Novelty Assessment Report

Paper: DreamPhase: Offline Imagination and Uncertainty-Guided Planning for Large-Language-Model Agents
PDF URL: https://openreview.net/pdf?id=81PJ2KPnmK
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-29

Abstract

Autonomous agents capable of perceiving complex environments, understanding instructions, and performing multi-step tasks hold transformative potential across domains such as robotics, scientific discovery, and web automation. While large language models (LLMs) provide a powerful foundation, they struggle with closed-loop decision-making due to static pretraining and limited temporal grounding. Prior approaches either rely on expensive, real-time environment interactions or brittle imitation policies, both with safety and efficiency trade-offs. We introduce DreamPhase, a modular framework that plans through offline imagination. A learned latent world model simulates multi-step futures in latent space; imagined branches are scored with an uncertainty-aware value and filtered by a safety gate. The best branch is distilled into a short natural-language reflection that conditions the next policy query, improving behavior without modifying the LLM. Crucially, DreamPhase attains its performance with substantially fewer real interactions: on WebShop, average API calls per episode drop from $\sim$40 with ARMAP-M (token-level search) to <10 with DreamPhase, a $4\times$ reduction that lowers latency and reduces executed irreversible actions by $\sim 5\times$ on WebShop (4.9$\times$ on ALFWorld) per incident logs. Across web, science, and embodied tasks, DreamPhase improves sample efficiency, safety, and cost over search-based and reward-based baselines. This offers a scalable path toward safe, high-performance autonomous agents via imagination-driven planning. Code: \url{https://anonymous.4open.science/r/DreamPhase-A8AD/README.md}.

Core Task Landscape

This paper addresses: **Offline Imagination and Uncertainty-Guided Planning for Language Model Agents**

A total of **46 papers** were analyzed and organized into a taxonomy with **23 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **World Model Learning and Latent Dynamics**
- **Uncertainty Quantification in Model-Based RL**
- **LLM-Based Planning and Reasoning**
- **Policy Learning and Value Estimation**
- **Multi-Agent and Safety-Critical Domains**
- **Specialized Planning and Auxiliary Methods**
- **DreamPhase and Core Framework**

Complete Taxonomy Tree

- Offline Imagination and Uncertainty-Guided Planning for Language Model Agents Survey Taxonomy
- World Model Learning and Latent Dynamics
  - Forward World Models for Embodied Navigation (4 papers)
  - [1] Dreamwalker: Mental planning for continuous vision-language navigation (Hanqing Wang, 2023) View paper
  - [10] Dreamernav: Learning-based autonomous navigation in dynamic indoor environments using world models (Stuart Shanks, 2025) View paper
  - [22] VISTAv2: World Imagination for Indoor Vision-and-Language Navigation (Yanjia Huang, 2025) View paper
  - [29] NavForesee: A Unified Vision-Language World Model for Hierarchical Planning and Dual-Horizon Navigation Prediction (Fei Liu, 2025) View paper
  - Long-Horizon and Dual-Process World Models (2 papers)
  - [12] Open-world reinforcement learning over long short-term imagination (Li, 2024) View paper
  - [25] DMWM: Dual-Mind World Model with Long-Term Imagination (Wang Ling-yi, 2025) View paper
  - Reverse and Bidirectional Imagination (1 papers)
  - [42] Offline Reinforcement Learning with Reverse Model-based Imagination (Jianhao Wang, 2021) View paper
  - Latent-Space and Low-Dimensional Dynamics (2 papers)
  - [33] Uncertainty Estimation Using Riemannian Model Dynamics for Offline Reinforcement Learning (Guy Tennenholtz, 2022) View paper
  - [38] L-MBOP-E: Latent-Model Based Offline Planning with Extrinsic Policy Guided Exploration (Imran Adham, 2024) View paper
- Uncertainty Quantification in Model-Based RL
  - Ensemble and Bayesian Uncertainty Estimation (4 papers)
  - [4] Uncertainty-aware model-based offline reinforcement learning for automated driving (Christopher Diehl, 2023) View paper
  - [30] Long-Horizon Model-Based Offline Reinforcement Learning Without Conservatism (Tianwei Ni, 2025) View paper
  - [31] Reflect-then-Plan: Offline Model-Based Planning through a Doubly Bayesian Lens (Jeong, 2025) View paper
  - [41] UMBRELLA: Uncertainty-Aware Model-Based Offline Reinforcement Learning Leveraging Planning (Diehl, 2021) View paper
  - Search-Based and Alternative Uncertainty Methods (1 papers)

- [15] SUMO: Search-Based Uncertainty Estimation for Model-Based Offline Reinforcement Learning (Lyu, 2025) View paper
- Uncertainty-Driven Exploration and Imagination (2 papers)
- [5] Uncertainty-driven imagination for continuous deep reinforcement learning (Gabriel Kalweit, 2017) View paper
- [18] Bayesian Uncertainty Estimation for Targeted Counterfactual Experience Generation in Reinforcement Learning (L Tong, 2025) View paper
- Penalty-Based Conservatism and Regularization (2 papers)
- [19] Maximization Operator Uncertainty Penalty for Model-Based Offline Reinforcement Learning (Wang Luo, 2025) View paper
- [21] Imagination-Limited Q-Learning for Offline Reinforcement Learning (Liu Wen-Hui, 2025) View paper

- LLM-Based Planning and Reasoning
  - LLM-Guided Navigation and Embodied Tasks (3 papers)
  - [6] Interactive planning using large language models for partially observable robotic tasks (Lingfeng Sun, 2024) View paper
  - [7] NavCoT: Boosting LLM-Based Vision-and-Language Navigation via Learning Disentangled Reasoning (Bingqian Lin, 2024) View paper
  - [23] Stairway to Autonomy: Hierarchical Decision-Making for LLM-Guided Planning, Bandit-Driven Exploration, and Multi-Agent Navigation (Nayak, 2025) View paper
  - LLM-Driven Dialogue and Goal-Oriented Interaction (2 papers)
  - [16] Zero-shot goal-directed dialogue via rl on imagined conversations (Hong, 2023) View paper
  - [46] Zero-Shot Goal Dialogue via Reinforcement Learning on Imagined Conversations (J Hong, n.d.) View paper
  - LLM Rollouts for Offline RL (2 papers)
  - [2] Knowledgeable agents by offline reinforcement learning from large language model rollouts (Pang, 2024) View paper
  - [3] Kalm: Knowledgeable agents by offline reinforcement learning from large language model rollouts (Xiong-Hui Chen, 2024) View paper
  - LLM Planning with Search and Uncertainty (1 papers)
  - [24] Monte Carlo Planning with Large Language Model for Text-Based Game Agents (Shi, 2025) View paper
  - Stress Testing and Failure Detection (1 papers)
  - [28] Adaptive Stress Testing Black-Box LLM Planners (Chakraborty, 2025) View paper
- Policy Learning and Value Estimation
  - Conservative Offline RL without World Models (1 papers)
  - [8] MOReL : Model-Based Offline Reinforcement Learning (Kidambi, 2020) View paper
  - Offline-to-Online Transition and Finetuning (2 papers)
  - [11] Planning without Search: Refining Frontier LLMs with Offline Goal-Conditioned RL (Hong, 2025) View paper
  - [26] A Simple Unified Uncertainty-Guided Framework for Offline-to-Online Reinforcement Learning (Guo Si-yuan, 2023) View paper
  - Policy Comparison and Evaluation (2 papers)
  - [37] Offline Policy Comparison with Confidence: Benchmarks and Baselines (Koul, 2022) View paper
  - [40] Discrete Uncertainty Quantification For Offline Reinforcement Learning (JosÃ© Luis PÃ©rez, 2023) View paper
- Multi-Agent and Safety-Critical Domains
  - Multi-Agent Coordination in Offline Settings (1 papers)
  - [13] A model-based solution to the offline multi-agent reinforcement learning coordination problem (Barde, 2023) View paper
  - Autonomous Driving with World Models (2 papers)
  - [9] VL-SAFE: Vision-Language Guided Safety-Aware Reinforcement Learning with World Models for Autonomous Driving (Huang Zi-lin, 2025) View paper
  - [39] Offline Learning for Stochastic Multi-Agent Planning in Autonomous Driving (Villaflor, 2024) View paper
- Specialized Planning and Auxiliary Methods
  - Diffusion-Based and Generative Planning (1 papers)
  - [14] Planning as in-painting: A diffusion-based embodied task planning framework for environments under uncertainty (Yang Chengâ□□Fu, 2023) View paper
  - Meta-RL and Context Adaptation (1 papers)
  - [34] Dream to Adapt: Meta Reinforcement Learning by Latent Context Imagination and MDP Imagination (Lu Wen, 2023) View paper
  - Human-AI Alignment and Interaction (1 papers)
  - [27] ICLR 2025 Workshop on Bidirectional Human-AI Alignment (H Shen, 2025) View paper
  - Auxiliary Techniques and Surveys (8 papers)
  - [17] Towards systematically engineering autonomous systems using reinforcement learning and planning (M. Wirsing, 2023) View paper
  - [20] DURABLE-RL: A Dynamic Uncertainty-aware Hybrid Reinforcement Learning Framework with Adaptive Buffer and Ensemble Modelling. (SH Abbood, 2025) View paper
  - [32] Clinically Motivated Sequential Decision Making Under Uncertainty in Offline Settings (Killian, 2024) View paper
  - [35] Building Reliable Autonomous Agents: A Causal Perspective (Deng, 2024) View paper
  - [36] Efficient Methods for Machine Learning in Sequential Decision Making (Osi□ski, 2023) View paper
  - [43] Offline replay supports planning in human reinforcement learning. (Ida Momennejad, 2019) View paper
  - [44] Tool Learning with Large Language Models (Youcef, n.d.) View paper
  - [45] View-Imagination: Enhancing Visuomotor Control with Adaptive View Synthesis (D Lee, n.d.) View paper
- DreamPhase and Core Framework ★ (1 papers)
  - [0] DreamPhase: Offline Imagination and Uncertainty-Guided Planning for Large-Language-Model Agents (Anon et al., 2026) View paper

Narrative

Core task: offline imagination and uncertainty-guided planning for language model agents. The field structure reflects a convergence of model-based reinforcement learning and large language model capabilities, organized into several complementary branches. World Model Learning and Latent Dynamics encompasses methods that build predictive models of environment transitions, often through learned representations or generative processes (e.g., Dreamwalker[1], Dreamernav[10]). Uncertainty Quantification in Model-Based RL addresses the challenge of estimating epistemic and aleatoric uncertainty to guide exploration and avoid overconfident predictions in learned dynamics (e.g., MOReL[8], Uncertainty Driven Imagination[5]). LLM-Based Planning and Reasoning leverages the reasoning and knowledge capabilities of language models for sequential decision-making (e.g., Interactive Planning LLM[6], NavCoT[7]), while Policy

Learning and Value Estimation focuses on deriving effective control policies from imagined or real trajectories. Multi-Agent and Safety-Critical Domains and Specialized Planning branches handle coordination, safety constraints, and domain-specific adaptations, ensuring robustness in complex or high-stakes settings.

A particularly active tension lies between purely model-free LLM planning approaches and hybrid methods that combine learned world models with uncertainty-aware rollouts. Works like Kalm[3] and Knowledgeable Agents[2] explore how language models can internalize domain knowledge for planning, yet they often lack explicit uncertainty estimates that guard against compounding errors in long-horizon imagination. DreamPhase[0] sits at the intersection of these themes, residing in the DreamPhase and Core Framework branch. It emphasizes offline imagination—generating hypothetical trajectories without environment interaction—paired with uncertainty quantification to prune unreliable rollouts, bridging the gap between classical model-based RL (e.g., MOReL[8]) and modern LLM reasoning. Compared to Kalm[3], which focuses on knowledge integration, DreamPhase[0] prioritizes uncertainty-driven selectivity in imagined futures, offering a principled mechanism to balance exploration breadth with epistemic caution in language-driven agents.

Related Works in Same Category

No sibling papers and no sibling subtopics were found under the same parent taxonomy node; the paper appears structurally isolated in the taxonomy.

Contributions Analysis

Overall novelty summary. DreamPhase proposes a modular framework combining offline imagination through a learned latent world model, uncertainty-aware value scoring, and safety-gated filtering to guide LLM-based agents. The paper occupies its own singleton leaf ('DreamPhase and Core Framework') within the taxonomy, indicating it is the sole representative of this specific integration approach. This isolated position suggests the work synthesizes elements from multiple established branches—world model learning, uncertainty quantification, and LLM-based planning—into a novel architectural combination not directly replicated by other surveyed methods.

The taxonomy reveals substantial activity in neighboring branches: 'World Model Learning and Latent Dynamics' contains four leaves with methods like Dreamwalker and Dreamernav focusing on forward dynamics for embodied tasks, while 'Uncertainty Quantification in Model-Based RL' includes ensemble-based and penalty-driven approaches (e.g., MOReL, MOPO). 'LLM-Based Planning and Reasoning' encompasses five leaves addressing navigation, dialogue, and search-augmented planning. DreamPhase diverges by integrating latent-space imagination with uncertainty-aware filtering specifically for LLM agents, whereas sibling branches typically address these components in isolation or apply them to non-LLM settings.

Among 26 candidates examined, the framework-level contribution (Contribution A) shows no clear refutation across 10 candidates, suggesting the specific architectural integration is relatively unexplored. However, uncertainty-aware value estimation (Contribution B) encounters three refutable candidates among six examined, indicating substantial prior work on uncertainty quantification in model-based RL. The language reflection mechanism (Contribution C) finds one refutable candidate among ten, pointing to some overlap with existing LLM prompting or self-refinement techniques. The limited search scope (26 candidates, not exhaustive) means these findings reflect top-K semantic matches rather than comprehensive field coverage.

Given the constrained literature search, DreamPhase appears to occupy a relatively sparse intersection—combining offline imagination, uncertainty filtering, and LLM control—though individual components draw on well-established techniques. The singleton taxonomy position and low refutation rate for the framework contribution suggest architectural novelty, while higher overlap for uncertainty and reflection mechanisms indicates these elements build incrementally on prior uncertainty quantification and LLM prompting research. The analysis captures top-30 semantic neighbors, leaving open the possibility of additional related work beyond this scope.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: DreamPhase framework for offline imagination-based planning

Description: The authors propose DreamPhase, a modular agent framework that uses a learned latent world model to simulate multiple future trajectories offline in latent space. These imagined branches are evaluated using uncertainty-aware value estimates and filtered through a safety gate before execution, enabling planning without real environment interactions.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Offline Trajectory Optimization for Offline Reinforcement Learning

URL: View paper

Brief Assessment

Offline Trajectory Optimization[59] focuses on model-based offline RL with trajectory augmentation for policy training, not on LLM-based agents with uncertainty-aware value filtering and language reflections for interactive decision-making tasks.

---

2. UMBRELLA: Uncertainty-Aware Model-Based Offline Reinforcement Learning Leveraging Planning

URL: View paper

Brief Assessment

UMBRELLA[41] focuses on model-based offline RL for autonomous driving using stochastic dynamics models and MPC planning, not on LLM agents with latent world models for web/embodied tasks.

---

3. Bachelor's Thesis Submitted in 2025

URL: View paper

Brief Assessment

Bachelor Thesis[62] appears to focus on sensor fusion and EKF frameworks for robotics, not on offline imagination-based planning with learned latent world models for LLM agents. The minimal context provided does not demonstrate prior work on the specific combination of latent world models, uncertainty-aware value filtering, and language-based reflections for agent planning.

---

4. Offline trajectory generalization for offline reinforcement learning

URL: View paper

Brief Assessment

Offline Trajectory Generalization[57] focuses on augmenting offline RL datasets through long-horizon trajectory simulation using world transformers, not on LLM-based agents with uncertainty-aware value filtering and language reflections as in DreamPhase.

---

5. Reflect-then-Plan: Offline Model-Based Planning through a Doubly Bayesian Lens

URL: View paper

Brief Assessment

Reflect-then-Plan[31] focuses on offline model-based planning through Bayesian posterior estimation over environment dynamics, whereas DreamPhase uses a learned latent world model to simulate trajectories in latent space with uncertainty-aware value filtering and language-based reflections for LLM agents.

---

6. Constrained latent action policies for model-based offline reinforcement learning

URL: View paper

Brief Assessment

Constrained Latent Actions[58] focuses on offline RL with learned latent world models for dynamics prediction and policy constraint, not on LLM-based agents with uncertainty-aware value filtering and natural-language reflections as in DreamPhase.

---

7. Offline Reinforcement Learning with Policy Guidance and Uncertainty Estimation

URL: View paper

Brief Assessment

Policy Guidance Uncertainty[61] focuses on offline RL with static datasets and distribution shift mitigation through policy constraints and Q-function generalization. DreamPhase addresses a different problem: online agent planning through learned latent world models that simulate future trajectories for LLM-based agents in interactive environments.

---

8. Integrating World Models into Vision Language Action and Navigation: A Comprehensive Survey

URL: View paper

Brief Assessment

World Models Survey[63] is a survey paper that reviews existing work on world models in vision-language-action domains. It does not present a novel framework for offline imagination-based planning with uncertainty-aware value filtering and safety gates as proposed in DreamPhase.

---

9. Uncertainty-aware model-based offline reinforcement learning for automated driving

URL: View paper

Brief Assessment

Uncertainty Automated Driving[4] focuses on automated driving with stochastic dynamics models for multi-agent traffic scenarios, not general-purpose LLM agents across web/science/embodied domains. The technical approaches differ fundamentally in application domain and architecture.

---

10. Offline Reinforcement Learning from Images with Latent Space Models

URL: View paper

Brief Assessment

Offline Images Latent[60] focuses on offline RL with latent-state dynamics models for visual observations in robotics tasks, not on LLM-based agents with uncertainty-aware value filtering and natural-language reflections for interactive decision-making environments.

Contribution 2: Uncertainty-aware value estimation with theoretical regret bound

Description: The authors develop a risk-aware branch selection mechanism that scores imagined trajectories using value estimates penalized by uncertainty measures. They provide a theoretical regret bound that relates cumulative regret to model approximation error and mis-gating rate, offering formal guarantees on decision quality.

This contribution was assessed against **6 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. UAMDP: Uncertainty-Aware Markov Decision Process for Risk-Constrained Reinforcement Learning from Probabilistic Forecasts

URL: View paper

Brief Assessment

UAMDP[68] focuses on Bayesian forecasting with CVaR constraints for financial/inventory domains, not latent world models for LLM agents. The regret analysis addresses posterior sampling under risk constraints, not imagination-based planning with mis-gating rates.

---

2. Multi-agent Uncertainty-Aware Pessimistic Model-Based Reinforcement Learning for Connected Autonomous Vehicles

URL: View paper

Prior Art Analysis

Multi-agent Uncertainty CAV[65] demonstrates prior work on uncertainty-penalized value estimation with theoretical regret bounds in model-based reinforcement learning. The candidate paper presents a pessimistic model-based framework that scores trajectories using value estimates penalized by uncertainty measures and provides formal regret bounds relating cumulative regret to model approximation error. This directly parallels the original paper's contribution of risk-aware branch selection with uncertainty penalties and regret bounds linking decision quality to model error and mis-gating rate.

Evidence

Evidence 1 - **Rationale**: Both papers employ pessimistic optimization frameworks that penalize uncertainty in value estimation to ensure safe decision-making in autonomous vehicle contexts. - **Original**: imagined trajectories are scored by value minus an uncertainty penalty and accepted only if a confidence gate passes; we provide a regret bound o(√ t ε) +bρtthat links decision quality to model error and mis-gating. - **Candidate**: we propose ma-pmbrl, a novel multi-agent pessimistic model-based reinforcement learning framework for cavs, incorporating a max-min optimization approach to enhance robustness and decision-making. to mitigate the inherent subjectivity of uncertainty estimation in mbrl and avoid incurring catastrophi...

Evidence 2 - **Rationale**: Both papers provide theoretical regret bounds that relate model approximation error to policy suboptimality, establishing formal guarantees on decision quality. - **Original**: the √ t εterm captures error from model approximation; bρt accounts for occasional acceptance of unreliable imagined branches. whenεis small andρis rare, regret grows sublinearly. - **Candidate**: by bounding the suboptimality of the resulting policy under mild theoretical assumptions, we successfully establish pac guarantees for ma-pmbrl, demonstrating that the proposed framework represents a significant step toward scalable, efficient, and reliable multi-agent decision-making for cavs.

---

3. No-Regret Replanning under Uncertainty

URL: View paper

Prior Art Analysis

No-Regret Replanning[69] demonstrates prior work on uncertainty-penalized value estimation with formal regret bounds in the context of model-based planning. The candidate paper presents UCB-replanning, which scores trajectories using value estimates penalized by uncertainty measures derived from Gaussian processes, and provides a theoretical regret bound that relates cumulative regret to model approximation error. This directly addresses the same technical contribution claimed in the original paper: combining uncertainty-aware value scoring with theoretical guarantees on decision quality through regret analysis.

Evidence

Evidence 1 - **Rationale**: Both papers claim to provide no-regret guarantees for uncertainty-aware planning algorithms. The candidate explicitly states 'appealing no-regret properties' for UCB-style algorithms adapted to path planning, establishing prior work on this contribution. - **Original**: imagined trajectories are scored by value minus an uncertainty penalty and accepted only if a confidence gate passes; we provide a regret bound o(√ t ε) +bρtthat links decision quality to model error and mis-gating. - **Candidate**: we propose ucb style algorithms that are popular in the bandit settings and show how those analyses can be adapted to the online robotic path planning problems. the proposed algorithm trades-off exploration and exploitation in near-optimal manner and has appealing no-regret properties.

Evidence 2 - **Rationale**: Both papers formalize regret as the difference between optimal and actual decisions. The candidate provides a concrete regret definition and proves convergence to zero, demonstrating prior theoretical work on regret bounds for uncertainty-aware planning. - **Original**: risk-aware branch selection with guarantees.imagined trajectories are scored by value minus an uncertainty penalty and accepted only if a confidence gate passes; we provide a regret bound o(√ t ε) +bρtthat links decision quality to model error and mis-gating. - **Candidate**: rt = 1 t [t∑ t=1 max i∗ t ∈[k] ft,i∗ t ({g(xi∗ t t,j)}l j=0) - t∑ t=1 ft,it ({g(xit t,j)}l j=0)] (2) namely, at each round t, we measure how much more reward the robot could gain if it could picki∗ t instead of it at xt. the goal is to make regret converges to zero so that in average the robot ...

Evidence 3 - **Rationale**: Both papers score trajectories using uncertainty-aware value estimation. The candidate explicitly constructs confidence intervals from GP uncertainty and uses them to compute upper confidence bounds on trajectory rewards, demonstrating the same technical approach claimed as novel in the original. - **Original**: we perform uncertainty-aware value estimation over these imagined branches, scoring each using predicted rewards and entropy-based confidence metrics - **Candidate**: to design the confidence interval for each trajectory's reward at every step, we first extract a confidence interval of the uncertain variable v from gp. we then use the lipschitz continuity of the reward function of each trajectory to transfer the confidence interval of the uncertain variable v to the ...

---

4. Uncertainty-driven exploration in sparse model-based reinforcement learning

URL: View paper

Prior Art Analysis

The candidate paper Uncertainty Sparse Exploration[66] presents a theoretical regret bound for model-based RL that explicitly relates cumulative regret to model approximation error and uncertainty quantification. The candidate establishes regret bounds that scale with model sparsity under sub-gaussian noise assumptions, providing formal guarantees on decision quality through uncertainty-penalized value estimation. This demonstrates prior work on uncertainty-aware value estimation with regret bounds in model-based planning, directly addressing the same theoretical contribution claimed by the original paper.

Evidence

Evidence 1 - **Rationale**: Both papers provide theoretical regret bounds that relate cumulative regret to model approximation error. The candidate explicitly derives regret bounds under uncertainty quantification, demonstrating prior work on this theoretical contribution. - **Original**: imagined trajectories are scored by value minus an uncertainty penalty and accepted only if a confidence gate passes; we provide a regret bound o(√ t ε) +bρtthat links decision quality to model error and mis-gating. - **Candidate**: under the assumptions of a sparse dynamical system with a bounded noise density function, the regret bound depends not only on the length of the trajectoriesh, the number of trajectoriesk, and the state space dimensionps, which are standard in the theoretical literature of rl, but also on the number...

Evidence 2 - **Rationale**: Both papers describe uncertainty-aware value estimation mechanisms. The candidate explicitly discusses using model uncertainty to score trajectories and guide exploration, which is the same core mechanism claimed as novel in the original paper. - **Original**: we perform uncertainty-aware value estimation over these imagined branches, scoring each using predicted rewards and entropy-based confidence metrics - **Candidate**: the idea of lower confidence-based continuous control (lc3) [30] algorithm is to exploit the uncertainty of the model to perform better exploration, resulting in higher sample efficiency. the algorithm is based onoptimism in the face of uncertaintythat takes advantage of optimistic policies to guide...

Evidence 3 - **Rationale**: Both papers use uncertainty-based gating mechanisms to accept or reject imagined trajectories. The candidate's confidence ball approach with radius β serves the same function as the original's confidence gate, demonstrating prior work on this mechanism. - **Original**: risk-aware branch selection with guarantees.imagined trajectories are scored by value minus an uncertainty penalty and accepted only if a confidence gate passes - **Candidate**: instead of simply using this linear approximation to generate data, the algorithm exploits the model uncertainty to build a confidence ball over parameters and obtain a set of dynamics, close to the approximated one, that could represent the real environment. ball = {w | ||(w - cw)(σ1/2)||2 2 ≤ β} w...

Evidence 4 - **Rationale**: Both papers claim sample efficiency as a key advantage of their uncertainty-aware model-based approach. The candidate establishes this benefit in the context of model-based RL with uncertainty quantification. - **Original**: this approach yields three key advantages: (i) sample efficiency:web-based tasks that previously required approximately 40 real clicks per episode now converge in fewer than 10 - **Candidate**: model-based reinforcement learning(mbrl) [41] refers to a class of rl algorithms that exploit a surrogate environment to reduce the number of interactions with the real one. the presence of a model of the dynamics allows the agent to simulate different trajectories without interacting with the real ...

---

5. Leveraging Learned Models for Robust Decision Optimization and Offline Reinforcement Learning

URL: View paper

Brief Assessment

Learned Models Optimization[67] focuses on decision optimization and offline RL with learned predictive models for uncertain parameters, not on uncertainty-penalized value estimation for model-based planning with regret bounds as in the original paper.

---

6. Model-Based Reinforcement Learning in Diffusion Environments: Value-Aware Estimation and Financial Application

URL: View paper

Brief Assessment

Diffusion Environments MBRL[64] focuses on diffusion-based stochastic environments with continuous state spaces, whereas the original paper addresses discrete interactive environments (web navigation, embodied tasks) with latent world models and language-based reflections. The candidate's regret analysis applies to uncertainty-penalized control in diffusion settings, not to the original's imagination-based planning with safety gates and language conditioning.

Contribution 3: Language reflection mechanism for zero-tuning policy control

Description: The authors introduce a mechanism that distills the best imagined trajectory into concise natural-language reflections and summaries. These are injected into the policy LLM prompt to guide action selection without fine-tuning the model, enabling interpretable behavior improvement while keeping the LLM frozen.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Reflexion: Language agents with verbal reinforcement learning

URL: View paper

Prior Art Analysis

Reflexion[47] demonstrates that a language reflection mechanism for steering frozen LLM policies without fine-tuning was already proposed and implemented. Both papers distill trajectory experiences into natural-language reflections that are injected into the policy LLM prompt to guide future actions while keeping the model frozen. Reflexion[47] explicitly describes converting feedback into textual summaries added as context for the LLM agent, which acts as semantic gradient signals to improve behavior without parameter updates —the same core mechanism claimed as novel in the original paper.

Evidence

Evidence 1 - **Rationale**: Both papers describe the same mechanism: using verbal reflections stored in memory to guide a frozen LLM policy without weight updates. This demonstrates prior work on the claimed contribution. - **Original**: language reflections for zero-tuning control.the selected branch is distilled into a short reflection and summary that are injected into the next prompt, steering a frozen llm toward higher-quality actions while remaining interpretable and requiring no parameter updates. - **Candidate**: we propose reflexion, a novel framework to reinforce language agents not by updating weights, but instead through linguistic feedback. concretely, reflexion agents verbally reflect on task feedback signals, then maintain their own reflective text in an episodic memory buffer to induce better decisio...

Evidence 2 - **Rationale**: This pair demonstrates that Reflexion[47] already implemented the mechanism of distilling trajectories into natural-language reflections stored in memory to guide future decisions of a frozen LLM, which is the exact contribution claimed. - **Original**: we distill the best low-risk trajectory into a natural-language reflection, which is injected back into the llm prompt to influence future decisions. crucially, the llm policy remains entirelyfrozenthroughout this process. its behavior evolves solely through internal simulation and language-based fe... - **Candidate**: the self-reflection model instantiated as an llm, plays a crucial role in the reflexion framework by generating verbal self-reflections to provide valuable feedback for future trials. given a sparse reward signal, such as a binary success status (success/fail), the current trajectory, and its persis...

---

2. Unveiling the Latent Directions of Reflection in Large Language Models

URL: View paper

Brief Assessment

Latent Directions Reflection[55] focuses on mechanistic interpretability of reflection in LLMs through activation steering and latent directions, not on distilling imagined trajectories into natural-language reflections for guiding frozen policy LLMs in interactive decision-making environments.

---

3. Re-rest: Reflection-reinforced self-training for language agents

URL: View paper

Brief Assessment

Re-rest[52] focuses on training-time reflection to improve self-training data quality, not on zero-tuning policy control. The reflector in Re-rest[52] is used during training to correct low-quality samples with environmental feedback, then the corrected samples are used to finetune the agent. This differs fundamentally from DreamPhase's approach of using reflections to steer a frozen LLM at inference time without any parameter updates.

---

4. Re2llm: reflective reinforcement large language model for session-based recommendation

URL: View paper

Brief Assessment

Re2llm[50] focuses on session-based recommendation using LLM self-reflection to generate hints for item recommendation, not on steering frozen LLM policies for multi-step decision-making in interactive environments without fine-tuning.

---

5. IROTE: Human-like Traits Elicitation of Large Language Model via In-Context Self-Reflective Optimization

URL: View paper

Brief Assessment

IROTE[51] focuses on eliciting human-like traits (personality, values) in LLMs through self-reflection for personalized applications and social simulations. The original paper uses reflection to distill imagined trajectories for guiding action selection in decision-making environments. These are fundamentally different application domains and mechanisms.

---

6. Self-Steering Language Models

URL: View paper

Brief Assessment

Self-Steering Models[54] focuses on generating inference programs that coordinate parallel search via sequential Monte Carlo, not on distilling imagined trajectories into natural-language reflections for zero-tuning policy control as in the original paper.

---

7. Seal: Steerable reasoning calibration of large language models for free

URL: View paper

Brief Assessment

Seal[49] focuses on calibrating chain-of-thought reasoning in LLMs by steering latent representations to reduce redundant reflection/transition thoughts, not on distilling imagined trajectories into language reflections for frozen policy control in interactive environments.

---

8. Instruction-Level Weight Shaping: A Framework for Self-Improving AI Agents

URL: View paper

Brief Assessment

Instruction-Level Weight Shaping[56] focuses on updating system instructions through reflection and user feedback in conversational agents, eventually distilling these into model weights. The original paper's mechanism distills imagined trajectories into natural-language reflections that condition a frozen LLM for sequential decision-making in interactive environments without any weight updates. These are fundamentally different approaches to using language-based reflection.

---

9. A Zero-Shot Language Agent for Computer Control with Structured Reflection

URL: View paper

Brief Assessment

Structured Reflection Agent[53] focuses on zero-shot computer control with structured reflection for iterative task improvement, not on distilling imagined trajectories from a latent world model into language reflections for frozen LLM steering as in the original paper.

---

10. Exploring large language models for communication games: An empirical study on werewolf

URL: View paper

Brief Assessment

Werewolf LLM Study[48] focuses on communication games (Werewolf) using reflection for question-answering and experience retrieval, not for distilling imagined trajectories into policy-guiding reflections in sequential decision-making environments.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] DreamPhase: Offline Imagination and Uncertainty-Guided Planning for Large-Language-Model Agents View paper
- [1] Dreamwalker: Mental planning for continuous vision-language navigation View paper
- [2] Knowledgeable agents by offline reinforcement learning from large language model rollouts View paper
- [3] Kalm: Knowledgeable agents by offline reinforcement learning from large language model rollouts View paper
- [4] Uncertainty-aware model-based offline reinforcement learning for automated driving View paper
- [5] Uncertainty-driven imagination for continuous deep reinforcement learning View paper
- [6] Interactive planning using large language models for partially observable robotic tasks View paper
- [7] NavCoT: Boosting LLM-Based Vision-and-Language Navigation via Learning Disentangled Reasoning View paper
- [8] MOReL : Model-Based Offline Reinforcement Learning View paper
- [9] VL-SAFE: Vision-Language Guided Safety-Aware Reinforcement Learning with World Models for Autonomous Driving View paper
- [10] Dreamernav: Learning-based autonomous navigation in dynamic indoor environments using world models View paper
- [11] Planning without Search: Refining Frontier LLMs with Offline Goal-Conditioned RL View paper
- [12] Open-world reinforcement learning over long short-term imagination View paper
- [13] A model-based solution to the offline multi-agent reinforcement learning coordination problem View paper
- [14] Planning as in-painting: A diffusion-based embodied task planning framework for environments under uncertainty View paper
- [15] SUMO: Search-Based Uncertainty Estimation for Model-Based Offline Reinforcement Learning View paper
- [16] Zero-shot goal-directed dialogue via rl on imagined conversations View paper
- [17] Towards systematically engineering autonomous systems using reinforcement learning and planning View paper
- [18] Bayesian Uncertainty Estimation for Targeted Counterfactual Experience Generation in Reinforcement Learning View paper
- [19] Maximization Operator Uncertainty Penalty for Model-Based Offline Reinforcement Learning View paper
- [20] DURABLE-RL: A Dynamic Uncertainty-aware Hybrid Reinforcement Learning Framework with Adaptive Buffer and Ensemble Modelling. View paper
- [21] Imagination-Limited Q-Learning for Offline Reinforcement Learning View paper
- [22] VISTAv2: World Imagination for Indoor Vision-and-Language Navigation View paper
- [23] Stairway to Autonomy: Hierarchical Decision-Making for LLM-Guided Planning, Bandit-Driven Exploration, and Multi-Agent Navigation View paper
- [24] Monte Carlo Planning with Large Language Model for Text-Based Game Agents View paper
- [25] DMWM: Dual-Mind World Model with Long-Term Imagination View paper
- [26] A Simple Unified Uncertainty-Guided Framework for Offline-to-Online Reinforcement Learning View paper
- [27] ICLR 2025 Workshop on Bidirectional Human-AI Alignment View paper
- [28] Adaptive Stress Testing Black-Box LLM Planners View paper
- [29] NavForesee: A Unified Vision-Language World Model for Hierarchical Planning and Dual-Horizon Navigation Prediction View paper
- [30] Long-Horizon Model-Based Offline Reinforcement Learning Without Conservatism View paper
- [31] Reflect-then-Plan: Offline Model-Based Planning through a Doubly Bayesian Lens View paper
- [32] Clinically Motivated Sequential Decision Making Under Uncertainty in Offline Settings View paper
- [33] Uncertainty Estimation Using Riemannian Model Dynamics for Offline Reinforcement Learning View paper
- [34] Dream to Adapt: Meta Reinforcement Learning by Latent Context Imagination and MDP Imagination View paper
- [35] Building Reliable Autonomous Agents: A Causal Perspective View paper
- [36] Efficient Methods for Machine Learning in Sequential Decision Making View paper
- [37] Offline Policy Comparison with Confidence: Benchmarks and Baselines View paper
- [38] L-MBOP-E: Latent-Model Based Offline Planning with Extrinsic Policy Guided Exploration View paper
- [39] Offline Learning for Stochastic Multi-Agent Planning in Autonomous Driving View paper
- [40] Discrete Uncertainty Quantification For Offline Reinforcement Learning View paper
- [41] UMBRELLA: Uncertainty-Aware Model-Based Offline Reinforcement Learning Leveraging Planning View paper
- [42] Offline Reinforcement Learning with Reverse Model-based Imagination View paper
- [43] Offline replay supports planning in human reinforcement learning. View paper
- [44] Tool Learning with Large Language Models View paper
- [45] View-Imagination: Enhancing Visuomotor Control with Adaptive View Synthesis View paper
- [46] Zero-Shot Goal Dialogue via Reinforcement Learning on Imagined Conversations View paper
- [47] Reflexion: Language agents with verbal reinforcement learning View paper
- [48] Exploring large language models for communication games: An empirical study on werewolf View paper
- [49] Seal: Steerable reasoning calibration of large language models for free View paper
- [50] Re2llm: reflective reinforcement large language model for session-based recommendation View paper
- [51] IROTE: Human-like Traits Elicitation of Large Language Model via In-Context Self-Reflective Optimization View paper
- [52] Re-rest: Reflection-reinforced self-training for language agents View paper
- [53] A Zero-Shot Language Agent for Computer Control with Structured Reflection View paper
- [54] Self-Steering Language Models View paper
- [55] Unveiling the Latent Directions of Reflection in Large Language Models View paper

- [56] Instruction-Level Weight Shaping: A Framework for Self-Improving AI Agents View paper
- [57] Offline trajectory generalization for offline reinforcement learning View paper
- [58] Constrained latent action policies for model-based offline reinforcement learning View paper
- [59] Offline Trajectory Optimization for Offline Reinforcement Learning View paper
- [60] Offline Reinforcement Learning from Images with Latent Space Models View paper
- [61] Offline Reinforcement Learning with Policy Guidance and Uncertainty Estimation View paper
- [62] Bachelor's Thesis Submitted in 2025 View paper
- [63] Integrating World Models into Vision Language Action and Navigation: A Comprehensive Survey View paper
- [64] Model-Based Reinforcement Learning in Diffusion Environments: Value-Aware Estimation and Financial Application View paper
- [65] Multi-agent Uncertainty-Aware Pessimistic Model-Based Reinforcement Learning for Connected Autonomous Vehicles View paper
- [66] Uncertainty-driven exploration in sparse model-based reinforcement learning View paper
- [67] Leveraging Learned Models for Robust Decision Optimization and Offline Reinforcement Learning View paper
- [68] UAMDP: Uncertainty-Aware Markov Decision Process for Risk-Constrained Reinforcement Learning from Probabilistic Forecasts View paper
- [69] No-Regret Replanning under Uncertainty View paper