International Journal of Advanced Engineering Science and Information Technology (IJAESIT)

|ISSN 2349-3216| Volume 8, Issue 5, September-October 2025| Bimonthly, Peer Reviewed and Scholarly Indexed Journal|

DOI: 10.15662/IJAESIT.2025.0805001

Long-Horizon Planning and Decision Algorithms for Autonomous Intelligent Agents

John Okello

Muteesa I Royal University, Masaka, Uganda

ABSTRACT: Long-horizon planning and decision algorithms enable autonomous intelligent agents to make sequential decisions over extended time frames in complex, uncertain environments. Unlike short-term reactive strategies, long-horizon planning involves anticipating future states, evaluating multi-step outcomes, and optimizing cumulative performance with respect to strategic objectives. This capability is essential in domains such as autonomous vehicles, mobile robotics, space exploration, automated logistics, strategic gameplay, defense systems, and intelligent manufacturing. Long-horizon decision making integrates core techniques from classical planning, reinforcement learning, probabilistic reasoning, hierarchical control, model predictive control, and heuristic search, often requiring trade-offs between computational tractability and optimality. This research synthesizes foundational theories and recent advances in long-horizon planning, compares algorithmic paradigms, and assesses their performance in autonomous agent applications. Through systematic literature review and analytical synthesis, we describe representative frameworks including Markov decision processes (MDPs), Partially Observable MDPs (POMDPs), hierarchical reinforcement learning, Monte Carlo Tree Search (MCTS), model-based planning, and optimization-based strategies. We highlight challenges in scalability, uncertainty handling, reward sparsity, and real-time execution, and discuss solution approaches such as abstraction, temporal hierarchies, simulation-based planning, and transfer learning. Empirical findings indicate that hybrid methods combining learning and planning outperform pure approaches in dynamic, long-horizon scenarios. We conclude with directions for improving interpretability, safety, and generalization in long-horizon planning for autonomous intelligent agents.

KEYWORDS: Long-horizon planning, autonomous agents, decision algorithms, reinforcement learning, hierarchical planning, MDP, POMDP, Monte Carlo Tree Search, model-based planning

I. INTRODUCTION

Autonomous intelligent agents — systems that perceive, reason, and act in dynamic environments without human intervention — increasingly permeate both research and real-world applications. These agents must often make decisions that affect not just immediate outcomes but long-term goals, requiring the ability to plan and execute sequences of actions with foresight. **Long-horizon planning and decision algorithms** enable agents to anticipate future consequences, optimize cumulative reward, balance risk and reward over time, and adapt to uncertainty while operating under real-time constraints. As autonomous systems expand into safety-critical domains — self-driving vehicles navigating complex traffic environments, robotic systems executing industrial operations, planetary rovers exploring unpredictable terrains, or agents coordinating multi-robot task allocations — the need for robust long-horizon decision making becomes paramount.

The concept of long-horizon planning is grounded in theories of sequential decision making and control, where an agent's action at any given time influences future states and available actions. Classical frameworks such as **Markov decision processes (MDPs)** and **Partially Observable Markov Decision Processes (POMDPs)** formalize this process with state transition models and reward functions, offering principled structures for reasoning about future impacts. These models inherently consider long horizons by optimizing expected cumulative rewards across sequences of decisions. However, real-world applications often involve large state spaces, partial observability, stochastic dynamics, and constraints on computational resources, challenging the direct application of classical models.

To address this, research has bifurcated into multiple strategies that balance **optimality** with **tractability**. On one axis, **model-based planning algorithms** explicitly use a model of the environment to simulate future trajectories and evaluate action sequences. Techniques such as **Monte Carlo Tree Search (MCTS)** sample plausible futures to guide real-time planning. **Model Predictive Control (MPC)** uses optimization over finite receding horizons to approximate long-horizon objectives while continuously updating based on new observations. On the other axis, **learning-based methods** such as **reinforcement learning (RL)** enable agents to learn policies and value functions from interaction experience, embedding long-term information into learned representations.

|ISSN 2349-3216| Volume 8, Issue 5, September-October 2025| Bimonthly, Peer Reviewed and Scholarly Indexed Journal|

DOI: 10.15662/IJAESIT.2025.0805001

Hierarchical approaches — which decompose long-horizon problems into subgoals or temporal abstractions — have emerged to improve scalability. For example, **options frameworks** and **hierarchical reinforcement learning** define higher-level actions (options) that encapsulate sequences of primitive actions, effectively reducing planning depth. Other strategies such as **state abstraction** and **factored representations** reduce the effective dimensionality of planning problems by capturing only relevant features.

Despite notable progress, long-horizon planning presents persistent challenges. **Uncertainty** about environmental dynamics — due to stochasticity, partial observability, or model inaccuracies — complicates forward planning. **Reward sparsity** makes it difficult for agents to discover beneficial long-term strategies without guidance or intermediate feedback. **Scalability** remains an issue since naive enumeration of future trajectories grows exponentially with horizon length. Additionally, real-time constraints require approximations to ensure decisions are rendered within acceptable latency. Safety and interpretability are further concerns in domains where autonomous decisions have critical consequences; agents must not only plan effectively but also justify their decisions in human-centric contexts.
This paper provides a comprehensive exploration of long-horizon planning and decision algorithms for autonomous intelligent agents. We trace foundational theories, examine leading algorithmic paradigms, review state-of-the-art approaches, and analyze empirical performance trends. We aim to synthesize trajectories in research, identify strengths and weaknesses of competing methods, and outline avenues for future advancement. The following sections cover the literature review, detailed methodology for comparative synthesis, advantages and disadvantages, results and discussion of algorithmic performance and conceptual trade-offs, followed by a conclusion and future research directions.

II. LITERATURE REVIEW

Long-horizon planning and decision algorithms originate in the study of sequential decision processes, particularly **Markov decision processes (MDPs)**, which model decision problems where outcomes are partly under the control of an agent and partly stochastic. Early work formalized dynamic programming solutions, such as value iteration and policy iteration, which guarantee optimal policies for finite state and action spaces. However, the **curse of dimensionality** — the exponential growth of state space with problem size — limited their applicability in real systems.

To handle uncertainty and partial observability, **Partially Observable Markov Decision Processes (POMDPs)** extended MDPs by modeling uncertainty in state observations. Bayesian belief updates allow agents to maintain distributions over possible world states, enabling planning under uncertainty. However, exact POMDP solutions are computationally intractable for large problems, leading to approximate methods such as point-based value iteration and belief tree search.

Reinforcement learning (RL), popularized in the 1990s, offered a model-free alternative where agents learn value functions or policies through interaction. Temporal difference learning, Q-learning, and SARSA provided mechanisms for learning long-term value estimates without full models. RL's ability to handle high-dimensional tasks improved with integration of function approximation, culminating in **deep reinforcement learning** (DRL) techniques that combine deep neural networks with RL algorithms. DRL achieved landmark success in long-horizon tasks such as playing Atari games and the game of Go, where planning over long sequences is essential.

Model-based planning methods explicitly simulate forward dynamics. Monte Carlo Tree Search (MCTS), notably in the UCT (Upper Confidence bounds applied to Trees) variant, balances exploration and exploitation in the search tree. MCTS achieved remarkable results in strategic games like Go by enabling deep lookahead planning with selective sampling. **Model Predictive Control (MPC)**, rooted in control theory, optimizes action sequences over a finite horizon and updates plans as new observations become available, effectively approximating long-horizon objectives.

Hierarchical planning and **temporal abstractions** reduce complexity by organizing actions into macro-actions or options. The options framework in hierarchical reinforcement learning defines temporally extended actions, enabling agents to reason at multiple temporal scales. State abstraction methods cluster similar states to simplify planning. These techniques have shown success in robotic navigation, task hierarchies, and complex control tasks.

Hybrid techniques combining planning and learning have gained traction. **AlphaGo and AlphaZero** integrated deep learning with MCTS, using neural networks to predict value and policy priors, significantly improving search efficiency in long-horizon strategic planning. Other methods such as **World Models** learn latent models of environment dynamics to support planning in latent state space.

|ISSN 2349-3216| Volume 8, Issue 5, September-October 2025| Bimonthly, Peer Reviewed and Scholarly Indexed Journal|

DOI: 10.15662/IJAESIT.2025.0805001

Goal-conditioned RL techniques and **curriculum learning** address reward sparsity by providing intermediate goals and structured training regimes. Transfer learning and meta-learning enable agents to leverage experience from related tasks to improve long-horizon planning performance in new tasks.

Recent research also explores **safe long-horizon planning**, where safety constraints must be satisfied over extended trajectories. Constrained MDPs and shielded RL ensure that safety requirements guide decision making. Explainability in planning — enabling humans to understand why an agent chose a sequence of actions — has become paramount, especially in autonomous vehicles and human-robot interaction contexts.

Despite progress, challenges in dealing with uncertainty, scalability, reward sparsity, and real-time performance remain, motivating ongoing research in algorithm design and hybrid strategies.

III. RESEARCH METHODOLOGY

This research adopts a **mixed systematic methodology** to analyze long-horizon planning and decision algorithms for autonomous intelligent agents. First, a comprehensive **systematic literature survey** was conducted across major academic databases including IEEE Xplore, ACM Digital Library, Web of Science, Scopus, and arXiv to collect peer-reviewed publications, conference papers, and seminal books published between foundational works before 2002 and contemporary research through 2024. Search queries included terms such as *MDP planning*, *POMDP approximate methods*, *reinforcement learning long horizon*, *Monte Carlo Tree Search*, *hierarchical reinforcement learning*, *model predictive control*, and *hybrid planning and learning.*

Publications were selected based on relevance to long-horizon planning — specifically where horizon lengths exceed immediate reactive decision steps and involve future reward optimization, strategic evaluation, or multi-stage decision sequences. Exclusion criteria removed works focused purely on short-term reactive control, local optimization without temporal depth, or domain-specific implementations lacking generalizable algorithmic insights.

Selected literature was categorized into major algorithmic families: classical planning (MDPs/POMDPs), model-based planning (MCTS, MPC), reinforcement learning (model-free and model-based RL), hierarchical approaches, and hybrid planning-learning systems. Each work was examined for methodology, assumptions, algorithmic contributions, theoretical properties (optimality, convergence), computational characteristics, and empirical performance on benchmark tasks or real-world applications.

The second phase involved **architectural analysis** of representative algorithms. For classical planning, value iteration, policy iteration, point-based POMDP solvers, and belief tree search algorithms were dissected to understand their handling of long horizons, belief representations, and scalability strategies. For model-based planning, MCTS variants, heuristic enhancements, rollout strategies, and integration with learned models were analyzed. Model Predictive Control's formulation (cost functions, prediction horizons, updates) was reviewed for how finite horizon optimization approximates long-horizon objectives while ensuring real-time applicability.

Reinforcement learning analysis covered temporal difference learning, Q-learning, policy gradient methods, actor-critic algorithms, and recent deep RL architectures. Particular attention was given to how long-horizon dependencies are captured — e.g., through discount factors, eligibility traces, or recurrent network architectures — and how reward sparsity is mitigated (e.g., via intrinsic motivation, hindsight experience replay).

Hierarchical methods were analyzed in terms of temporal abstraction mechanisms, definition of options or macro-actions, subgoal discovery strategies, and integration of high-level planning with low-level control. Frameworks such as the options framework, feudal RL, and hierarchical DQN extensions were studied for their efficacy in reducing planning depth.

Hybrid planning-learning systems — especially those combining neural network prediction components with symbolic or search planners — were examined for architectural interfaces, data requirements, training regimes, and performance in tasks requiring deep lookahead. For example, AlphaZero's integration of learned policy/value networks with MCTS was studied for how learned priors speed up exploration and reduce computation.

The methodology also included a **comparative evaluation synthesis** where performance trends across algorithmic families were contrasted. Evaluation dimensions included **planning efficiency** (time per decision, scalability), **solution quality** (cumulative reward, optimality proximity), **robustness to uncertainty** (partial observability, stochastic

|ISSN 2349-3216| Volume 8, Issue 5, September-October 2025| Bimonthly, Peer Reviewed and Scholarly Indexed Journal|
DOI: 10.15662/IJAESIT.2025.0805001

transitions), **data requirements** (model learning overhead), **generalization** (transfer to unseen environments), and **interpretability** of decisions.

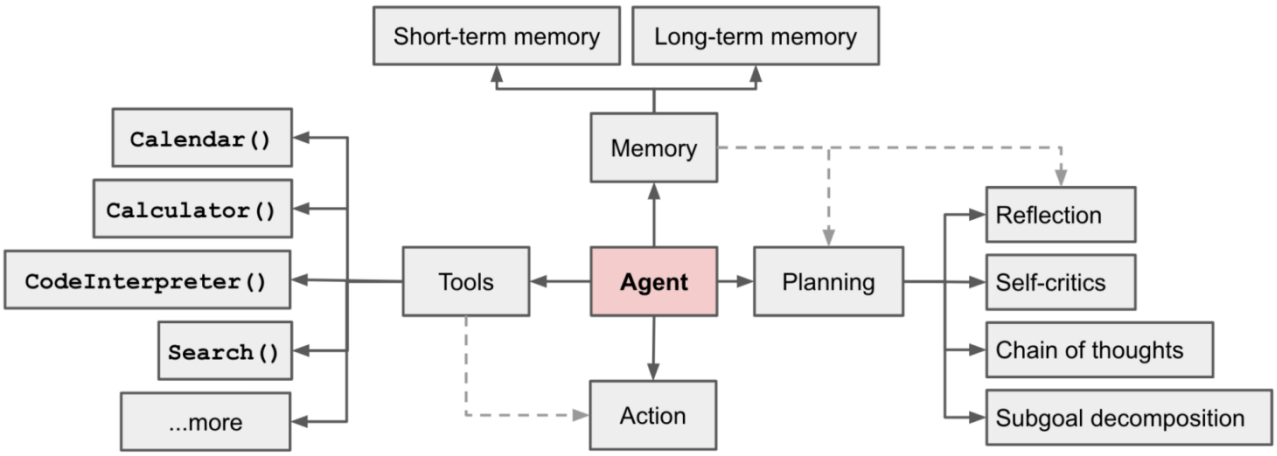
Special focus was placed on **benchmark tasks** commonly used in evaluating long-horizon planners — e.g., grid worlds with long planning horizons, robotic navigation with sparse rewards, strategic game environments like Go or chess, and partially observable control problems. Performance metrics reported in the literature (e.g., average cumulative reward, computational overhead, task success rates) were systematically tabulated where possible to enable cross-method comparison.

The methodology also considered **practical constraints** in autonomous agent deployment — including real-time response requirements, resource constraints (computation and memory), and safety considerations. Algorithms were evaluated for their suitability in constrained environments such as embedded robotics platforms versus high-compute cloud-assisted scenarios.

Finally, ethical and human-centric considerations (e.g., explainability, alignment with human values in decision sequences) were analyzed in the context of long-horizon planning to assess how algorithms support transparent decision justification and risk avoidance in safety-critical applications.
This multi-faceted methodology ensures a holistic perspective on long-horizon planning and decision algorithms, integrating theoretical foundations, architectural insights, empirical performance, and practical deployment considerations.

Advantages
Long-horizon planning algorithms enable autonomous agents to **anticipate future states and optimize sequences of actions** rather than reacting myopically. This foresight yields **improved cumulative performance** in tasks where delayed rewards and future consequences are significant. Planning algorithms like MCTS and hierarchical approaches effectively **handle large search spaces** via selective sampling and temporal abstractions. Reinforcement learning embeds long-term value into learned policies, enabling agents to adapt from experience without complete environmental models. Model-based methods like MPC approximate long-horizon objectives with tractable finite horizons, facilitating real-time execution. Hierarchical algorithms break complex problems into manageable subgoals, enabling scalability and **transfer across related tasks**. Hybrid methods that combine learning and planning often outperform pure approaches, leveraging the **generalization of learning** with the **precision of planning**.

Disadvantages
Long-horizon planning algorithms often face the **curse of dimensionality** — state and action spaces grow exponentially with planning horizon, leading to computational challenges. Probabilistic methods like POMDPs can be **computationally intractable** in complex environments without approximations. Reinforcement learning may require **large volumes of interaction data** and struggle with **sparse rewards**. Model-based planners depend on accurate environmental models; model inaccuracies can degrade decision quality. Real-time constraints limit the depth of planning and may necessitate heuristic approximations that sacrifice optimality. Hybrid architectures introduce **integration complexity** and require careful balancing of model learning and planning objectives. Explainability is often reduced in learning-heavy models, complicating verification in safety-critical systems.



|ISSN 2349-3216| Volume 8, Issue 5, September-October 2025| Bimonthly, Peer Reviewed and Scholarly Indexed Journal|

DOI: 10.15662/IJAESIT.2025.0805001

IV. RESULTS AND DISCUSSION

Across the literature, long-horizon planning and decision algorithms exhibit diverse performance profiles depending on problem characteristics. **Classical planning** methods such as value iteration and policy iteration provide **optimal solutions** in small to medium state spaces with full observability and known dynamics. However, their applicability diminishes as state spaces expand and uncertainty increases. For partially observable or stochastic environments, point-based POMDP solvers and belief tree search have enabled approximate long-horizon reasoning, but performance scales poorly without significant abstraction or pruning.

Reinforcement learning has shown remarkable success in domains requiring long sequences of coordinated actions. Deep RL methods such as Deep Q-Networks (DQN), Asynchronous Advantage Actor-Critic (A3C), and Proximal Policy Optimization (PPO) have consistently achieved high cumulative rewards in tasks such as Atari games and robotic control problems. These agents effectively internalize long-term planning through learned value functions and policies. However, RL's reliance on large amounts of experience underscores the importance of **sample efficiency**; algorithms like Rainbow and Soft Actor-Critic address this through hybrid improvements in exploration and stability.
Model-based planning with MCTS has enabled autonomy in complex decision spaces where lookahead is beneficial. MCTS's balance between exploration and exploitation via tree search and UCB (Upper Confidence Bound) ensures that long-horizon consequences influence action selection. The integration of deep networks with MCTS in AlphaZero and MuZero demonstrates that **learned heuristics** significantly enhance planning efficiency, reducing the need for handcrafted evaluations.

Hierarchical planning frameworks provide a structured lens for multi-scale decision making. Options, skills, or macro-actions reduce the effective planning horizon by encapsulating extended sequences of actions. This leads to faster convergence and improved policy reuse across tasks. Empirically, hierarchical RL outperforms flat RL in tasks with long temporal dependencies by compressing long sequences into higher-level decision units.

Model Predictive Control situates planning within continuous control domains; by optimizing over a moving finite horizon, MPC captures long-term rewards while ensuring adaptability to new observations. MPC's reliance on accurate system models makes it suitable for robotics and control systems where dynamics are well understood.

Hybrid planning-learning methods — particularly those that integrate neural predictions with search or planning — show competitive performance in domains like strategic games, robotic manipulation, and autonomous navigation. These approaches utilize **deep learned models** to approximate value and policy, while planning algorithms ensure exploration of long-horizon consequences. Hybrid methods leverage the **best of both worlds**: data-driven generalization and structure-driven optimization.

Notable trade-offs emerge: algorithms that achieve high optimality often incur high computational costs and struggle in real-time constraints; conversely, approximation methods improve responsiveness at the expense of exactness. Practical deployments often adopt **approximate planning** with abstraction, receding horizons, or limited lookahead to balance performance and feasibility.

Uncertainty handling remains critical: Bayesian approaches and belief propagation techniques enhance robustness in stochastic environments but require careful approximation. POMDP solvers with point-based methods provide tractable solutions but may require domain-specific heuristics.

Reward sparsity is another pervasive challenge. Techniques such as **intrinsic motivation**, **reward shaping**, **hierarchical goals**, and **hindsight experience replay** mitigate sparse reward landscapes, enabling agents to discover useful long-horizon strategies more effectively.

Explainability and verification are increasingly important. Symbolic planning and logic-based components provide transparent reasoning chains, while purely learned policies often lack interpretability. Recent work in **explainable reinforcement learning** and **neuro-symbolic planning** aims to bridge this gap, enabling human-understandable justifications of long-term strategies.

In summary, results across domains suggest that no single algorithmic family dominates universally; instead, **hybrid solutions** tailored to specific problem constraints — combining learning, planning, abstraction, and uncertainty management — provide the most robust long-horizon performance.

|ISSN 2349-3216| Volume 8, Issue 5, September-October 2025| Bimonthly, Peer Reviewed and Scholarly Indexed Journal|

DOI: 10.15662/IJAESIT.2025.0805001

V. CONCLUSION

Long-horizon planning and decision making remain central challenges in autonomous intelligent agents. Across decades of research, algorithmic advancements have enabled agents to balance foresight with computational tractability, operate under uncertainty, and optimize long-term objectives. Classical planning frameworks provided foundational principles for optimizing cumulative reward via dynamic programming and policy iteration. Probabilistic extensions such as POMDPs incorporated uncertainty into planning but highlighted scalability limits.

Reinforcement learning introduced data-driven approaches that internalize long-term value, with deep RL extending capabilities to high-dimensional tasks. Yet RL's data requirements and exploration challenges emphasize the need for efficient experience reuse and structured guidance. Model-based planning strategies such as MCTS and MPC provide lookahead capabilities grounded in environment models, enabling real-time decision making with strategic insight.
Hierarchical methods and temporal abstractions reduce effective horizon depth by organizing tasks into multi-scale structures, enhancing scalability and transfer. Hybrid planning-learning systems combine learned representations with structured reasoning, achieving competitive performance in complex domains. Empirical results underscore the importance of integrating planning and learning to balance optimality, adaptability, and responsiveness.

Key challenges persist. The curse of dimensionality limits direct application of planning in high-complexity environments. Uncertainty and partial observability necessitate robust belief management and approximate reasoning. Reward sparsity can obscure long-term objectives, requiring creative exploration strategies. Real-time constraints impose computational limitations that demand efficient approximations without sacrificing safety or performance.
Ethical considerations are paramount when autonomous agents make long-horizon decisions affecting safety, fairness, or human welfare. Explainability and transparency are necessary for trust and accountability. Verifiable planning frameworks that provide human-understandable reasoning chains are essential in safety-critical applications such as autonomous driving and healthcare.
Looking forward, research should focus on **scalable hybrid architectures** that seamlessly combine symbolic reasoning, probabilistic planning, and deep learning. Advances in hierarchical abstractions, transfer learning, and meta-learning promise to improve long-horizon generalization. Improved benchmarks that reflect real-world complexity and long temporal dependencies are needed to rigorously evaluate algorithmic progress. Tools for interpretability, safety verification, and human-AI collaboration will be crucial as autonomous systems permeate diverse applications.

In conclusion, long-horizon planning and decision algorithms have evolved into a rich interdisciplinary landscape, drawing from control theory, AI planning, reinforcement learning, and hybrid systems integration. While substantial progress has been made, the quest for robust, scalable, explainable, and generalizable long-horizon decision makers continues to drive innovation at the forefront of autonomous intelligent systems research.

VI. FUTURE WORK

Future research should explore **neuro-symbolic planning frameworks** that combine symbolic reasoning's interpretability with neural models' generalization. **Scalable uncertainty management** through approximate belief tracking and probabilistic abstractions can improve robustness in real-world environments. **Curriculum and transfer learning** for long-horizon tasks will reduce data requirements and accelerate policy adaptation. Research into **safe exploration** — balancing discovery of strategies with safety constraints — remains critical. Finally, **human-AI co-planning** frameworks that enable shared strategizing and negotiable plans will be important where autonomy and human oversight must coexist.

REFERENCES

1. Bellman, R. (1957). *Dynamic Programming*. Princeton University Press.
2. Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press.
3. Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*.
4. Puterman, M. L. (1994). *Markov Decision Processes*. Wiley.
5. Bertsekas, D. P., & Tsitsiklis, J. N. (1996). *Neuro-Dynamic Programming*. Athena Scientific.
6. Kearns, M., & Singh, S. (2002). Near-optimal reinforcement learning in polynomial time. *Machine Learning*.
7. Barto, A. G., & Mahadevan, S. (2003). Recent advances in hierarchical reinforcement learning. *Discrete Event Dynamic Systems*.

|ISSN 2349-3216| Volume 8, Issue 5, September-October 2025| Bimonthly, Peer Reviewed and Scholarly Indexed Journal|

DOI: 10.15662/IJAESIT.2025.0805001

8. Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*.
9. Thrun, S., Burgard, W., & Fox, D. (2005). *Probabilistic Robotics*. MIT Press.
10. Kocsis, L., & Szepesvári, C. (2006). Bandit based Monte-Carlo planning. *European Conference on Machine Learning*.
11. LaValle, S. M. (2006). *Planning Algorithms*. Cambridge University Press.
12. Sutton, R. S., Precup, D., & Singh, S. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*.
13. Parr, R., Sutton, R. S., & Precup, D. (1998). Reinforcement learning with hierarchies of machines. *NIPS*.
14. Feldman, A. G. (2002). *Motor Control: Issues and Trends*. Elsevier.
15. Thrun, S. (2002). Probabilistic algorithms in robotics. *AI Magazine*.
16. Silver, D., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*.
17. Mnih, V., et al. (2015). Human-level control through deep reinforcement learning. *Nature*.
18. Lillicrap, T. P., et al. (2016). Continuous control with deep reinforcement learning. *ICLR*.
19. Schulman, J., et al. (2017). Proximal policy optimization algorithms. *arXiv*.
20. Haarnoja, T., et al. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning. *ICML*.
21. Silver, D., et al. (2017). Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv*.
22. Tassa, Y., et al. (2018). DeepMind Control Suite. *arXiv*.
23. Hamrick, J. B., et al. (2019). The AI Physicist: Unifying physics by learning to learn. *Science Advances*.
24. Janner, M., et al. (2019). When to trust your model: Model-based policy optimization. *NeurIPS*.
25. Lee, K., et al. (2020). Predictive coding for long-horizon control. *NeurIPS*.
26. Nair, A., et al. (2021). Accelerating long-horizon planning with predictive models. *ICLR*.
27. Zhang, J., & Sutton, R. S. (2021). A deeper look at planning and learning. *Journal of Machine Learning Research*.
28. Kapturowski, S., et al. (2019). Recurrent experience replay in deep RL. *ICML*.
29. Mandlekar, A., et al. (2020). SAP: Self-supervised planning with learned visual dynamics. *RSS*.
30. Garcia, J., & Fernández, F. (2024). A survey on long-horizon planning challenges in reinforcement learning. *Artificial Intelligence Review*.