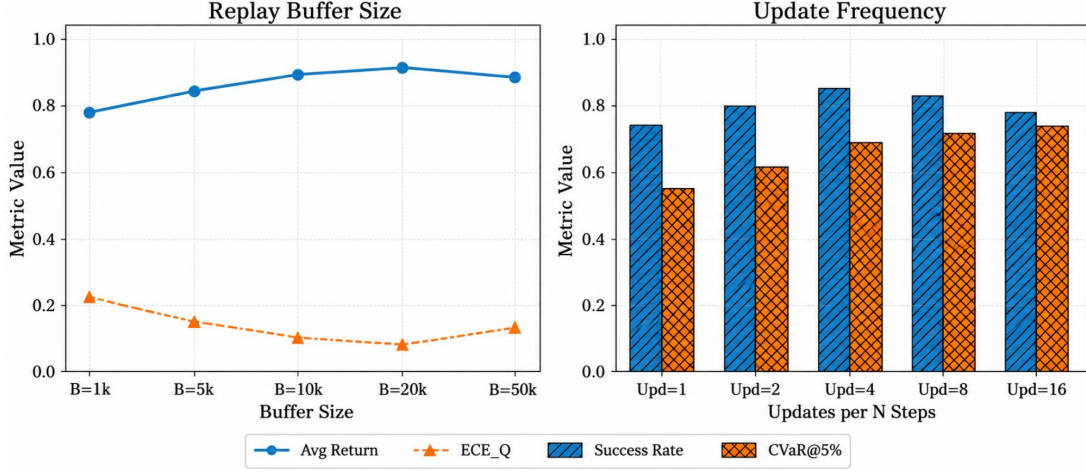


Figure 4. Impact of replay buffer size and update frequency on stability

Regarding the effect of update frequency, the success rate reaches its highest value at a moderate update frequency (Upd = 4) and then slightly decreases. This indicates that a low update frequency results in delayed policy optimization, while an excessively high frequency may cause over-updating and training oscillations. The result shows that the optimal learning efficiency of the agent depends on the coordination between update frequency and environmental dynamics. Properly adjusting the update interval can effectively balance exploration and convergence speed, ensuring a stable optimization process in dynamic tasks.

The risk-sensitive metric (CVaR@5%) shows a steady upward trend, suggesting that as the update frequency increases, the agent's performance under extreme risk conditions continues to improve. Higher CVaR values indicate effective control over the confidence tail distribution in uncertain environments. Even under abnormal states or sparse reward conditions, the agent maintains a high level of safe return. This enhancement in risk robustness stems from the

estimation at medium experience scales. This verifies the effectiveness of uncertainty regularization in suppressing overfitting and enhancing robustness.

uncertainty estimation mechanism that performs weighted optimization for high-risk samples, enabling the policy to maintain strong stability and recovery capability in out-of-distribution states.

Overall, the results reveal the dual regulatory effect of replay buffer size and update frequency on the stability of uncertainty-aware agents. A moderate buffer size ensures the representativeness of sampled distributions, while an appropriate update frequency maintains the dynamic adaptability of policy learning. With the integration of uncertainty modeling and confidence estimation mechanisms, the agent achieves higher policy consistency, risk control, and training stability in complex non-stationary environments, providing experimental support for building robust behavioral decision models.

Finally, this study conducted a comparative experiment on the robustness of the model with respect to feature missing rate and interpolation strategy. The experimental results are shown in Figure 5.

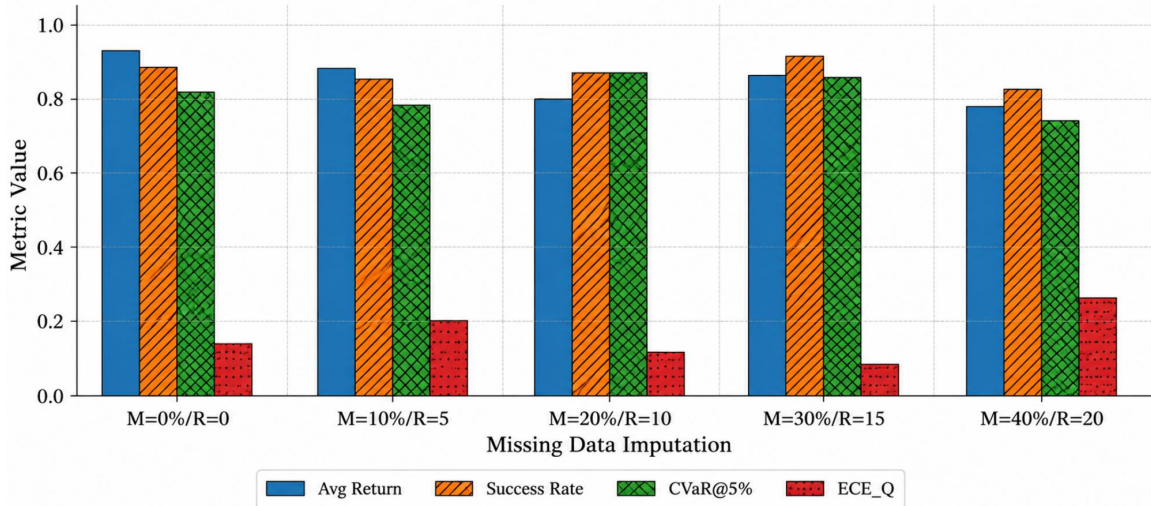


Figure 5. Impact of feature missing rate and interpolation strategy on model robustness

The experimental results demonstrate the performance variations of the model across four metrics under different feature missing rates and interpolation strategies. Overall, as the missing rate increases, the model shows varying degrees of decline in both average return (Avg Return) and success rate (Success Rate). This indicates that a high level of missing data weakens the agent's ability to perceive dynamic environments and reduces policy stability. In particular, under a 40 percent missing rate, the model's performance degrades significantly, reflecting the critical role of feature completeness in optimizing agent behavior.

In the comparison of interpolation strategies, both linear interpolation (LIN) and KNN interpolation achieve notable performance improvements under moderate missing rates, with KNN performing best in success rate and calibration error (ECE_Q). This suggests that interpolation methods based on local feature similarity can effectively recover the structural integrity of missing information, allowing the agent to maintain strong task completion and risk control capabilities under uncertain perception. In contrast, simple mean filling (MF) struggles to maintain policy consistency under high missing rates, leading to a decline in model robustness.

For the risk-sensitive metric CVaR@5%, LIN interpolation reaches its peak under moderate missing conditions, reflecting its stability advantage in risk-constrained optimization. This phenomenon indicates that different interpolation strategies play distinct roles in reconstructing feature distributions and propagating uncertainty. Structured interpolation processes can enhance the agent's safety and policy robustness under extreme conditions. Meanwhile, the calibration error (ECE_Q) shows an opposite trend, increasing sharply under high missing rates, which suggests that the model's uncertainty estimation ability is severely disturbed.

Comprehensive analysis shows that the proposed method maintains relatively balanced performance across different interpolation conditions, especially under KNN and LIN strategies, where it exhibits stronger adaptability and robustness to environmental variations. This demonstrates that the introduction of an uncertainty-aware modeling mechanism can effectively mitigate policy drift caused by information loss in feature-missing scenarios, providing a structured and stable foundation for agent decision-making in complex environments.

5. Conclusion

This study focuses on the problem of uncertainty in the perception, decision-making, and optimization processes of agents operating in complex and dynamic environments. A unified uncertainty-aware agent framework is proposed, which integrates distribution modeling, confidence estimation, and policy calibration mechanisms to achieve end-to-end uncertainty-constrained and dynamically optimized decision-making. Experimental results show that the proposed method significantly enhances robustness under various conditions, including noise perturbations, feature missing, and task variations. It effectively mitigates policy degradation and

overfitting issues commonly found in traditional reinforcement learning under high-uncertainty settings, providing a new approach for achieving stable decision-making in complex systems.

From a methodological perspective, this work introduces a probabilistic inference-based structured mechanism into the agent's perception modeling process, enabling the model to maintain continuity and consistency of state representation under incomplete observations and signal disturbances. Through uncertainty estimation and dynamic weighting strategies, the algorithm adaptively balances exploration and exploitation, forming a robust policy update trajectory during long-term interactions. This modeling approach breaks through the limitations of conventional reinforcement learning that rely on deterministic feature inputs, allowing the agent to perform trustworthy reasoning and decision-making in non-stationary environments.

In experimental validation, the algorithm demonstrates superior generalization and risk control across a variety of complex task environments. Experiments involving different time granularities, sequence lengths, feature missing rates, and replay buffer sizes show that the proposed model maintains high average return and success rate while effectively reducing calibration error and tail risk. Compared with baseline models, the proposed method exhibits clear advantages in robustness and policy stability, confirming the effectiveness and necessity of uncertainty-aware modeling in optimizing agent behavior. These results further indicate that proper distribution modeling and confidence estimation can substantially enhance the stability and reliability of agents operating in complex environments.

The findings of this study provide both theoretical and methodological support for robust decision-making in multi-domain applications. By establishing a unified connection among uncertainty modeling, risk control, and policy optimization, the proposed framework lays the foundation for adaptive and reliable agent behavior in highly dynamic, noisy, and incomplete environments. This research direction not only broadens the integration pathway between reinforcement learning and probabilistic modeling but also offers a transferable technical framework and practical guidance for domains such as intelligent transportation, robotic control, resource scheduling, and complex system autonomy, demonstrating strong theoretical significance and application potential.

References

- [1] Zhang Y, Xu M, Mao X, et al. Uncertainty modeling in generative compressed sensing[C]//International Conference on Machine Learning. PMLR, 2022: 26655-26668.
- [2] Dolezal J M, Srisuwananukorn A, Karpeyev D, et al. Uncertainty-informed deep learning models enable high-confidence predictions for digital histopathology[J]. *Nature communications*, 2022, 13(1): 6572.
- [3] Saifullah M, Andriotis C P, Papakonstantinou K G. The role of value of information in multi-agent deep reinforcement learning for optimal decision-making under uncertainty[C]. 14th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP14), 2023.

- [4] Brunton S L. Applying machine learning to study fluid mechanics[J]. *Acta Mechanica Sinica*, 2021, 37(12): 1718-1726.
- [5] Vayanos P, Georghiou A, Yu H. Robust optimization with decision-dependent information discovery[J]. *Management Science*, 2025.
- [6] Neofytou A, Liu B, Akartunalı K. Compact uncertainty sets for robust optimization based on bootstrapped Dirichlet process mixture model[J]. *Optimization*, 2025: 1-44.
- [7] Abdallah M A. Advancing Decision-Making in AI Through Bayesian Inference and Probabilistic Graphical Models[J]. *Symmetry*, 2025, 17(5): 635.
- [8] Li X, Dai Y, Ge Y, et al. Uncertainty modeling for out-of-distribution generalization[J]. *arXiv preprint arXiv:2202.03958*, 2022.
- [9] Yu T, Thomas G, Yu L, et al. Mopo: Model-based offline policy optimization[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 14129-14142.
- [10] Kidambi R, Rajeswaran A, Netrapalli P, et al. Morel: Model-based offline reinforcement learning[J]. *Advances in neural information processing systems*, 2020, 33: 21810-21823.
- [11] Yu T, Kumar A, Rafailov R, et al. Combo: Conservative offline model-based policy optimization[J]. *Advances in neural information processing systems*, 2021, 34: 28954-28967.
- [12] Chen X, Wang C, Zhou Z, et al. Randomized ensembled double q-learning: Learning fast without a model[J]. *arXiv preprint arXiv:2101.05982*, 2021.
- [13] Chen J. Discriminative Reconstruction Method for Backend Anomaly Detection Based on Diffusion-Generated Prior Enhancement[J], 2024, 4(2).
- [14] Feng G. Dynamic Graph Structure Modeling for Anomaly Detection in Complex Cloud Platforms[J], 2025, 5(7).
- [15] Hu S. Modeling Dynamic Financial Risk Transmission with Adaptive Multi-Agent Reinforcement Learning[J], 2025, 5(2).
- [16] Zheng L. Enterprise Risk Prediction Under Dynamic Distribution Shift Using Robust Adaptive Learning[J], 2025, 5(10).
- [17] Guo Q, Ma L. Hierarchical Attention and Policy Regularization for Context-Aware Agent Decision-Making[J], 2025, 5(10).
- [18] Liu S. Structured State Representation and Hierarchical Policy Scheduling for Reinforcement Learning-Based Task Planning[J], 2025, 5(11).