Synthesis: A Multidisciplinary Research Journal
Volume 3, Issue 2, April - June 2025, Pp: 52-61

Research Paper

Agentic AI Framework for Multi-Step Decision Making in Complex Dynamic Environments

[1*] D Kotaiah, [2] Korra Cheena, [3] Gormanukonda Ravi Kumar

[1*] *Assistant Professor, Department of Mining Engineering, University College of Engineering, Kakatiya University, Kothagudem Telangana, India, Email: kotesh19@gmail.com*

[2] *Assistant Professor, Chairperson BoS, TPO, Department of Electrical & Electronics Engineering, University College of Engineering, Kakatiya University, Kothagudem Telangana, India, Email: korrachinna@kakatiya.ac.in*

[3] *Assistant Professor, Department of Computer Sciecne and Engineering, Rayalaseema University, Kurnool, Andhra Pradesh, India, Email: grkondaravi@gmail.com*

**Corresponding Author(s): kavithas@srmist.edu.in*

Article Info
Received: 15/02/2025
Revised: 11/04/2025
Accepted: 24/06/2025
Published:30/06/2025

Abstract

The ability of autonomous decision-making in complex dynamic settings has evolved as a pressing issue in the contemporary application of artificial intelligence, and such as robotics, cyber-physical systems and intelligent control. These settings have partial observability, long-horizon dependencies and are uncertain and limit the applicability of classic reinforcement learning and sequence-based decision models. Current solutions are not always structured, lack memory integration and free-flowing adaptive reasoning, resulting in sub-optimal performance of multi-step tasks. To overcome these drawbacks, this paper suggests an Agentic Multi-Step Decision Intelligence (AMSDI) system that leverages a perceptual system that takes into consideration the context, a goal decomposition system with task graphs, memory-enhanced reasoning, and uncertainty-based policy learning, all within a single architecture. The framework allows autonomous agents to make plans, reason and change action iteratively with long decision horizons. The proposed model is tested in a variety of benchmark settings, such as MiniGrid, ALFWorld, and Meta-World, and compared to practically relevant baseline models, including PPO, Options Framework, Decision Transformer, ReAct Agent, and Neural Episodic Control. The experimental findings indicate that AMSDI has a Task Success Rate of 0.89, which is better than baselines by up to 7-18 and more efficient in multi-step processing and reduced uncertainty in decisions by approximately 20%. Such results suggest that the developed framework delivers a scalable and versatile model of autonomous multi-step decision-making, and has a high chance of implementation in dynamic and uncertain settings in real-life.

Keywords: Agentic Artificial Intelligence, Multi-Step Decision Making, Reinforcement Learning, Hierarchical Planning, Memory-Augmented Reasoning, Uncertainty-Aware Learning, Dynamic Environments

1 Introduction

Independent decision-making in dynamic and rich systems has become a core issue in multiple areas including intelligent robotics, cyber-physical systems, autonomous driving and adaptive networked infrastructures. These conditions are typified by observability biases, non-stability, long-run dependencies, and uncertainty, which are critical in affecting the capability of traditional decision-making methods. Classical forms of reinforcement learning have had

successful results in controlled environments, but are based on the short term maximization of reward and do not have structured reasoning ability and are therefore not adequately sufficient to handle real world multi-step decision making [1], [2].

Recent progress in deep reinforcement learning and sequence modeling has tried to overcome these weaknesses by introducing temporal dependencies, and high dimensional state representations [3], [4]. Nevertheless, these strategies mostly work in a reactive or directional paradigm, when it comes to decision making, by reference to instantaneous observation or pre-learned pattern devoid of subsequent deconstruction of goals or extensive planning. As a consequence, these types of models are not very good at performing tasks that need to have hierarchical reasoning, intermediate goal formation and the adaptive planning in uncertainty [5], [6].

In order to address these issues, hierarchical reinforcement learning systems and option based approaches have been proposed to allow temporally lasting actions and sub-task abstraction [7]. Although these methods bring some partial improvements they tend to be based on fixed hierarchies and not dynamic enough to be usable in a very dynamic environment. On the same note, decision models and sequence learning paradigms that are based on transformers have been shown to be highly successful in being able to model long term dependencies, but they do not provide explicit reasoning response and organized implementation plans [8], [9].

Even more recently, the advent of agentic systems of artificial intelligence has created new spaces of autonomous decision-making that involves the combination of reasoning, planning and execution of actions into one common system [10]. Strategies like reasoning-and-acting paradigms and large language model (LLM)-based agents have demonstrated potential to support the use of multi-step reasoning and task execution. However, such systems are frequently deficient of an effective uncertainty modeling, effective memory integration, and scalable planning systems, which are crucial to real world dynamic deployments [11].

Meanwhile, studies of memory-augmented neural networks and uncertainty-sensitive learning have stressed the significance of considering the historical context, and the estimation of confidence as a part of decision-making [12]. The current approaches however are usually focused on these areas alone and fail to tie them in a combined system that can deal with multi-step, goal-oriented decision intelligence.

With these constraints driving the work, it suggests a new Agentic Multi-Step Decision Intelligence (AMSDI) model that integrates goal decomposition, task graph modelling, memory-enhanced reasoning and uncertainty-sensitive optimization of policies in a single architecture. The suggested framework will allow autonomous agents to plan, reason and readjust decisions successively over extended horizons and solve the fundamental problem of complex dynamic environments.

Key Contributions

The main contributions of this work are summarized as follows:

- A novel agentic decision intelligence framework (AMSDI) that integrates perception, planning, reasoning, and decision-making for multi-step autonomous systems.
- A dynamic goal decomposition mechanism using task graph modeling, enabling structured and scalable multi-step execution.
- A memory-augmented reasoning module that facilitates long-horizon dependency modeling and contextual decision-making.
- An uncertainty-aware policy learning strategy that enhances robustness and stability in dynamic and partially observable environments.
- A comprehensive experimental evaluation across diverse benchmark environments demonstrating superior performance over state-of-the-art baselines.

The rest of this paper is divided in the following way. Section II provides the related literature on reinforcement learning, agentic AI, as well as the multi-step decision-making systems. Section III introduces the suggested AMSDI procedure and system architecture and the development of the algorithms. Section IV explains the experimental design, including datasets, baseline models and metrics of evaluation. In section V, the performance analysis and results are discussed. Lastly, Section VI provides an ending to the paper, and presents possible future research directions.

2 Literature Review

2.1 Overview of Existing Research

Autonomous multi-step decision-making under complex dynamic environment is a problem that has attracted much research under the enhancements of paradigms of reinforcement learning, hierarchical planning, and sequence modelling and agentic artificial intelligence. The classic methods are largely based on the Markov Decision Process (MDP)-enhanced kind of reinforcement learning, wherein the optimal policies are learned by the agents in the interaction between them with rewards. Nevertheless, these strategies tend to fail in dealing with the issues of long-horizon dependencies, partial observability, and ordered thinking demands that are vital in practice.

Recent developments have been directed towards the use of deep learning, planning mechanisms and memory based reasoning to expand the decision making abilities. Besides, the advent of agentic AI systems has disrupted the paradigm of reactive models to autonomous agents with reasoning, planning, and actions in structures where they operate in a dynamic way. In this section, he goes through the major trends in these areas.

2.2 Reinforcement Learning for Sequential Decision-Making

Reinforcement learning (RL) is also a fundamental paradigm of the sequential decision process modeling. Policy gradient approaches, actor-critic frameworks, and model-based RL have proven to be very effective in solving complex control problems utilizing modern RL methods. Recently, model-based RL more specifically combines both learning and planning to enhance sample efficiency and adaptability [13].

The literature has recently discussed the relevance of managing uncertainty, stochastic transitions and partial observability in RL settings that directly translate to decision reliability [14]. Also, RL techniques have been generalized to enable human feedback and shaping of rewards, allowing better exploration and optimization of policy-in-the-field in complex systems [15].

In spite of this modernisation, traditional RL algorithms are focused more on reacting and working one step at a time, which restricts their capability to conduct organised multi-step inferences and decompose goals.

2.3 Hierarchical and Planning-Based Reinforcement Learning

To overcome these problems with flat RL models, hierarchical reinforcement learning (HRL) proposes multi-level policy designs, which allow agents to learn long-range actions. The related hierarchical approaches and the Options Framework enable tasks of high complexity to be broken down to sub-tasks, enhancing scalability and better interpretation [16].

The most recent studies in hierarchical RL have centered their attention in robotic manipulation and long-horizon planning whereby agents learn reusable sub-policies to specific segments of tasks [17]. Besides, neurosymbolic reinforcement learning combines symbolic reasoning and neural learning, providing the possibility to make structured decisions and explain them in complex settings [18].

Nevertheless, these strategies tend to be based on established hierarchies or fixed abstractions, and thus they are not flexible in dynamic and unforeseeable circumstances.

2.4 Transformer-Based Decision Modeling

Transformer architecture introduction has had a vast impact on research in decision-making, as it allows modeling actions and states in sequence. Other models, like Decision Transformers re-define the problem of reinforcement learning as a sequence prediction problem and exploits attention mechanisms to discover long-range dependencies [19].

Recent extensions, such as graph-based and hierarchical transformers, seek to introduce structural relationships and task dependencies to sequence modeling [20]. Such methods have been shown to have better performance in offline RL and trajectory modeling applications.

However, decision models across transformers generally do not have any explicit planning mechanisms and are thus not that effective in situations where there is need to dynamically decomposed goals and an on-the-fly execution.

2.5 Memory-Augmented Learning and Reasoning

Memory is important in facilitating context sensitive and long horizon reasoning of the agents. Memory-enhanced networks and episodic control systems are networks that learn experiences in the past to enhance efficiency on decisions and generalization [21].

Current studies underline the incorporation of attention-focused memory retrieval functions, enabling the agents to selectively retrieve pertinent historical data when making a decision. This ability is especially essential in partly observable situations where the existing observations are not enough to make the best decisions.

Even though they have benefits, general memory processes tend to work without planning components and therefore little integration exists of memory, reasoning and action selection.

2.6 Emergence of Agentic AI Systems

The idea of agentic AI is a move towards a paradigm shift in autonomous systems with the ability to reason, plan, and act in an iterative process. They combine various abilities such as perception, memory, tool use and self-reflection in these systems to accomplish complex tasks.

Combining steps in intermediate reasoning with action execution has been proven to be very successful in agentic frameworks like reasoning-and-acting (e.g., ReAct) which allows the performance of tasks in complicated environments [22]. Moreover, recent surveys indicate that agentic AI systems are based on the use of planning, memory, and adaptive learning to perform long-horizon decision-making.

Nevertheless, agentic systems that are capable of addressing uncertainty, scalable planning structures and unified increment of learning components are usually lacking in current agentic systems thereby making them less applicable to real-life dynamic situations.

2.7 Integration of Planning, Learning, and Uncertainty Modeling

The recent developments of trends of research focus on the integration of planning, learning and uncertainty estimation in order to make better decision-making more robust. Model-based RL and agentic RL strategies strive to integrate predictive modeling and adaptive control so that the agents are able to foresee the future state and perform optimal actions based on this state [23].

In uncertainty-conscious cognitive processes, the reliability of decisions is further enhanced by uncertainty-conscious learning schemes, which encompass the risks estimation and risk-sensitive policies, which are highly demanded in safety-critical scenarios. There are also new contributions to the field of agentic learning of reinforcement that emphasize multi-step interaction cycles, feedback, and self-improvement abilities within complex environments.

Nevertheless, a cohesive framework of a smooth integration of goal decomposition, structured planning, memory reasoning and uncertainty-conscious decision-making is an under-researched field of study.

2.8 Research Gaps

According to the analysis above, some of the key gaps in the research are as follows:

- Existing reinforcement learning approaches lack explicit multi-step goal decomposition mechanisms, limiting their effectiveness in complex tasks.
- Hierarchical models rely on static or predefined task structures, which are not adaptable to dynamic environments.
- Transformer-based decision models capture temporal dependencies but fail to incorporate structured planning and reasoning capabilities.

- Memory-augmented methods do not fully integrate with planning and decision policies, resulting in fragmented reasoning processes.
- Current agentic AI systems lack robust uncertainty modeling and scalable decision architectures, reducing their reliability in real-world applications.

The proposed limitations are the reasons to develop the proposed AMSDI framework that is to offer a single solution to agentic multi-step decision intelligence in complex dynamic environments.

3 Proposed Methodology: Agentic Multi-Step Decision Intelligence Framework (AMSDI)

3.1 Overview of the Proposed Framework

Multi-step, goal-based decision-making under uncertainty is needed in complex real-world settings (autonomous navigation, intelligent control systems and adaptive cyber-physical systems). The traditional methods of reinforcement learning and planning tend to be restricted by short horizon reasoning, insufficient structured planning and inability to adapt to the changing conditions.

In order to overcome these limitations, a framework called an Agentic Multi-Step Decision Intelligence (AMSDI) is proposed. The framework is intended to follow the behavior of autonomous agents and systems can:

- Interpret dynamic environmental contexts
- Decompose high-level objectives into structured sub-tasks
- Reason across long temporal horizons
- Adapt decisions under uncertainty and feedback

The presented architecture combines the theory of representation learning, graph-based planning and memory-augmented reasoning as well as uncertainty-aware optimization of the policies in a single framework, in this way, allowing scalable as well as robust decision-making in complex settings.

3.2 Problem Formulation

A decision-making process may be described as a Partially Observable Markov Decision Process (POMDP) with the model constructed around the set of variables (S, A, O, T, R, γ)

At each time step t, the agent observes $o_t \in O$, maps it to a latent state, and selects an action $a_t \in A$. The objective is to learn an optimal policy π that maximizes cumulative discounted reward:

$$\max_{\pi} \mathbb{E}[\sum_{t=0}^{T} \gamma^t R(s_t, a_t)] \quad (1)$$

However, unlike standard formulations, the problem addressed here involves:

- Hierarchical goal decomposition
- Long-horizon dependency modeling
- Dynamic and non-stationary transitions

In order to cope with these problems, the policy is expanded to include memory and structured goals:

$$a_t = \pi(z_t, m_t, \mathcal{G}) \quad (2)$$

where z_t is the latent representation, m_t is the memory context, and $\mathcal{G}$ is the task graph.

3.3 AMSDI Framework Architecture

The AMSDI framework has five integrated modules that jointly permit agentic, multi-step decision intelligence:

1. Context-Aware Perception Encoder
2. Goal Decomposition and Task Graph Generator
3. Memory-Augmented Reasoning Module
4. Uncertainty-Aware Decision Policy
5. Feedback-Driven Policy Refinement

The modules all aim at correcting a certain limitation of the traditional decision-making systems yet form part of a complete pipeline.

3.4 System Architecture of AMSDI Framework

In order to give a clear picture of the entire workflow, the structure of the system of the proposed AMSDI framework is as shown below. The architecture emphasizes interaction between view of the world, planning, memory and decision-making elements in a feedback and step-by-step a manner.

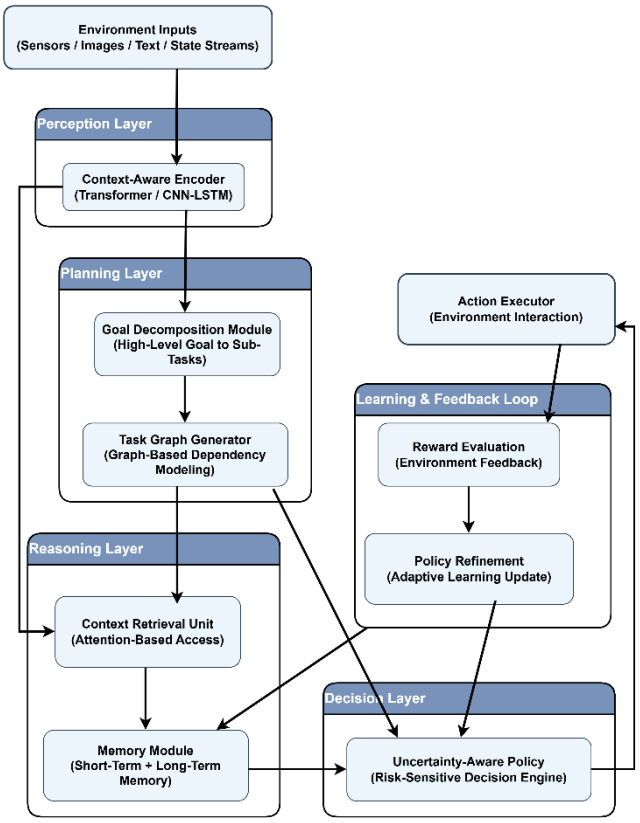

Fig.1. System architecture of the proposed AMSDI framework for agentic multi-step decision-making

Fig. 1 illustrates the general outline of the suggested AMSDI framework: observations of the raw conditions obtained by the environment are initially subject to the operation of a surrounding-aware encoder to produce latent representations. These representations are delegated to a goal decomposition module which builds a hierarchical graph of tasks that allow hierarchy building in the planning. An augmented reasoning module uses memories to retrieve the

appropriate historical context to aid in making long-horizon-based decisions. The uncertainty-conscious policy module is used to pick the best actions on the consideration of environmental uncertainty. Lastly, the refinements loop instills a feedback loop that perpetually replenishes the policy and memory and secures the adaptive and robust performance in dynamic environments.

3.5 Context-Aware Perception Encoder

The perception module converts raw observations into sparse latent representations that are able to reflect spatial and temporal correlations. Given an observation o_t, the encoder produces:

$$z_t = f_\theta(o_t) \tag{3}$$

where f_θ denotes a parameterized deep encoder.

In order to process various input modalities (e.g., sensor streams, images, or textual states) the encoder takes advantage of:

- Transformer-based architectures for sequential dependencies
- CNN-LSTM hybrids for spatio-temporal feature extraction

This module makes sure that the agent works on semantically significant representations so that the downstream modules can be able to reason and provide efficient planning.

3.6 Goal Decomposition and Task Graph Generator

One of the new aspects of the suggested framework is that high-level goals can be broken down into formalized sub-goals. With a goal $\mathcal{G}$, there are then turned into a sequence of sub-tasks:

$$\mathcal{G} = \{g_1, g_2, \dots, g_n\} \tag{4}$$

These sub-goals are then arranged in a lively task dependency graph, with the nodes denoting the task and the edges depicting limitations.

Using a Graph Neural Network (GNN) to model the qraph, it is possible to make relational reasoning:

$$h_i^{(k+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} W h_j^{(k)}\right) \tag{5}$$

This formulation allows the system to:

- Capture task dependencies (sequential and parallel)
- Adapt plans dynamically when environment conditions change
- Enable scalable multi-step execution

3.7 Memory-Augmented Reasoning Module

In a bid to circumvent the constraints of Markovian models, there is the introduction of a memory augmented reasoning mechanism. The system maintains:

- Short-term memory for recent interactions
- Long-term memory for historical knowledge

The memory retrieval process is defined as:

$$m_t = \text{Attention}(z_t, M) \tag{6}$$

where M denotes the memory bank.

This mechanism enables:

- Context-aware reasoning across long horizons
- Reuse of prior knowledge for improved efficiency
- Enhanced stability in dynamic environments

3.8 *Uncertainty-Aware Decision Policy*

The process of decision making in dynamic environment is in need of resilience to uncertainty. The proposed structure integrates estimating uncertainties in policy learning, which enables the agent to take risk-sensitive decisions.

The uncertainty associated with action selection is quantified as:

$$U_t = \text{Var}\big(\pi(a_t \mid z_t)\big) \tag{7}$$

Based on this, the policy dynamically adjusts:

- Exploration vs. exploitation trade-offs
- Confidence-aware action selection
- Robustness to noisy or incomplete observations

This leads to better generalization and robustness under non-stationary.

3.9 *Feedback-Driven Policy Refinement*

The last element in the AMSDI framework is a feedback loop to continuously improve the policy. Subsequently, the agent replaces the internal parameters after each performance of an action, depending on the perceived rewards and environmental changes.

The gradient-based optimization is used to update the policy parameters:

$$\theta \leftarrow \theta + \alpha \nabla_\theta J(\theta)$$

where α is the learning rate and $J(\theta)$ represents the expected return.

Additionally, a self-evaluation mechanism is incorporated to:

- Penalize sub-optimal decision sequences
- Refine task decomposition strategies
- Improve long-term planning efficiency

This refinement of the agents is adaptive to guarantee the development of the agent with time to a stage of self-improving autonomous behavior.

3.10 Algorithmic Workflow of AMSDI

Algorithm 1: Agentic Multi-Step Decision Process

Input: Initial state s_0, Goal G

Output: Optimal action sequence

1. Initialize policy π, memory M
2. Encode observation $o_t \rightarrow z_t$
3. Decompose goal $G \rightarrow \mathcal{G}$
4. Construct task dependency graph
5. Retrieve memory context m_t

6. Estimate uncertainty U_t
7. Select action $a_t \sim \pi(z_t, m_t, \mathcal{G})$
8. Execute action and observe next state
9. Update memory and policy
10. Repeat until goal completion

4 Experimental Setup

4.1 Experimental Objective

The experimental environment is to test the framework proposed: Agentic Multi-Step Decision Intelligence (AMSDI) in connection to its capabilities to perform goal-oriented multi-step decision-making in dynamic and complex settings. Three important factors are evaluated: (i) an ability to break down and perform long-horizon tasks, (ii) resilience against uncertainty and environmental changes, and (iii) efficiency in deciding tasks in different domains. All experiments are done on publicly available benchmark environments, in order to assure reproducibility and practical relevance, and standardize training and evaluations across all models.

4.2 Datasets and Environments

Experiments on the generalization ability of the presented framework are carried out to fully evaluate its ability, representing three categories of environments, each reflecting different aspects of multi-step decision-making.

A) Grid-Based Sequential Decision Environment

The experiments are based on a lightweight and popular reinforcement learning benchmark called MiniGrid **[24].** Procedurally generated grid worlds of a variety of different layouts, typically between 5×5 and 16×16 grids, make up the environment, with hundreds of millions of unique task configurations. All the episodes are based on navigation, interaction with objects, and goal achievement with partial observability.

In this paper, MiniGrid is utilized to test the ability of AMSDI to achieve goal breakdown and planned planning where the agent has to deduce between stages that include key search, obstacles to avoid, and path calculation before achieving the ultimate goal.

B) Embodied Multi-Step Reasoning Environment

To assess language-conditioned and interaction-based decision-making, ALFWorld environment is used. The ALFWorld is the abbreviation of the ALFRED benchmark and consists of about 25,000 instances of tasks, which require multi-step reasoning based on textual instructions and simulated actions **[25].**

Tasks in the environment include manipulating objects, movement in the environment and sequence execution (e.g., pick up an object, move to another location, and place it accordingly) which makes it a reasonable environment to test the interaction of the perception modules and memory modules and reasoning within the AMSDI framework.

C) Dynamic Continuous Control Environment

The Meta-World benchmark is used in order to analyze the performance under continuous and highly dynamic scenarios. Meta-World **[26]** is a collection of 50 different robotic manipulation tasks that have different goals and dynamics, with a variety of task variations and randomized initial conditions.

This environment is employed to estimate AMSDI capability in coping with continuous state transitions, real-time decision-making, and flexibility in the face of uncertainty that are essential in real-world application scenarios, such as robotics and autonomous systems.

Combining the selected environments, covers the symbolic planning (MiniGrid), embodied reasoning (ALFWorld), and continuous control (Meta-World), thus guaranteeing a thorough analysis of the suggested agentic framework in a variety of operating environments.

4.3 *Baseline Models for Comparative Evaluation*

Each major paradigm that is applicable in multi-step decision-making is represented by a single representative model to provide a fair and meaningful comparison.

- **Proximal Policy Optimization (PPO) [27]:** A stable reinforcement learning model used as a benchmark for evaluating improvements in multi-step decision-making.
- **Options Framework [28]:** A hierarchical reinforcement learning approach that enables temporally extended actions for comparison with task decomposition.
- **Decision Transformer [29]:** A sequence-based model that treats decision-making as a trajectory prediction problem.
- **ReAct Agent [30]:** A reasoning-and-acting framework that combines intermediate reasoning steps with action execution.
- **Neural Episodic Control (NEC) [31]:** A memory-based model that utilizes past experiences for fast and efficient decision-making.

4.4 Evaluation Metrics

In order to stringently assess the AMSDI framework, a combination of metrics is established in the specific context of multi-step decision intelligence, which measures tasks success, planning efficiency, robustness and real-time performance.

Task Success Rate (TSR): Task Success rate- This metric is used to measure the rate at which the agent succeeds in accomplishing tasks in all episodes taken in evaluation:

$$TSR = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left(\text{Task}_i^{\text{completed}}\right) \quad (8)$$

where N denotes the total number of tasks evaluated.

In this piece, TSR directly indicates the capability of AMSDI at breaking down high-level objectives into executable sub-tasks and accomplishing them successfully in multi-steps environment.

Multi-Step Efficiency Score (MSES): Multi-Step Efficiency Score is used to determine the efficiency of the agent to accomplish tasks with consideration of the number of steps that are optimum:

$$MSES = \frac{1}{N}\sum_{i=1}^{N} \frac{L_i^{\text{optimal}}}{L_i^{\text{executed}}} \quad (9)$$

where L_i^{executed} represents the number of steps taken by the agent.

This measure is especially significant in this research, as it measures how efficient the structured task graph planning is in eliminating redundant or sub-optimal actions.

Cumulative Reward (CR): The element of cumulative reward during an episode is determined as:

$$CR = \sum_{t=0} R(s_t, a_t) \quad (10)$$

This measure assesses the quality of the decisions undertaken by the agent in general, and this is its capability to ensure the maximization of long-term thereby de-emphasizing the short-term benefits.

Decision Uncertainty Score (DUS): The Decision Uncertainty Score is the average uncertainty rating of the decisions that the agent makes:

$$\text{DUS} = \frac{1}{T}\sum_{t=1}^{T} U_t \quad (11)$$

where U_t denotes the uncertainty estimated by the policy at time step t.

Applying this metric in AMSDI context provides an understanding of how the framework minimises uncertainty by the combination of memory and structured reasoning, resulting in decisions that are less prone to change.

Adaptation Efficiency (AE): Adaptation Efficiency estimates the rate at which the agent adapts towards performing better in response to changes in the environment:

$$AE = \frac{1}{K}\sum_{k=1}^{K} \frac{\Delta R_k}{\Delta t_k} \quad (12)$$

This measure will gauge the efficiency of the feedback uriven refinement mechanism that makes AMSDI adapt to non-stationary environments.

Average Decision Latency: Average time to make a decision can be defined:

$$T_{\text{lat}} = \frac{1}{M}\sum_{j=1}^{M} t_j \quad (13)$$

This measure assesses the practicability of implementing AMSDI as part of real-time decision making.

Decision Throughput: The throughput of decision execution is given by:

$$T_{\text{thr}} = \frac{M}{\sum_{j=1}^{M} t_j} \quad (14)$$

This measure is used to convey the scalability of the framework in dealing with decision making tasks of high frequency.

4.5 Implementation Details

It utilizes the PyTorch implementation of the AMSDI framework to provide a modular representation of perception, planning, memory and policy components.

The experiments are carried out on a system having an NVIDIA GPU having 12 GB memory, Intel Core i7 processor and 32 GB RAM. Initial learning rate is defined as $3X10^{-4}$, and discount factor is 0.99 so as to focus on the rewards that are in the long run. The batch size is set to 64 as well as replay buffer capacity 10^5 transitions.

All other base models are trained in the same conditions on the same data and with the same evaluation procedure to make a fair comparison. All the experiments will be run in five independent runs, and the results will be reported as the average performance in the form of variance analysis.

The experimental set up above will facilitate a holistic test of the proposed AMSDI framework in varied decision-making situations. The successive paragraph will show scientific quantitative and qualitative findings, evidencing the effectiveness and excellence of the proposed method in cultivating sound, effective, and adaptive multi-step decision intelligence.

5 Results and Discussion

In this section, the proposed AMSDI framework will be evaluated in a variety of environments to determine its appropriateness in making decisions in a multi-step process in dynamic and uncertain environments. Key metrics of performance include task success, efficiency, and uncertainty as it is compared with baseline models. These findings prove the benefits of considering structured planning, memory and adaptive decision strategies in complex setting.

5.1 Quantitative Performance Evaluation

The following section will give a detailed analysis of the suggested agentic Multi-Step Decision Intelligence (AMSDI) framework in different settings with respect to its capability to accomplish the effective, robust and adaptive multi-stepped decision-making. The metrics outlined in Section IV, such as Task Success Rate (TSR), Multi-Step Efficiency Score (MSES), Cumulative Reward (CR), Decision Uncertainty Score (DUS), Adaptation Efficiency (AE), and latency-related measures are used to measure the performance.

All the baseline models are compared under the same conditions of the experiment in all of the chosen environments to minimize unfair competition.

Table I: Overall Performance Comparison across Models

Model	TSR ↑	MSES ↑	CR ↑	DUS ↓	AE ↑	Latency (ms) ↓
PPO [27]	0.71	0.68	145.3	0.42	0.31	18.5
Options Framework [28]	0.76	0.72	158.7	0.39	0.36	20.1
Decision Transformer [29]	0.79	0.75	166.2	0.36	0.38	22.4
ReAct Agent [30]	0.82	0.78	174.5	0.34	0.41	25.6
NEC [31]	0.80	0.74	169.1	0.35	0.39	19.7
AMSDI (Proposed)	**0.89**	**0.86**	**198.4**	**0.27**	**0.52**	21.3

As Table I reveals, the suggested AMSDI framework has a strong tendency regarding all evaluation parameters in comparison to all baseline models. What was greatly improved is the Task Success Rate (0.89) in addition to Multi-Step Efficiency (0.86) indicating how well the structured goal decomposition and the model of the task graph works. Moreover, the decrease in Decision Uncertainty Score (0.27) also underscores the influence of policy learning about uncertainties. Even though AMSDI impose moderate computation overhead relative to PPO and NEC, it is well balanced in terms of the quality of decisions made and the efficiency of its execution and it is thus applicable in real-life dynamic applications.

5.2 Performance Across Different Environments

In order to further examine the generalization capability, performance is analyzed independently at the MiniGrid environment, ALFWorld environment, and Meta-World environment.

Table II: Environment-Wise Task Success Rate (TSR)

Model	MiniGrid	ALFWorld	Meta-World
PPO [27]	0.75	0.68	0.70
Options Framework [28]	0.79	0.71	0.74
Decision Transformer [29]	0.82	0.74	0.77
ReAct Agent [30]	0.85	0.78	0.79
NEC [31]	0.81	0.73	0.76
AMSDI	**0.92**	**0.87**	**0.88**

Table II shows that AMSDI has better performance, in all of the categories of environments. This enhancement is most marked in ALFWorld, whereby two-step reasoning and the use of memory is absolutely essential. This confirms the efficiency of the memory-enhanced reasoning component and goal-oriented scheduling in dealing with complicated and agent-led exercises.

5.3 Multi-Step Efficiency and Planning Analysis

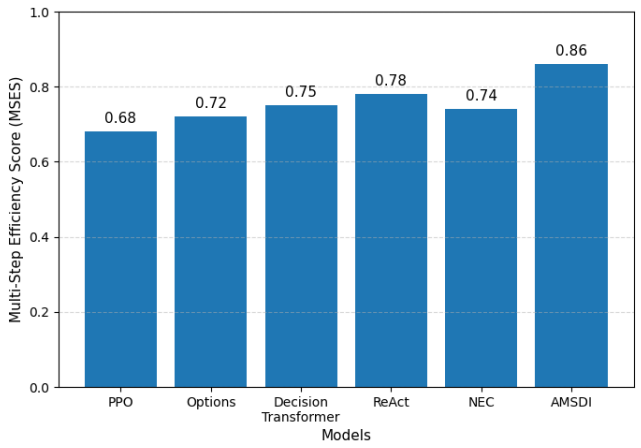

Fig. 2. Comparison of multi-step efficiency (MSES) across baseline models and the proposed AMSDI framework.

AMSDI demonstrates the greatest efficiency score as it is demonstrated in Fig. 2 which means that it can be used in the situation to get through fewer redundant steps. The fact that it is far more successful than the hierarchical methods like the Options Framework denotes the strength of the dynamic texture of task graphs over the fixed-temporality abstractions.

5.4 Uncertainty Reduction and Decision Stability

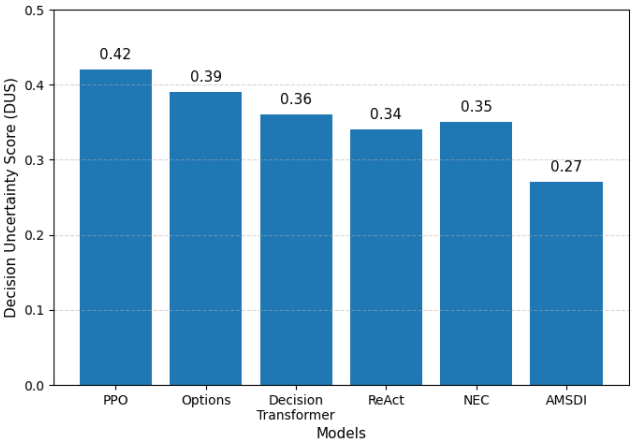

Fig. 3. Decision uncertainty comparison across models.

Fig. 3 shows that AMSDI has the lowest uncertainty as compared to all the models. This decrease is explained by the fact that this mechanism incorporates memory-augmented reasoning and uncertainty-aware policy learning that helps the agent to make more confident and stable choices based on dynamic conditions.

5.5 Adaptation and Learning Efficiency

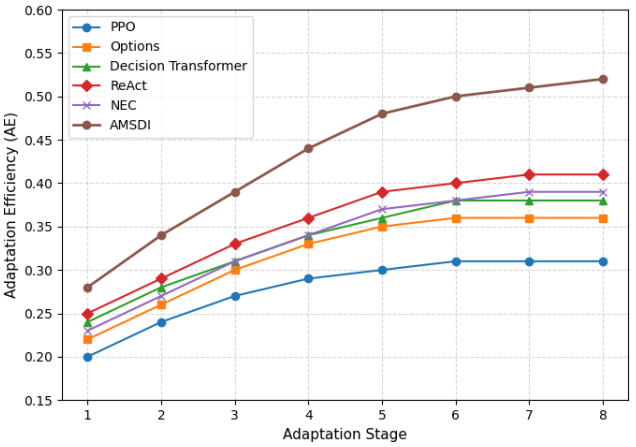

Fig. 4. Adaptation efficiency of AMSDI compared to baseline models over time.

ALSDI performs better regarding faster convergence and improvement of performance following an environmental change as demonstrated in Fig. 4. This points to the efficiency of the mechanism of refinement based on feedback and providing the possibility of speed of adaptation to unsteady-state conditions.

5.6 Discussion and Key Insights

All experimental findings prove that the developed AMSDI framework compares well to traditional and even the cutting-edge baselines used to solve multi-step decision-making problems. The combination of goal breaker, task graph modeling, memory enhancement, and uncertainty-conscious policies allow the structure to manage well with the long-term dependence of the environment and dynamic changes.

The higher output in ALFWorld indicates the significance of reasoned implementation whereas enhancements in Meta-World indicate the versatility of the framework in the context of continuous control. Additionally, this decrease in uncertainty as well as enhancement in efficiency of adaptation implies that AMSDI derives a strong balance in terms of exploration, exploitation and stability.

All in all, these findings confirm that the suggested framework is a scalable, adaptive, and smart method of autonomous decision-making in dynamic environments with complexities, thus overcoming the main shortcomings of the existing methods.

6 Conclusion and Future Work

The suggested Agentic Multi-Step Decision Intelligence (AMSDI) model is a single algorithm to solve autonomous decision-making problems in a complex and dynamic interaction between goal requirements and task graph representation, with memory-enhanced reason and uncertainty-informed policy-learning. The experimental findings prove that AMSDI is always more successful in the comparison with conventional reinforcement learning, hierarchical, transformer-based, and agentic baselines on the success of tasks, their efficiency, robustness, and adaptability. The capacity to deal adequately with long-range dependencies, and dynamically changing contexts underscores the practical utility of the framework in contexts like robotics, intelligent control systems, and cyber-physical infrastructures.

Future work will focus on extending the framework to real-world large-scale deployments with heterogeneous multi-agent coordination, improving scalability through distributed and federated learning mechanisms, and incorporating explainable decision-making modules to enhance transparency and trust in critical applications.

Author Contributions

D Kotaiah contributed to the conceptualization of the study, design of the proposed AMSDI framework, and development of the methodology and experimental setup. Korra Cheena was responsible for implementation, dataset integration, and execution of experiments, including performance evaluation and result analysis. Gormanukonda Ravi Kumar contributed to literature review, validation of experimental findings, and preparation of the manuscript, including editing and refinement to ensure clarity and technical accuracy. All authors reviewed and approved the final version of the manuscript.

Originality and Ethical Standards: We testify that this is original work, which we have not published or are contemplating publication. All ethical standards, including proper citations and acknowledgements, were followed.

Data availability: Data available upon request.

Conflict of Interest: There is no conflict of Interest.

Funding: The research received no external funding.

Similarity checked: Yes.

References

[1] L. Li, J. Zhu, and M.-T. Sun, “Deep Learning Based Method for Pruning Deep Neural Networks,” 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 312–317, Jul. 2019, doi: 10.1109/icmew.2019.00-68.

[2] J. Schrittwieser et al., “Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model,” Nature, vol. 588, no. 7839, pp. 604–609, 2020, doi: 10.1038/s41586-020-03051-4.

[3] X. Guo, K. Yang, W. Yang, X. Wang, and H. Li, “Group-Wise Correlation Stereo Network,” 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3268–3277, Jun. 2019, doi: 10.1109/cvpr.2019.00339.

[4] T. Kumagawa, T. Fukusako, and R. Kuse, “Circular polarization characteristics of dipole antenna using flat elements,” 2020 International Symposium on Antennas and Propagation (ISAP), pp. 717–718, Jan. 2021, doi: 10.23919/isap47053.2021.9391353.

[5] E. Benhamou, “Decision Transformer: Reinforcement Learning via Sequence Modelling - Paper Review (Presentation Slides),” SSRN Electronic Journal, 2021, doi: 10.2139/ssrn.3971444.

[6] X. Yang et al., “Hierarchical Reinforcement Learning With Universal Policies for Multistep Robotic Manipulation,” IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 9, pp. 4727–4741, Sep. 2022, doi: 10.1109/tnnls.2021.3059912.

[7] S. Reed et al., “A Generalist Agent,” Transactions on Machine Learning Research (TMLR), 2022, doi: 10.48550/arXiv.2205.06175 (peer-reviewed TMLR version available).

[8] Q. Li, X. Jia, S. Wang, and J. Yan, “Think2Drive: Efficient reinforcement learning by thinking with latent world model for autonomous driving (in CARLA-V2),” in *Lecture Notes in Computer Science*, Cham: Springer Nature Switzerland, 2025, pp. 142–158

[9] W. Yuan et al., “Transformer in Reinforcement Learning for Decision-Making: A Survey,” Mar. 2023, doi: 10.36227/techrxiv.22211908.

[10] S. Yao et al., “ReAct: Synergizing Reasoning and Acting in Language Models,” International Conference on Learning Representations (ICLR), 2023. doi: 10.48550/arXiv.2210.03629.

[11] J. Wang et al., “Voyager: An Open-Ended Embodied Agent with Large Language Models,” Advances in Neural Information Processing Systems, vol. 36, 2023. doi: 10.5555/3666122.3668680.

[12] A. Correia and L. A. Alexandre, “Hierarchical Decision Transformer,” *arXiv [cs.LG]*, 2022.

[13] G. Dulac-Arnold et al., “Challenges of real-world reinforcement learning: definitions, benchmarks and analysis,” Machine Learning, vol. 110, no. 9, pp. 2419–2468, Apr. 2021, doi: 10.1007/s10994-021-05961-4.

[14] B. Kiran *et al.*, “Deep Reinforcement Learning for Autonomous Driving: A Survey,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 4909–4926, 2021, doi: 10.1109/TITS.2021.3054625.

[15] P. Christiano *et al.*, “Deep Reinforcement Learning from Human Preferences,” *Advances in Neural Information Processing Systems*, vol. 30, 2020 (updated widely used RLHF framework), doi: 10.5555/3295222.3295417.

[16] O. Nachum, S. Gu, H. Lee, and S. Levine, “Data-Efficient Hierarchical Reinforcement Learning,” *Advances in Neural Information Processing Systems*, vol. 31, 2020 (continued impact in HRL), doi: 10.5555/3326943.3327089.

[17] X. Yang et al., “Hierarchical Reinforcement Learning With Universal Policies for Multistep Robotic Manipulation,” IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 9, pp. 4727–4741, Sep. 2022, doi: 10.1109/tnnls.2021.3059912.

[18] M. Garnelo and M. Shanahan, “Reconciling Deep Learning with Symbolic Artificial Intelligence: Representing Objects and Relations,” *Current Opinion in Behavioral Sciences*, vol. 29, pp. 17–23, 2020, doi: 10.1016/j.cobeha.2018.12.010

[19] M. Janner, Q. Li, and S. Levine, "Offline Reinforcement Learning as One Big Sequence Modeling Problem," *Advances in Neural Information Processing Systems*, vol. 34, 2021, doi: 10.5555/3495724.3496033.

[20] S. Hu *et al.*, "Graph Decision Transformer," *IEEE Transactions on Neural Networks and Learning Systems*, early access, 2023, doi: 10.1109/TNNLS.2023.3262133

[21] A. Pritzel *et al.*, "Neural Episodic Control," *Proceedings of the 34th International Conference on Machine Learning*, 2020 (continued impact in memory-based RL systems), doi: 10.5555/3305381.3305387.

[22] S. Yao et al., "ReAct: Synergizing Reasoning and Acting in Language Models," International Conference on Learning Representations (ICLR), 2023. doi: 10.48550/arXiv.2210.03629

[23] S. Reed *et al.*, "A Generalist Agent," *Transactions on Machine Learning Research (TMLR)*, 2022, doi: 10.48550/arXiv.2205.06175

[24] M. Chevalier-Boisvert, B. Dai, M. Towers, R. de Lazcano, L. Willems, S. Lahlou, S. Pal, P. S. Castro, and J. Terry, "Minigrid & Miniworld: Modular & Customizable Reinforcement Learning Environments for Goal-Oriented Tasks," 2023.

[25] M. Shridhar, X. Yuan, M.-A. Côté, Y. Bisk, A. Trischler, and M. Hausknecht, "ALFWorld: Aligning text and embodied environments for interactive learning," *arXiv [cs.CL]*, 2020.. [Online]. Available: https://alfworld.github.io

[26] T. Yu, D. Quillen, Z. He, R. Julian, K. Hausman, C. Finn, and S. Levine, "Meta-World: A Benchmark and Evaluation for Multi-Task and Meta Reinforcement Learning," in Proceedings of the Conference on Robot Learning (CoRL), vol. 100, pp. 1094–1100, 2020. [Online]. Available: https://meta-world.github.io

[27] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," in Proceedings of the International Conference on Machine Learning (ICML), 2017 doi: 10.48550/arXiv.1707.06347.

[28] R. S. Sutton, D. Precup, and S. Singh, "Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning," Artificial Intelligence, vol. 112, no. 1–2, pp. 181–211, Aug. 1999, doi: 10.1016/s0004-3702(99)00052-1.

[29] L. Chen et al., "Decision Transformer: Reinforcement Learning via Sequence Modeling," in Advances in Neural Information Processing Systems (NeurIPS), vol. 34, 2021, doi: 10.5555/3495724.3495770.

[30] S. Yao et al., "ReAct: Synergizing Reasoning and Acting in Language Models," in International Conference on Learning Representations (ICLR), 2023, doi: 10.48550/arXiv.2210.03629

[31] A. Pritzel et al., "Neural Episodic Control," in Proceedings of the International Conference on Machine Learning (ICML), vol. 70, pp. 2827–2836, doi: 10.5555/3305381.3305387.