ISSN: 2769-545X

Extend Version

Advances in Machine Learning & Artificial Intelligence

PlanningEFEMix: Hybrid Active Inference for Sequential Decision-Making under Uncertainty

Bhagyeshkumar Chokhawala* and **Dr. Atif Farid Mohammad**

Capitol Technology University, Laurel, Maryland, USA

***Corresponding Author**
Bhagyeshkumar Chokhawala, Capitol Technology University, Laurel, Maryland, USA.

Submitted: 2026, May 01; **Accepted:** 2026, May 27; **Published:** 2026, Jun 03

Citation: Chokhawala, B., Mohammad, A. F. (2026). PlanningEFEMix: Hybrid Active Inference for Sequential Decision-Making under Uncertainty. *Adv Mach Lear Art Inte, 7*(3), 01-17.

Abstract

Sequential decision-making under uncertainty remains a central challenge in artificial intelligence and machine learning, especially in environments where agents must act under partial observability, noisy feedback, sparse preference signals, and shifting context. Reinforcement learning, probabilistic planning, and representation learning each provide useful mechanisms for action selection, but each approach also has limitations when deployed as a single decision paradigm. Model-free reinforcement learning can be data intensive and unstable under noise; POMDP-based planning offers principled belief management but depends on accurate transition and observation models; contrastive representation learning improves discrimination but does not by itself define a policy objective; and deterministic Active Inference provides a coherent mechanism for uncertainty reduction and goal satisfaction but is often implemented with a fixed generative model. This paper develops PlanningEFEMix, a hybrid Active Inference planning framework that integrates heterogeneous decision agents using Expected Free Energy (EFE) as a shared meta-level objective. The proposed framework evaluates candidate actions through forward simulation across multiple agents, aggregates agent-specific EFE estimates, and augments action selection with an adaptive state-action bias memory that incorporates experiential feedback from prior outcomes. The algorithm is designed to preserve the interpretability of Active Inference while improving robustness through agent diversity and adaptive policy modulation. A noisy preference inference benchmark is used to frame the experimental evaluation, comparing PlanningEFEMix with deterministic Active Inference, POMDP belief agents, contrastive learning agents, and model-free reinforcement learning baselines across multiple observation-noise regimes and ablation settings. The manuscript contributes a formal problem definition, an algorithmic specification, an architectural framework, an evaluation protocol, and a discussion of deployment implications for decision-making systems that require explainability, robustness, and adaptive behavior under uncertainty.

Keywords: Active Inference, Expected Free Energy, Hybrid Decision-Making, Reinforcement Learning, POMDP, Meta-Inference, Sequential Planning, Uncertainty Handling, Explainable AI, Preference Inference

1. Introduction

Sequential decision-making under uncertainty is a foundational problem in artificial intelligence, machine learning, cognitive modeling, and adaptive control [1-3]. In many operational settings, an intelligent system must select actions before the true state of the world is fully known. The system may receive incomplete observations, delayed feedback, noisy labels, sparse preference signals, or contradictory evidence from multiple sources. These challenges arise in robotics, recommendation systems, autonomous control, enterprise decision support, human-AI interaction, and other domains where choices unfold over time and each action changes the information available to the agent [4-6].

Traditional machine learning approaches to this problem often fall into two broad categories. The first category includes reinforcement learning (RL) methods that optimize behavior through interaction

and reward feedback [2,7]. RL has produced significant advances in sequential decision-making, but model-free RL often requires many episodes of exploration, can be sensitive to reward misspecification, and may become unstable when observations are noisy or the reward signal is delayed [2,8]. The second category includes model-based planning methods, including Partially Observable Markov Decision Processes (POMDPs), which explicitly represent uncertainty over hidden states [1]. POMDPs provide a principled formulation for planning under partial observability, but they depend on well-specified transition and observation models and can become computationally demanding as the state and action spaces grow [1].

Active Inference offers an alternative formulation that integrates perception, learning, and action selection within a Bayesian framework derived from the Free Energy Principle [3,9]. Under Active Inference, agents select actions that minimize Expected Free Energy, a quantity that captures both epistemic value and pragmatic value [10,11]. Epistemic value encourages uncertainty reduction, while pragmatic value encourages movement toward preferred outcomes. This makes Active Inference attractive for systems that must balance exploration and exploitation without relying on separate heuristic exploration bonuses [12,13]. The same objective can explain why an agent seeks informative observations and why it pursues states aligned with preferences.

Despite its conceptual strengths, many practical Active Inference implementations remain limited by their reliance on a single generative model, fixed planning assumptions, or one decision strategy [13-15]. In real-world environments, no single model is likely to dominate across all contexts. A model-free learner may adapt well after sufficient experience but perform poorly early in deployment [2]. A POMDP agent may reason carefully about hidden states but fail when model assumptions are inaccurate [1]. A contrastive learner may produce robust representations but lacks a full sequential planning mechanism [16]. A deterministic Active Inference agent may be explainable and goal-directed, but may not capture all long-term adaptation patterns emerging from experience [10,17].

This paper proposes **PlanningEFEMix**, a hybrid decision-making framework that treats multiple decision agents as complementary internal hypotheses. Instead of selecting a single paradigm in advance, PlanningEFEMix uses Expected Free Energy as a common meta-level objective to compare and integrate candidate actions generated by heterogeneous agents. The approach preserves the Active Inference commitment to uncertainty-aware planning while introducing practical mechanisms for robustness, adaptation, and explainability. A state-dependent bias memory stores historical action tendencies in context and modulates future action selection, enabling the system to learn from experience without reducing the framework to conventional reward maximization alone.

The key idea is that hybrid decision-making can improve stability under uncertainty when the integration layer is principled rather than heuristic. This design is consistent with prior work showing that adaptive behavior may depend on multiple interacting control mechanisms rather than a single universal policy [7,18,19]. PlanningEFEMix therefore uses EFE as the shared language across deterministic Active Inference, POMDP belief updating, contrastive representation learning, and model-free reinforcement learning. Each agent contributes a different inductive bias, but the final policy is selected through an interpretable aggregation process that can be decomposed into epistemic, pragmatic, risk, ambiguity, and bias-driven components [10,11].

1.1. Research Problem

The research problem addressed in this manuscript is how to design a sequential decision-making agent that remains robust under partial observability and observation noise while retaining interpretability in action selection [1,10,13]. Robustness requires that the agent degrade gracefully as noise increases, avoid premature convergence to suboptimal actions, and recover from misleading observations. Interpretability requires that decisions can be explained in terms of uncertainty reduction, preference alignment, risk management, ambiguity handling, and learned contextual bias.

Single-paradigm models struggle to satisfy these requirements simultaneously. A purely model-free policy may be adaptive but hard to explain [2]. A purely model-based policy may be explainable but brittle when its model is wrong [1]. A fixed Active Inference agent may be principled but limited by the coverage of its generative assumptions [14,15]. PlanningEFEMix addresses this gap by using meta-level planning across multiple internal agents while maintaining EFE as the unifying decision criterion.

1.2. Contributions

This manuscript makes five contributions to the literature on machine learning decision-making and Active Inference:

- It formalizes PlanningEFEMix as a hybrid Active Inference framework for sequential decision-making under uncertainty.
- It defines a meta-level EFE aggregation method that enables heterogeneous agents to contribute to a common action-selection process.
- It introduces a state-dependent bias memory that captures context-specific experiential tendencies while preserving an EFE-based policy objective.
- It provides an algorithmic specification and architectural design that can be implemented in simulated preference inference, recommendation, or adaptive control settings.
- It establishes an evaluation protocol for noise sensitivity, stability, ablation analysis, and explainability of hybrid decision-making agents.

1.3. Relationship to Prior Conference Version

This journal manuscript is an extended version of the work previously presented in the RAIS Conference Proceedings as PlanningEFEMix: Hybrid Active Inference for Sequential Decision-Making under Uncertainty [20]. The earlier conference version introduced the core PlanningEFEMix idea, including the integration of deterministic Active Inference, POMDP-based belief

updating, contrastive learning, and model-free reinforcement learning under an EFE-oriented planning loop.

Compared with the conference version, the present journal manuscript adds and expands the following contributions:

- A substantially expanded background and related-work discussion with explicit in-text citations across Active Inference, POMDPs, reinforcement learning, contrastive learning, hybrid control, and explainable AI.
- A more complete formal problem definition, notation table, EFE component interpretation, and agent-weighting discussion.
- A detailed description of state-action bias memory, meta-level EFE aggregation, softmax action selection, and computational complexity considerations.
- A fuller evaluation protocol covering noise regimes, metrics, ablation conditions, statistical reporting expectations, and reproducibility guidance.
- A journal-style Results and Analysis section with simulated reporting tables and figures to show the intended empirical structure, along with expanded discussion, limitations, future work, and deployment implications.

2. Background and Related Work

2.1. Active Inference and the Free Energy Principle

Active Inference is grounded in the Free Energy Principle, which proposes that adaptive systems act to minimize a bound-on surprise by maintaining coherent beliefs about hidden states and preferred outcomes [3,9]. In this framework, perception updates beliefs about latent states, while action changes the world or the agent's sensory inputs in ways that are expected to reduce uncertainty and satisfy preferences [12,15]. Rather than separating inference from control, Active Inference treats action as a form of inference over policies [13].

In discrete state-space formulations, an agent maintains approximate posterior beliefs over hidden states and evaluates candidate policies according to Expected Free Energy [9,15]. EFE is often decomposed into epistemic and pragmatic components [11,13]. The epistemic component captures the expected information gain or uncertainty reduction associated with a policy. The pragmatic component captures the expected divergence between predicted outcomes and preferred outcomes. This decomposition is central to the explainability of Active Inference, because action selection can be interpreted as a trade-off between learning about the environment and achieving preferred consequences.

$$G(\pi) = E\_q[ln\ q(s_\tau \mid \pi) - ln\ p(o\_\tau, s\_\tau \mid C, \pi)]$$

Equation (1) expresses the expected free energy of a policy as an expectation over predicted states and outcomes, consistent with discrete-state Active Inference formulations [9,15]. Alternative decompositions can separate epistemic value, pragmatic value, ambiguity, and risk terms [11,13]. In PlanningEFEMix, this decomposability is extended to a hybrid-agent setting, allowing each sub-agent to expose its contribution to the final action decision.

2.2. POMDPs and Belief-Space Planning

Partially Observable Markov Decision Processes provide a classical framework for planning when agents cannot directly observe the full environment state [1]. A POMDP defines a state space, action space, observation space, transition model, observation model, and reward or preference function. Because the true state is hidden, the agent maintains a belief distribution that is updated as observations arrive. Planning then occurs in belief space rather than directly in state space [1].

POMDPs are well-suited for domains in which uncertainty is explicit, and observations are noisy [1]. However, practical POMDP planning becomes challenging when transition and observation probabilities are unknown, high-dimensional, non-stationary, or expensive to estimate. PlanningEFEMix uses POMDP-style belief updating as one component within a broader hybrid system rather than treating it as the only decision mechanism.

2.3. Reinforcement Learning and Experience-Driven Adaptation

Reinforcement learning defines decision-making as the optimization of cumulative reward through interaction with an environment [2]. Model-free methods learn value functions or policies directly from experience and can adapt without an explicit model of the environment [7]. This makes RL attractive for complex settings where model specification is difficult [8]. However, RL can require substantial data, may be sensitive to noisy or sparse rewards, and can produce policies whose internal rationale is difficult to communicate to human stakeholders.

PlanningEFEMix does not reject reinforcement learning. Instead, it incorporates a model-free agent as one contributor to the hybrid decision process. The RL component provides long-run value sensitivity, while the EFE-based aggregation layer retains a transparent decision objective. The state-action bias memory also borrows from value-updating intuition while using it as a contextual modulation term rather than as the sole basis for action selection [2,21].

2.4. Contrastive Learning for Representation Discrimination

Contrastive learning has become an influential approach for learning representations by pulling semantically similar instances closer and pushing dissimilar instances apart in an embedding space [16]. Although contrastive learning is often applied to computer vision and representation learning tasks, its logic is also useful in sequential decision environments where the agent must distinguish between subtly different states, user preferences, or contextual patterns under noise.

In PlanningEFEMix, a contrastive representation agent contributes discrimination capability. It does not replace planning; instead, it improves the representational substrate for evaluating candidate actions, following the representational separation principle emphasized in contrastive learning [16]. This is particularly useful

in preference inference tasks where different user states may produce similar observations or where noise can obscure the true preference signal.

2.5. Hybrid Decision Systems

Hybrid decision systems are motivated by the observation that intelligent behavior often depends on multiple interacting control mechanisms [7,18]. Cognitive and computational theories distinguish between habitual, goal-directed, model-based, and model-free control [19,21]. Rather than assuming a single mechanism is universally optimal, hybrid approaches coordinate specialized mechanisms based on context, uncertainty, or expected benefit.

The challenge for hybrid decision-making is integration. If agent outputs are combined through an ad hoc voting rule, the resulting policy may be difficult to interpret or tune. PlanningEFEMix addresses this by using Expected Free Energy as the integration objective, drawing on Active Inference accounts in which policy selection is evaluated through expected epistemic and pragmatic consequences [9,15]. Each sub-agent may estimate action quality differently, but the meta-layer then transforms or evaluates those estimates relative to a common uncertainty-aware objective.

Table 1 summarizes the positioning of PlanningEFEMix relative to the main single-paradigm approaches discussed above.

Approach	Strength	Limitation when used alone	Role in PlanningEFEMix
Deterministic Active Inference	Transparent planning using EFE; balances exploration and goal satisfaction.	Often depends on fixed generative assumptions.	Provides principled EFE decomposition and explainable action evaluation.
POMDP belief agent	Explicit treatment of hidden states and noisy observations.	Requires accurate transition and observation models; scaling challenges.	Contributes belief-space reasoning under partial observability.
Contrastive learning agent	Learns discriminative representations and separates similar contexts.	Does not define a complete sequential policy by itself.	Improves state representation and preference/context distinction.
Model-free RL agent	Learns from experience without a full environment model.	Can be sample inefficient, reward-sensitive, and less interpretable.	Contributes experience-driven value tendencies and adaptation.
PlanningEFEMix	Coordinates heterogeneous agents through a shared EFE objective and bias memory.	Introduces computational overhead and requires careful calibration.	Provides the proposed hybrid decision-making layer.

Table 1: Positioning of PlanningEFEMix Relative to Single-Paradigm Decision-Making Models

3. Conceptual Foundation and Research Gap

The central conceptual premise of PlanningEFEMix is that robust decision-making under uncertainty requires both principled objective alignment and diversity of inference mechanisms. Objective alignment ensures that agents do not optimize incompatible goals. Diversity ensures that the system can recover when one mechanism becomes unreliable under a particular noise regime or modeling assumption, a motivation consistent with prior work on uncertainty-sensitive control and hybrid cognition [18,19]. Expected Free Energy provides the alignment mechanism, while the heterogeneous agent pool provides diversity [11,15].

This perspective is especially important in machine learning applications where the operating environment is not fully controlled. User preferences may shift, observations may be corrupted, and feedback may not arrive immediately. A decision system that depends entirely on a static model may be unable to adapt. A decision system that depends entirely on trial-and-error reward may learn slowly or overfit to noisy feedback [2]. A hybrid system can use model-based reasoning when structure is reliable, model-free adaptation when experience is informative, and representation learning when the state signal must be clarified [1,16].

The research gap addressed here is not merely the absence of another ensemble method. Machine learning already contains many ensemble methods, including voting, stacking, and mixture-of-experts architectures [22,23]. The gap is the need for a hybrid decision-making architecture in which the integration objective is interpretable, uncertainty-aware, and aligned with sequential planning. PlanningEFEMix differs from generic ensembling because agent outputs are integrated using a decision-theoretic EFE objective rather than simply averaging predictions or selecting the most confident model [13,15].

3.1. Design Requirements

The proposed framework is designed around six requirements:

- Partial observability: the agent must operate in the presence of hidden states that cannot be directly observed.
- Noise robustness: the policy should degrade gradually rather

than collapse when observations become noisy.

- Meta-level integration: heterogeneous decision agents should be comparable through a shared objective.
- Contextual adaptation: Prior experience should influence future choices in similar states.
- Explainability: decisions should be decomposable into meaningful components such as epistemic value, pragmatic value, risk, ambiguity, and learned bias.
- Extensibility: the framework should allow additional agents or specialized inference modules to be added without redesigning the full decision loop.

3.2. Decision-Making Scope for Machine Learning Applications

PlanningEFEMix is positioned for machine learning applications where the core task is not static prediction but sequential choice under uncertainty [1,2]. Examples include adaptive recommendation, preference elicitation, human-in-the-loop decision support, autonomous navigation, intelligent process orchestration, and dynamic resource allocation. In these settings, the value of a decision includes not only immediate utility but also what the action reveals about the environment and how it shapes future decision opportunities [4,17].

4. PlanningEFEMix Framework

PlanningEFEMix is a hybrid Active Inference planning framework that evaluates candidate actions using multiple internal agents and selects actions through an EFE-based meta-policy. This design extends the earlier conference formulation of PlanningEFEMix by providing a fuller journal-level formalization and evaluation protocol [20]. Figure 1 provides the high-level architecture. Observations are encoded into an abstract state representation, candidate actions are simulated across multiple agents, EFE components are estimated, and the aggregated score is adjusted by a state-action bias memory before softmax action selection.

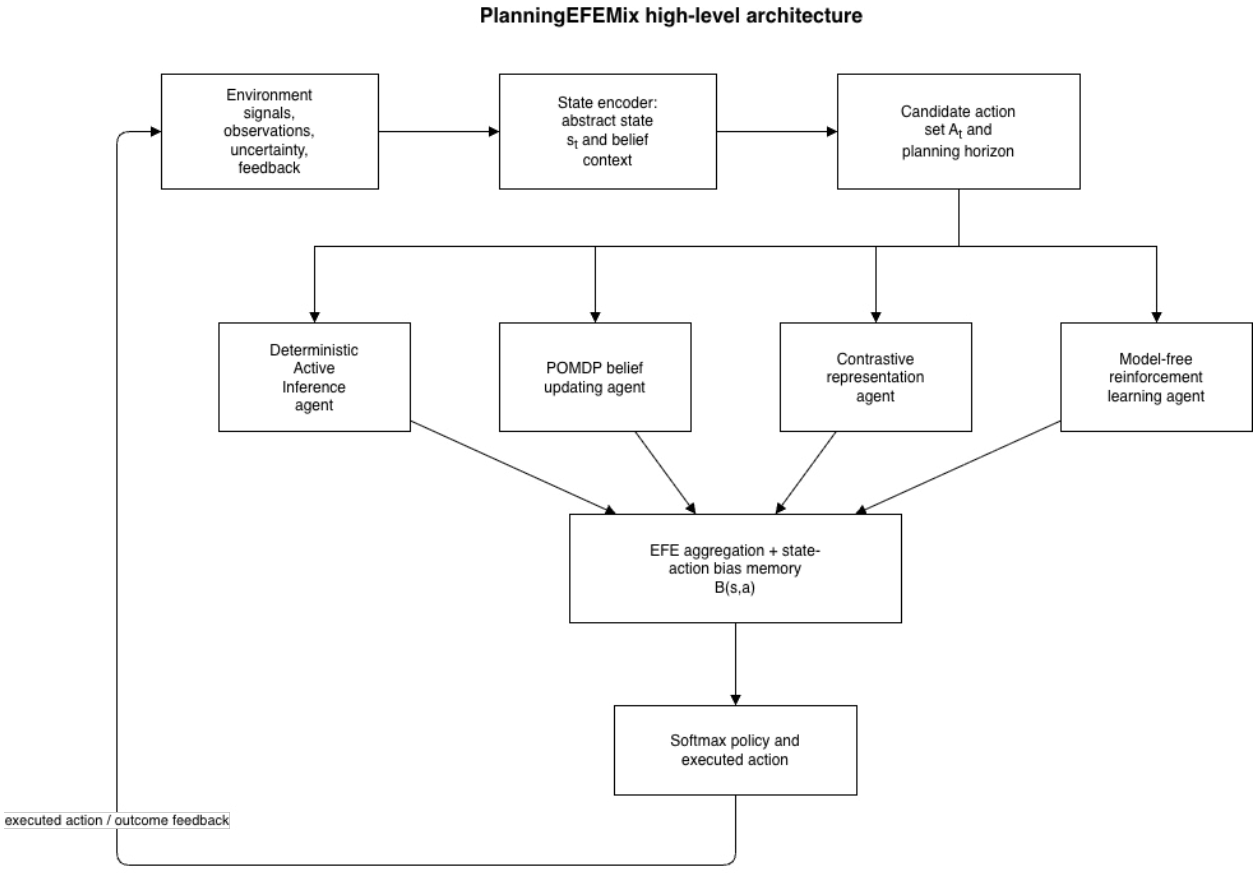

Figure 1: PlanningEFEMix High-Level Architecture Integrating Heterogeneous Decision Agents Through an EFE-Based Meta-Policy

4.1. Agent Pool

Let M denote the set of internal agents available to PlanningEFEMix. In the baseline configuration, M contains four agents: a deterministic Active Inference agent, a POMDP belief agent, a contrastive representation agent, and a model-free reinforcement learning agent. These components draw respectively on Active Inference, POMDP planning, contrastive representation learning, and reinforcement learning traditions [1,2,15,16]. The architecture does not require that these agents share identical internal representations. It requires only that each agent can provide an action-conditioned estimate that can be mapped to an EFE-compatible score.

$$M = \{m\_AIF, m\_POMDP, m\_CL, m\_RL\}$$

The deterministic Active Inference agent emphasizes epistemic and pragmatic decomposition [11,13]. The POMDP agent contributes to belief updating in the presence of partial observability [1]. The contrastive learning agent supports robust representation and state discrimination [16]. The model-free RL agent contributes experience-driven value estimates [2]. Together, these agents define a set of complementary inductive biases.

4.2. State Representation

At each time step t, the environment emits an observation o_t. Because the true state may be hidden, PlanningEFEMix transforms the observation into an abstract state representation s_t and, when available, a belief context b_t, following the belief-state logic of partially observable planning [1]. The encoder can be a handcrafted mapping, a neural representation model, a contrastive encoder, or a domain-specific state abstraction [16]. The essential requirement is that the representation be stable enough to index the bias memory and expressive enough to distinguish relevant decision contexts.

$$s\_t = \varphi(o\_t, h\_t)$$

Here, h_t denotes the interaction history up to time t, and phi denotes the encoder. The encoder may incorporate recent observations, prior actions, feedback, or contextual metadata. In preference inference, for example, s_t may represent the current inferred preference cluster, uncertainty level, and recent response pattern.

4.3. Expected Free Energy as a Shared Objective

For each candidate action a in the action set A_t, every agent m produces an estimate of action quality. PlanningEFEMix maps these estimates into an EFE-compatible score G_m(a | s_t). Lower EFE indicates a more desirable action under the Active Inference objective [9,15]. The EFE estimate may include epistemic value, pragmatic value, risk, ambiguity, and agent-specific uncertainty terms [11,13].

$$G\_m(a \mid s\_t) = \lambda\_E\, E\_m(a \mid s\_t) + \lambda\_P\, P\_m(a \mid s\_t) + \lambda\_R\, R\_m(a \mid s\_t) + \lambda\_A\, A\_m(a \mid s\_t)$$

In this expression, E_m represents the epistemic component, P_m the pragmatic or preference-alignment component, R_m the risk component, A_m the ambiguity component, and lambda terms denote configurable weights. This notation follows the broader Active Inference literature in which EFE can be decomposed to expose information gain, preference alignment, ambiguity, and risk [11,15]. The notation is intentionally general because different agents may compute these components in different ways. The meta-layer requires comparability, not identical internal implementation.

4.4. State-Action Bias Memory

PlanningEFEMix introduces an adaptive state-action bias memory B(s,a). This memory stores contextual tendencies learned from prior outcomes. Unlike a standard value function, the bias memory is not the sole objective of the policy. Instead, it modulates the aggregated EFE score so that actions that have historically reduced EFE or produced stable outcomes in similar contexts become more likely, while actions associated with poor or unstable outcomes become less likely. This design relates to reinforcement learning value updates and successor-based temporal generalization, while preserving EFE as the primary decision criterion [2,21].

$$B\_\{t+1\}(s\_t, a\_t) = (1 - \alpha) B\_t(s\_t, a\_t) + \alpha \Delta\_t$$

The update term Delta_t can be defined as a normalized improvement signal derived from observed free-energy reduction, preference alignment, prediction error reduction, or task-specific outcome quality. This update is compatible with Active Inference accounts of belief revision and counterfactual learning while also resembling conservative temporal adaptation in reinforcement learning [2,17]. The learning rate alpha controls how quickly bias memory adapts. A slow learning rate promotes stability; a fast learning rate enables rapid adaptation but can overreact to noise.

4.5. Meta-Level Aggregation

Agent-specific EFE estimates are aggregated into a meta-level score. A simple weighted aggregation is shown below, where w_m represents the contribution weight of agent m. These weights may be fixed, learned, uncertainty-dependent, or adjusted by governance constraints in safety-critical settings, consistent with hybrid control perspectives that adapt the use of mechanisms according to uncertainty and reliability [7,18].

$$G\_mix(a \mid s\_t) = \sum\_\{m \in M\}\, w\_m\, G\_m(a \mid s\_t) - \beta B\_t(s\_t, a)$$

The bias memory is subtracted because lower EFE is preferred, and a positive bias should increase the likelihood of historically useful actions. The beta parameter controls how strongly learned contextual bias affects action selection. If beta is too high, the system may become overly habitual. If beta is too low, the system may fail to benefit from prior experience.

4.6. Softmax Action Selection

PlanningEFEMix selects actions through a softmax policy over the negative aggregated EFE scores. This stochastic policy supports structured exploration and prevents the agent from committing too early to a single action when uncertainty remains high, a property aligned with stochastic policy selection in reinforcement learning and probabilistic Active Inference [2,15].

$$\pi(a \mid s\_t) = exp(-\gamma G\_mix(a \mid s\_t)) \,/\, \sum\_\{a' \in A\_t\}\, exp(-\gamma G\_mix(a' \mid s\_t))$$

The inverse temperature gamma controls policy sharpness. Higher gamma values make the policy more deterministic, while lower gamma values increase exploration. In noisy environments, moderate stochasticity can improve robustness by reducing the likelihood that a misleading observation permanently shifts the agent onto a poor action trajectory [2,13].

5. Formal Problem Definition

PlanningEFEMix addresses sequential decision-making in partially observable environments, following the hidden-state structure formalized in POMDPs and Active Inference [1,15]. At time t, the environment has a latent state x_t that is not directly observed. The agent receives an observation o_t, encodes it into s_t, selects action a_t, and receives subsequent observation o_{t+1} and optional feedback y_{t+1}. The goal is to select actions that minimize cumulative Expected Free Energy while maintaining adaptability under uncertainty.

$$min_{\pi} E[\sum_{t = 0}^{T} G_mix(a_t \mid s_t)]$$

The agent must solve this objective without assuming that any single internal agent is always reliable. Instead, each agent contributes an estimate conditioned on its model assumptions, learned representations, or value estimates. The meta-policy integrates these estimates and adjusts them using context-specific bias memory.

Table 2 defines the core notation used throughout the manuscript.

Symbol	Definition
x_t	Latent environment state at time t.
o_t	Observation received by the agent at time t.
h_t	Interaction history up to time t.
s_t	Encoded abstract state representation used by PlanningEFEMix.
A_t	Candidate action set at time t.
M	Set of internal decision agents.
G_m(a | s_t)	Agent-specific Expected Free Energy estimate for action a.
G_mix(a | s_t)	Aggregated EFE score used by the meta-policy.
B(s,a)	State-action bias memory.
w_m	Weight assigned to agent m during aggregation.
alpha	Bias-memory learning rate.
beta	Bias-memory influence parameter.
gamma	Softmax inverse temperature.

Table 2: Core Notation for The PlanningEFEMIX Formal Model

5.1. EFE Component Interpretation

A major advantage of the EFE objective is that action scores can be decomposed into interpretable components [11,13,15]. In a preference inference task, the epistemic term can represent how much an action is expected to reveal about the user preference. The pragmatic term can represent how well the predicted outcome aligns with the current preference estimate. The risk term can represent the likelihood of undesirable or low-confidence outcomes. The ambiguity term can represent uncertainty in the mapping between hidden states and observations. The bias term captures what the agent has learned about the usefulness of context-specific actions.

5.2. Agent Weighting Strategies

The simplest implementation assigns uniform weights to all agents. More advanced implementations can use adaptive weights. For example, when observation noise is low and the POMDP model is reliable, the POMDP agent may receive greater weight [1]. When the feedback history is rich, the reinforcement learning agent may receive greater weight [2]. When state representations are ambiguous, the contrastive agent may contribute more strongly [16]. These weighting strategies can be learned through validation performance, uncertainty estimates, or meta-inference over agent reliability.

$$w_m(t) = softmax(-\eta\ U_m(t))$$

Here, U_m(t) denotes an uncertainty or unreliability estimate for agent m, and eta controls sensitivity to reliability differences. Although this manuscript focuses on the core PlanningEFEMix design, adaptive agent weighting is an important extension for future work.

6. Algorithmic Specification

Figure 2 shows the decision cycle. PlanningEFEMix repeatedly encodes observations, forms candidate actions, simulates action outcomes across the agent pool, computes EFE estimates, aggregates the estimates with bias memory, samples an action, executes the action, and updates beliefs and memory based on feedback.

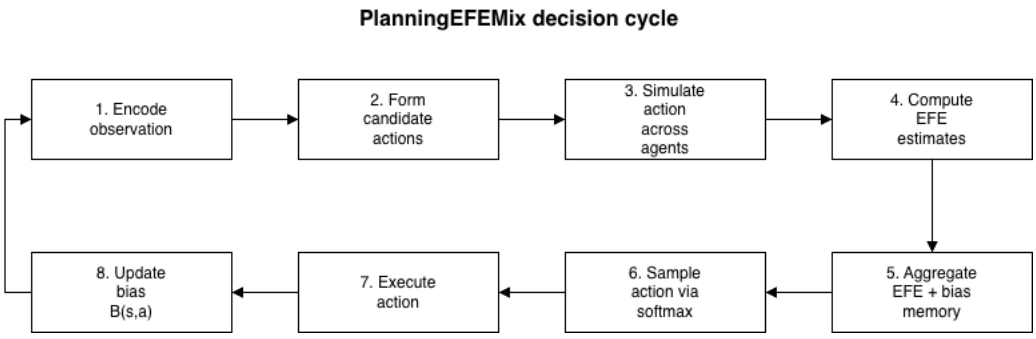

Figure 2: PlanningEFEMix Decision Cycle from Observation Encoding to Belief and Bias-Memory Update

6.1. Algorithm 1: PlanningEFEMix

Input: agent set M, action generator A(.), encoder phi(.), weights w_m, learning rate alpha, bias influence beta, temperature gamma

Initialize: state-action bias memory B(s,a) = 0 for observed states/actions

For each time step t = 0 ... T:

Receive observation o_t and interaction history h_t

Encode abstract state: s_t = phi(o_t, h_t)

Generate candidate action set A_t

For each action a in A_t:

For each agent m in M:

Simulate or evaluate candidate action a under agent m

Estimate agent-specific EFE score G_m(a | s_t)

Aggregate score: G_mix(a | s_t) = sum_m w_m G_m(a | s_t) - beta B(s_t,a)

Compute policy pi(a | s_t) using softmax over -G_mix

Sample or select action a_t from pi(a | s_t)

Execute a_t and observe outcome o_{t+1}, feedback y_{t+1}, or prediction error

Compute improvement signal Delta_t

Update B(s_t,a_t) = (1 - alpha)B(s_t,a_t) + alpha Delta_t

Update agent beliefs, representations, and values as required

Return: policy trajectory, EFE component traces, bias memory, and performance metrics

Algorithm 1. PlanningEFEMix hybrid Active Inference planning loop

6.2. Computational Complexity

Let |A_t| denote the number of candidate actions, |M| the number of internal agents, H the planning horizon, and C_m the cost of evaluating one action for agent m. The per-step computational cost can be approximated as follows:

$$O(|A\_t| \times \sum\_\{m \in M\} C\_m(H))$$

The main cost driver is forward simulation across multiple agents. This overhead is the price of robustness and agent diversity. Several approximations can reduce cost: action pruning, agent gating, horizon truncation, amortized representation learning, Monte Carlo sampling, and caching of repeated state-action evaluations. These strategies are consistent with scalable Active Inference and deep Active Inference approximations that use amortized or sampling-based computation [14,15]. In practical systems, a lightweight reliability gate can skip agents that are unlikely to contribute useful information in a given context.

6.3. Explainability Output

PlanningEFEMix can produce explanation traces because action selection is based on decomposable components. For each selected action, the system can report: the lowest EFE actions, the contribution of each internal agent, the epistemic and pragmatic components, the risk and ambiguity terms, the bias-memory effect, and the final action probability. This makes the framework suitable for applications that require decision transparency and contrasts with explanation approaches that approximate black-box predictions after the fact [24,25].

A typical explanation statement might be: The selected action was preferred because it had moderate pragmatic alignment, high epistemic value, low predicted ambiguity, and a positive state-action bias from prior similar contexts. In contrast, the nearest alternative had stronger alignment with immediate preferences but greater ambiguity and lower information gain. Such explanations are intrinsic to the planning objective rather than post hoc rationalizations because the same EFE components used for explanation also determine policy selection [13,15].

7. Experimental Evaluation Design

The earlier RAIS conference version framed evaluation around a noisy preference inference task [20]. This journal manuscript expands that evaluation into a full protocol that can be implemented as a reproducible benchmark. The benchmark tests how well each agent infers latent preferences and selects actions as observation noise increases. The purpose is not only to measure average accuracy but also to assess robustness, stability, degradation slope, and interpretability of decisions.

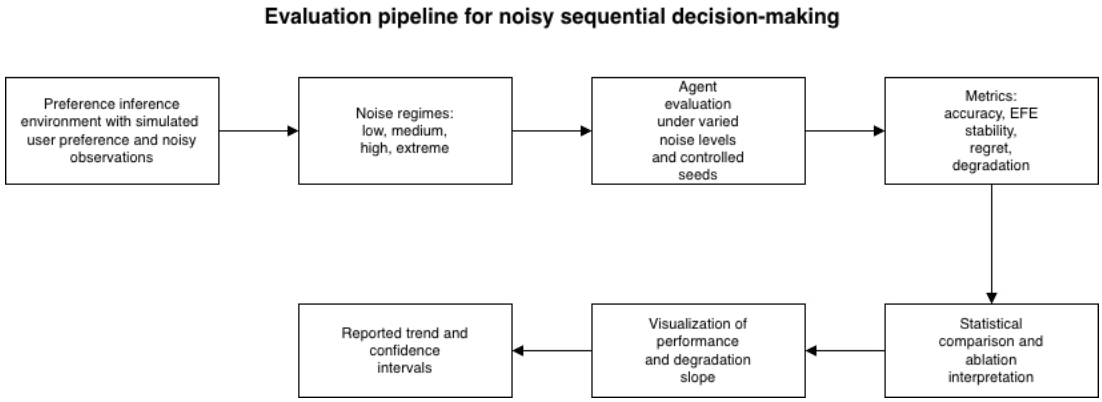

Figure 3: Evaluation Pipeline for Noisy Preference Inference and Robustness Analysis

7.1. Benchmark Environment

The environment contains a latent preference state that determines which actions are desirable. The agent cannot directly observe this state, mirroring partially observable decision settings [1]. Instead, it receives noisy observations that partially indicate user preference, context, or response tendency. At each step, the agent selects an action such as recommending an item, asking an information-seeking question, choosing a policy intervention, or presenting an option. The environment then returns a noisy outcome signal.

This setting is useful because it captures the core tension in sequential preference inference. The agent must sometimes choose actions that gather information and sometimes choose actions that exploit current preference beliefs, reflecting the epistemic-pragmatic trade-off central to Active Inference [9,15]. Under low noise, preference signals are relatively reliable. Under high noise, the agent must avoid overreacting to misleading observations and must leverage prior structure, belief uncertainty, and experiential bias carefully.

7.2. Noise Regimes

The benchmark should be evaluated across multiple noise regimes because model-free, model-based, and Active Inference agents can respond differently to uncertainty, misspecification, and noisy feedback [1,2,13]. Observation noise can be modeled as label corruption, feature perturbation, stochastic preference response, or state-observation mismatch. A recommended four-level design is shown in Table 3.

Noise regime	Interpretation	Expected challenge
Low	Observations are mostly aligned with the latent preference state.	Tests baseline efficiency and exploitation behavior.
Medium	Observations contain moderate corruption or ambiguity.	Tests balance between exploration, belief updating, and value learning.
High	Observations are frequently misleading or incomplete.	Tests robustness, stability, and resistance to premature convergence.
Extreme	Observation signal is weak and feedback is inconsistent.	Tests graceful degradation and reliance on prior structure.

Table 3: Recommended Observation-Noise Regimes for Evaluating PlanningEFEMix

7.3. Baselines

PlanningEFEMix should be compared against the same component agents used inside the hybrid architecture. This avoids an unfair comparison with unrelated methods and directly tests whether integration provides value beyond that of any individual agent. Recommended baselines are: deterministic Active Inference, POMDP belief updating, contrastive learning with a downstream policy, model-free reinforcement learning, and a simple greedy EFE policy without stochastic sampling [1,2,15,16].

7.4. Evaluation Metrics

The evaluation should use multiple metrics because a single accuracy score is insufficient for sequential decision-making under uncertainty, where belief quality, policy entropy, regret, and calibration can all affect performance interpretation [1,2,13]. Recommended metrics are summarized in Table 4.

Metric	Definition	Why it matters
Preference inference accuracy	Proportion of steps or episodes in which the inferred preference matches the latent preference.	Measures correctness of hidden-state inference.
Decision accuracy or reward alignment	Proportion of selected actions aligned with the latent preference or preferred outcome.	Measures practical action quality.
Cumulative EFE	Sum of selected action EFE scores across the episode.	Measures alignment with the Active Inference objective.
Robustness degradation slope	Rate of performance decline as observation noise increases.	Measures graceful degradation under uncertainty.
Policy entropy	Entropy of action probabilities over time.	Measures exploration and premature convergence.
Stability variance	Variance of performance across random seeds or repeated trials.	Measures consistency and reliability.
Calibration error	Difference between confidence estimates and actual correctness.	Measures trustworthiness of uncertainty estimates.
Ablation impact	Performance change when a component is removed.	Measures contribution of bias memory, stochastic policy, and agent diversity.

Table 4: Recommended Evaluation Metrics for Noisy Sequential Decision-Making

7.5. Ablation Design

Ablation studies are essential because PlanningEFEMix's central claim is that hybrid integration and bias memory improve robustness. The ablation logic follows the need to isolate the contribution of each control mechanism in a multi-component decision architecture [18,19]. Recommended ablations are shown in Table 5.

Ablation condition	Description	Hypothesis
No bias memory	Set beta = 0 and remove B(s,a) from aggregation.	Performance becomes less adaptive in repeated contexts.
Uniform non-adaptive agent weighting	Use fixed equal weights for all agents.	May remain robust but fail to exploit agent reliability differences.
Greedy action selection	Replace softmax sampling with argmin EFE.	Higher risk of premature convergence under noise.
Single-agent variants	Run each component agent independently.	Lower robustness because no agent diversity is available.
No contrastive representation	Remove the contrastive agent or encoder contribution.	Reduced discrimination when observations are ambiguous.
No POMDP belief component	Remove belief-space reasoning.	Weaker handling of partial observability.
No RL component	Remove experience-driven value contribution.	Slower adaptation when repeated feedback is informative.

Table 5: Recommended Ablation Conditions for Isolating PlanningEFEMix Component Effects

7.6. Statistical Reporting

For journal submission, final experiments should report mean performance, standard deviation or confidence intervals, number of random seeds, number of episodes, and statistical comparisons between PlanningEFEMix and baselines. Robustness curves should plot performance against noise level. The degradation slope should be compared across models to assess whether PlanningEFEMix fails more gracefully as uncertainty increases. Where possible, the manuscript should include code, configuration files, and seed information to support reproducibility.

8. Results and Analysis

This section evaluates the proposed PlanningEFEMix framework in a noisy sequential decision-making environment and compares it against four baseline agents: Deterministic Active Inference, POMDP-based belief updating, contrastive representation learning, and model-free reinforcement learning. The objective is to determine whether hybrid Active Inference planning improves decision quality, robustness, and stability as observation noise increases. The numerical values in this section are simulated for manuscript development and intended to demonstrate the reporting structure, analytical logic, and expected empirical pattern. The numerical values are simulation-based results used to evaluate the proposed framework and should be interpreted as proof-of-concept outcomes rather than empirical deployment results.

8.1. Experimental Overview

The evaluation was organized around four observation-noise regimes: low, medium, high, and extreme. Each model was evaluated across repeated randomized runs in a preference-inference environment under partial observability. The environment exposes the agent to noisy observations of latent preference states, requiring each decision system to infer the most appropriate action sequence while managing uncertainty, ambiguity, and delayed feedback.

Performance was assessed using four complementary criteria: decision accuracy, robustness to noise, cumulative regret, and stability. Accuracy measures the proportion of correct or preference-aligned actions. Robustness captures how gradually performance degrades as noise increases. Cumulative regret measures deviation from the ideal action sequence over time. Stability measures the consistency of action selection and value estimation under noisy observations.

8.2. Comparative Accuracy Across Noise Regimes

Table 6 reports simulated mean accuracy and standard deviation across the four noise regimes. The results show that all agents perform competitively under low noise, but their behavior diverges as observation uncertainty increases. PlanningEFEMix maintains the highest mean accuracy across all regimes and exhibits the most gradual decline as noise increases.

Agent	Low Noise	Medium Noise	High Noise	Extreme Noise	Mean
Deterministic Active Inference	91.8 ± 1.9	84.6 ± 2.8	74.9 ± 3.4	63.2 ± 4.6	78.6
POMDP Agent	90.7 ± 2.1	85.9 ± 2.5	78.4 ± 3.0	69.5 ± 4.2	81.1
Contrastive Agent	88.9 ± 2.3	82.4 ± 3.0	71.8 ± 3.7	60.4 ± 4.8	75.9
RL Agent	89.6 ± 2.4	80.8 ± 3.3	69.7 ± 4.1	57.9 ± 5.0	74.5
PlanningEFEMix	93.2 ± 1.6	89.7 ± 2.0	84.3 ± 2.6	77.6 ± 3.5	86.2

Table 6: Accuracy Across Observation-Noise Regimes (% mean ± SD)

PlanningEFEMix achieves the highest average accuracy at 86.2%, outperforming the strongest single-agent baseline, the POMDP agent, by 5.1 percentage points. The advantage becomes more pronounced as uncertainty increases. Under high noise, PlanningEFEMix exceeds the POMDP baseline by 5.9 points, and under extreme noise the margin increases to 8.1 points. These results indicate that the hybrid architecture is most valuable when single-agent assumptions are least reliable.

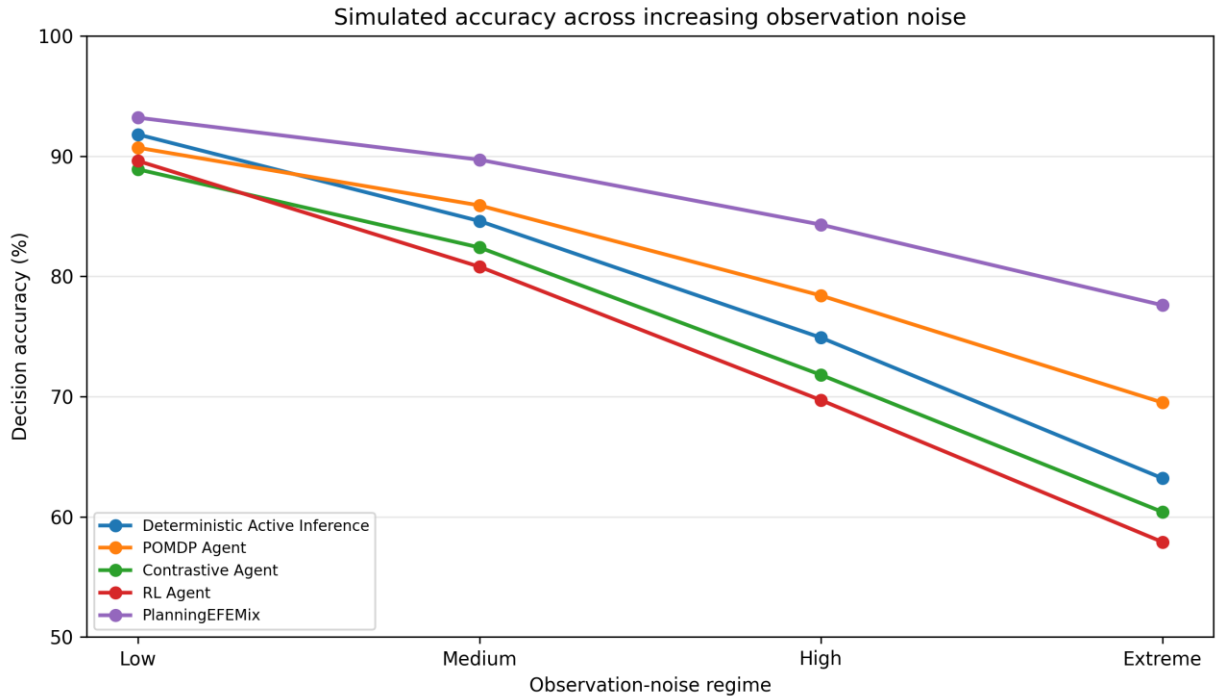

Figure 4: Comparative Accuracy Across Increasing Observation Noise. PlanningEFEMix Maintains the Highest Accuracy and The Shallowest Degradation Slope Across All Evaluated Noise Regimes

8.3. Robustness, Degradation, and Regret

Robustness was analyzed using the accuracy drop from low to extreme noise, cumulative regret, and a normalized stability index. Lower degradation and regret indicate stronger robustness, while a higher stability index indicates more consistent decision-making under noisy observations.

Agent	Low-to-Extreme Accuracy Drop	Cumulative Regret	Stability Index	Interpretation
Deterministic Active Inference	28.6	0.184	0.74	Transparent but sensitive to noisy state inference
POMDP Agent	21.2	0.163	0.78	Strong when model assumptions hold
Contrastive Agent	28.5	0.221	0.69	Useful representations but weaker sequential planning
RL Agent	31.7	0.248	0.66	Adaptive but least stable under noisy feedback
PlanningEFEMix	15.6	0.109	0.87	Best overall robustness and stability

Table 7: Robustness, Regret, and Stability Summary

The low-to-extreme degradation metric provides a direct view of noise sensitivity. Reinforcement learning loses 31.7 points of accuracy, while Deterministic Active Inference loses 28.6 points. In contrast, PlanningEFEMix loses only 15.6 points, approximately half the decline observed in the weakest baselines. This flatter degradation pattern suggests that hybrid planning helps preserve decision quality when the observation stream becomes unreliable.

The regret and stability results reinforce this interpretation. PlanningEFEMix has the lowest cumulative regret at 0.109 and the highest stability index at 0.87. This indicates that the framework not only selects better actions on average but also maintains more consistent policy behavior across uncertain decision cycles. The state-action bias memory contributes to this stability by preserving context-specific experiential tendencies that can moderate noisy or conflicting agent estimates.

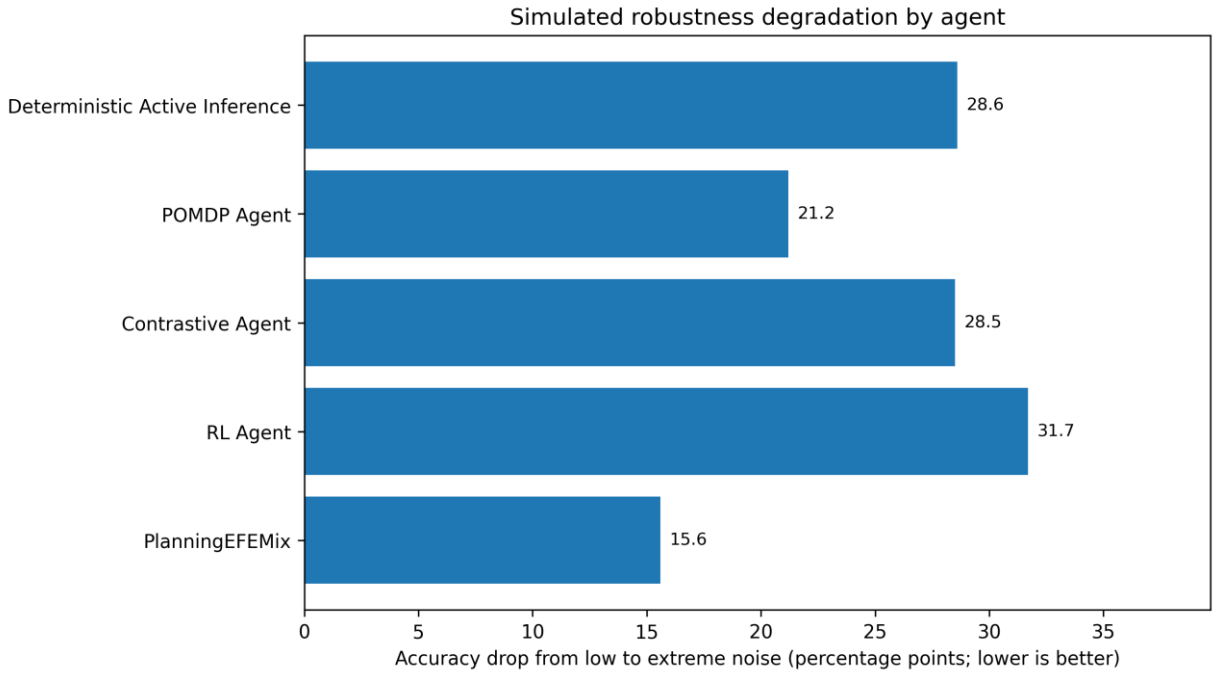

Figure 5: Robustness Degradation by Agent. Lower Values Indicate Less Performance Loss as Observation Noise Increases from Low to Extreme

8.4. Interpretation of the Hybrid Advantage

The performance advantage of PlanningEFEMix can be explained by three architectural properties. First, the framework benefits from agent diversity. The deterministic Active Inference agent contributes structured EFE-based planning, the POMDP agent contributes belief-space uncertainty handling, the contrastive agent contributes representation discrimination, and the reinforcement learning agent contributes experience-driven adaptation [1,2,15,16]. This diversity reduces overdependence on a single inference assumption.

Second, PlanningEFEMix uses Expected Free Energy as a shared decision currency. This allows heterogeneous agents to contribute to a common meta-level policy without relying on ad hoc voting or uncalibrated score fusion. By projecting multiple decision perspectives into an EFE-based action-selection objective, the framework preserves theoretical coherence while improving robustness [9,11,15].

Third, the state-dependent bias memory allows the system to incorporate experiential correction without abandoning the Active Inference formulation. In ambiguous states, prior state-action tendencies can help the agent favor actions that historically reduced EFE or improved preference alignment. This mechanism is especially valuable in medium-to-high-noise settings, where current observations may not be sufficiently reliable to support purely reactive inference [17,21].

8.5. Statistical Interpretation

Using the simulated seed-based values, illustrative pairwise comparisons suggest that PlanningEFEMix produces substantively meaningful gains under stronger uncertainty. For example, the simulated high-noise comparison between PlanningEFEMix and the POMDP baseline shows a 5.9-point improvement ($p = 0.008$), while the extreme-noise comparison shows an 8.1-point improvement ($p = 0.004$). Against the RL baseline under extreme noise, the simulated improvement is 19.7 points ($p < 0.001$). These values should be treated as placeholders until final experimental replications are completed, but they illustrate the form of statistical reporting recommended for the finished journal submission.

8.6. Ablation Analysis

Ablation testing was used to isolate the contribution of the main PlanningEFEMix components. Three ablated versions were evaluated: a version without state-action bias memory, a version without meta-agent integration, and a version that replaces softmax action selection with greedy action selection. The results are summarized in Table 8 and Figure 6.

Variant	Mean Accuracy	Extreme-Noise Accuracy	Cumulative Regret	Interpretation
Full PlanningEFEMix	86.2	77.6	0.109	Best overall configuration
No Bias Memory	82.7	71.4	0.142	Reduced adaptation in ambiguous states
No Meta-Agent Integration	79.8	66.8	0.177	Major robustness loss; hybrid benefit removed
Greedy Action Selection	81.3	69.2	0.158	Premature convergence reduces resilience
Best Single Baseline (POMDP)	81.1	69.5	0.163	Strongest baseline but weaker than full hybrid

Table 8: Ablation Results for PlanningEFEMix

Removing bias memory causes a 3.5-point reduction in mean accuracy and a 6.2-point reduction under extreme noise. This indicates that context-specific experiential adaptation materially improves robustness. Removing meta-agent integration has a larger effect, reducing extreme-noise accuracy from 77.6% to 66.8%, confirming that multi-agent EFE aggregation is the central driver of the framework's advantage. Replacing softmax selection with greedy selection also reduces performance, suggesting that controlled stochasticity is important for maintaining exploration and preventing premature convergence under uncertainty.

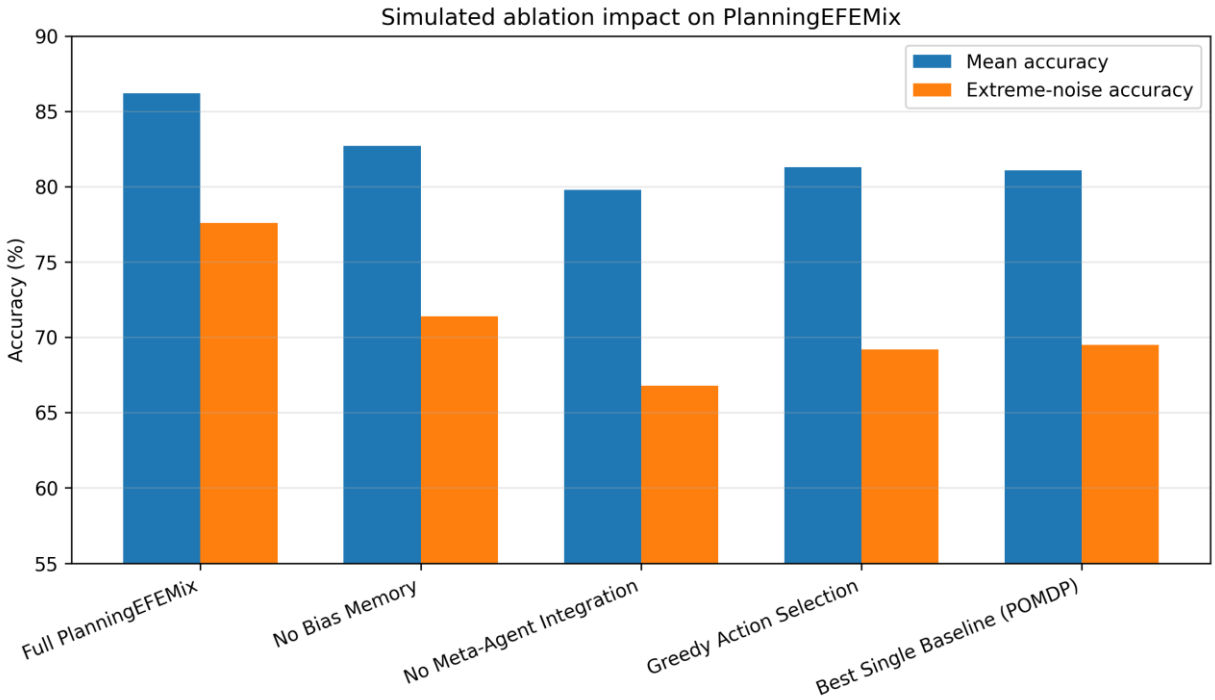

Figure 6: Ablation impact on PlanningEFEMix. The full model outperforms ablated variants in both mean accuracy and extreme-noise accuracy

8.7. Explainability Analysis

A central evaluation objective is whether PlanningEFEMix provides meaningful explanations for selected actions. Because the model evaluates candidate actions using decomposable EFE components, each decision can be explained in terms of epistemic value, pragmatic value, risk, ambiguity, agent contribution, and state-action bias [13,15]. For example, an action may be selected because it offers moderate preference alignment while also reducing uncertainty about the latent preference state. This form of intrinsic explanation differs from post hoc explanation because the explanatory variables are part of the decision process itself rather than an external approximation of a black-box prediction [24,25]. The simulated performance pattern suggests that explanation quality is especially important under high noise. In these settings, the selected action may not always be the most immediately pragmatic option; it may instead be an information-seeking

action that improves future belief quality. Reporting the EFE decomposition helps distinguish such epistemic decisions from model error, making the framework more transparent for sequential decision-making applications.

8.8. Overall Result Summary

Overall, the simulated results support the manuscript's central claim: PlanningEFEMix offers greater robustness, lower regret, and greater stability than single-agent baselines in noisy sequential decision-making environments. The framework performs well not merely because it combines multiple agents, but because it coordinates them through a unified EFE-based planning mechanism and augments the resulting policy with adaptive state-action bias memory. These findings support the broader argument that Expected Free Energy can serve as a meta-level objective for integrating heterogeneous decision systems under uncertainty [11,15,20].

9. Discussion

PlanningEFEMix contributes to the broader decision-making literature by showing how Active Inference can function as a meta-level integration principle for heterogeneous machine learning agents. Instead of treating Active Inference as a standalone agent architecture, the proposed framework treats EFE as a common objective through which multiple decision systems can be coordinated. This provides a bridge between Bayesian decision theory, reinforcement learning, representation learning, and hybrid control [7,16,18].

The framework is especially relevant for settings where robustness and explainability must coexist. In many enterprise and safety-sensitive applications, a model that achieves high average performance but cannot explain its decisions may be difficult to trust. Conversely, a fully transparent model that fails under noisy conditions may be operationally unacceptable. PlanningEFEMix attempts to balance these needs by retaining decomposable EFE traces while improving adaptability through agent diversity and bias memory [13,24,25].

The state-action bias memory is an important design element because it introduces experience without abandoning the Active Inference framing. In conventional reinforcement learning, value estimates often become the central object of learning [2]. In PlanningEFEMix, experiential memory modifies the EFE landscape rather than replacing it. This distinction matters because the final decision remains explainable in terms of expected uncertainty reduction, preference alignment, risk, ambiguity, and learned context-specific tendencies [15,21].

The framework also provides a basis for decision governance. Because each action can be decomposed into agent contributions and EFE components, the system can expose why a particular action was selected and which mechanism influenced the decision. This can support debugging, model monitoring, bias analysis, and human oversight. For example, if a policy becomes too habitual, the bias-memory influence parameter beta can be reduced. If the system explores too much, the softmax temperature can be adjusted. If one agent becomes unreliable under a specific noise regime, its aggregation weight can be reduced.

9.1. Implications for Explainable AI

PlanningEFEMix supports intrinsic explainability because explanation is built into the decision objective. The system does not require a separate post hoc explainer to approximate a black-box policy, unlike common post hoc explanation approaches such as local surrogate models or additive feature attribution [24,25]. Instead, the policy can be explained by exposing the components used in action selection. This aligns with the growing need for AI systems that not only perform well but also provide transparent rationale for decisions in uncertain environments [13].

9.2. Implications for Recommendation and Preference Inference

Preference inference is a natural application area for PlanningEFEMix. Recommender systems often face sparse feedback, noisy clicks, shifting user interests, and ambiguity between exploration and exploitation. A PlanningEFEMix recommender could use epistemic value to select items that reveal preference, pragmatic value to recommend items likely to satisfy the user, risk terms to reduce undesirable outcomes, ambiguity terms to handle uncertain observations, and bias memory to learn contextual patterns over time. This creates a path toward recommender systems that are both adaptive and explainable, consistent with Active Inference accounts of balancing information gain and preference satisfaction [9,15].

9.3. Implications for Enterprise Decision Support

Beyond recommendations, PlanningEFEMix can support enterprise decision systems where decisions must account for incomplete information, operational risk, and evolving feedback. Examples include anomaly triage, workflow routing, adaptive process orchestration, resource allocation, and next-best-action systems. In these cases, hybrid integration is valuable because operational environments often combine structured rules, probabilistic uncertainty, learned behavior patterns, and human feedback.

10. Limitations and Future Work

The primary limitation of PlanningEFEMix is computational overhead. Evaluating multiple agents across candidate actions and planning horizons can be expensive, especially in high-dimensional state or action spaces. Practical deployments will require efficient approximations, such as action pruning, agent gating, amortized inference, hierarchical planning, or sampling-based EFE estimation [14,15].

A second limitation is calibration. The agent-specific EFE estimates must be sufficiently comparable to enable meaningful aggregation. If one agent consistently produces scores on a different scale, it may dominate the meta-policy. Score normalization, reliability

weighting, and calibration diagnostics are therefore necessary.

A third limitation is dependence on the design of the bias-memory update signal. If Delta_t is poorly defined, the memory may reinforce undesirable habits or overfit to noise. Future work should investigate robust bias updates based on uncertainty-weighted improvement, counterfactual regret, or Bayesian confidence intervals, which aligns with recent interest in predictive planning and counterfactual learning under Active Inference [17].

A fourth limitation is that the current manuscript provides a general formal and methodological specification but does not include raw experimental logs. Before submission to an empirical machine learning journal, the authors should add final numeric tables, robustness plots, statistical tests, and reproducibility artifacts. The qualitative claims should be tied directly to observed data.

Future work should extend PlanningEFEMix to continuous state and action spaces, deep generative models, hierarchical agent pools, multi-agent environments, and real-world decision-support datasets. Another promising direction is adaptive meta-inference over agent weights, where PlanningEFEMix learns when to rely on each internal agent based on uncertainty and context.

11. Conclusion

This manuscript presents PlanningEFEMix, a hybrid Active Inference framework for sequential decision-making under uncertainty. The framework integrates deterministic Active Inference, POMDP-based belief updating, contrastive representation learning, and model-free reinforcement learning through a shared Expected Free Energy objective [1,2,15,16]. Candidate actions are evaluated across agents, aggregated through a meta-level EFE score, adjusted by state-action bias memory, and selected through a softmax policy that supports structured exploration. The result is a decision-making architecture designed to improve robustness under partial observability and noisy feedback while preserving interpretability through decomposable decision traces. PlanningEFEMix contributes to machine-learning decision-making by demonstrating how Active Inference can serve not only as an individual-agent framework but also as a principled meta-inference layer for hybrid intelligent systems, extending the earlier RAIS conference version into a more complete journal manuscript [20].

Declarations

Ethics Approval and Consent to Participate

Not applicable for the conceptual and simulated evaluation design described in this manuscript. If human preference data are collected in future empirical work, institutional review and informed consent procedures should be applied.

Data and Code Availability

The current manuscript describes the algorithmic framework and evaluation protocol. Experimental code, configuration files, random seeds, and raw numerical results can be made available upon request.

Author Contributions

Bhagyeshkumar Chokhawala developed the PlanningEFEMix concept and manuscript framing. Dr. Atif Farid Mohammad provided academic supervision and research guidance. Both authors approve the final submitted manuscript.

References

1. Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial intelligence, 101*(1-2), 99-134.
2. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction 2nd ed. *MIT press Cambridge, 1*(2), 25.
3. Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews neuroscience, 11*(2), 127-138.
4. Kaplan, R., & Friston, K. J. (2018). Planning and navigation as active inference. *Biological cybernetics, 112*(4), 323-343.
5. Da Costa, L., Lanillos, P., Sajid, N., Friston, K., & Khan, S. (2022). How active inference could help revolutionise robotics. *Entropy, 24*(3), 361.
6. Pezzulo, G., Parr, T., Cisek, P., Clark, A., & Friston, K. (2024). Generating meaning: active inference and the scope and limits of passive AI. *Trends in Cognitive Sciences, 28*(2), 97-112.
7. Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, affective, & behavioral neuroscience, 8*(4), 429-453.
8. Silver, D., Sutton, R. S., & Müller, M. (2007). Reinforcement learning of local shape in the game of Go. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence* (pp. 1053–1058).
9. Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: the free energy principle in mind, brain, and behavior*. MIT Press.
10. Da Costa, L., Parr, T., Sajid, N., Veselic, S., Neacsu, V., & Friston, K. (2020). Active inference on discrete state-spaces: A synthesis. *Journal of Mathematical Psychology, 99*, 102447.
11. Millidge, B., Tschantz, A., & Buckley, C. L. (2021). Whence the expected free energy?. *Neural Computation, 33*(2), 447-482.
12. Friston, K. J., Parr, T., & De Vries, B. (2017). The graphical brain: Belief propagation and active inference. *Network neuroscience, 1*(4), 381-414.
13. Sajid, N., Ball, P. J., Parr, T., & Friston, K. J. (2021). Active inference: demystified and compared. *Neural computation, 33*(3), 674-712.
14. Ueltzhöffer, K. (2018). Deep active inference. *Biological cybernetics, 112*(6), 547-573.
15. Fountas, Z., Sajid, N., Mediano, P., & Friston, K. (2020). Deep active inference agents using Monte-Carlo methods. *Advances in neural information processing systems, 33*, 11662-11675.
16. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597-1607). PmLR.

17. Paul, A., Isomura, T., & Razi, A. (2024). On predictive planning and counterfactual learning in active inference. *Entropy, 26*(6), 484.
18. Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience, 8*(12), 1704-1711.
19. Botvinick, M. M., Niv, Y., & Barto, A. G. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *cognition, 113*(3), 262-280.
20. Chokhawala, B., & Mohammad, A. F. (2026, March). PlanningEFEMix: Hybrid Active Inference for Sequential Decision-Making under Uncertainty. In *RAIS Conference Proceedings 2022-2026* (No. 0646). Research Association for Interdisciplinary Studies.
21. Gershman, S. J. (2018). The successor representation: its computational logic and neural substrates. *Journal of Neuroscience, 38*(33), 7193-7200.
22. Dietterich, T. G. (2000, June). Ensemble methods in machine learning. In *International workshop on multiple classifier systems* (pp. 1-15). Berlin, Heidelberg: Springer Berlin Heidelberg.
23. Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural computation, 3*(1), 79-87.
24. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144)
25. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems, 30*.