A Diffusion Model-Based Approach to Active Inference

Elias Aoun Durand, Mateus Joffily, Mehdi Khamassi

A Diffusion Model-Based Approach to Active Inference

Elias AOUN DURAND
ISIR / GATE
CNRS, Sorbonne Université
Paris, France
elias.aoundurand@gmail.com

Mateus JOFFILY
GATE UMR 5229
CNRS, Université de Lyon
69003 Lyon, France
joffily@gate.cnrs.fr

Mehdi KHAMASSI
ISIR UMR 7222
CNRS, Sorbonne Université
75005 Paris, France
mehdi.khamassi@upmc.fr

***Abstract*—One of the current challenges in machine learning is the acquisition of generative models to enable decision making in complex, uncertain environments. In Bayesian methods like Active Inference, this requires the estimation of posterior distributions, which is often intractable. Variational Free Energy (VFE) provides a tractable approximation method for this estimation. This paper proposes a novel framework that leverages the expressive power of diffusion models to estimate free energy quickly and efficiently. Our approach, merging concepts from statistical physics, machine learning, and computational neuroscience, aims to improve the estimation of posterior distributions. We establish a theoretical link between active inference and diffusion models, emphasizing their shared strategies for uncertainty management, action selection, and model optimization. We propose using diffusion processes to approximate environmental generative models and employ score-based guidance for action selection through expected free-energy minimization. Simulations in increasingly complex grid world environments demonstrate the framework's ability to handle partial observability and stochastic transitions, showing enhanced goal-directed behavior and uncertainty resolution compared to traditional methods. This suggests a path toward scalable active inference agents that could reason about uncertainty in real-world contexts, with possible applications to robotics, healthcare, and adaptive control systems.**

***Index Terms*—Reinforcement learning, Active inference, Diffusion models, Decision making, Generative models, Free energy principle.**

I. Introduction

The concept of free energy in active inference, introduced by Friston et al. [1], provides a unifying principle for understanding perception, learning, and action as inferential processes in artificial and biological systems [2]. Central to this paradigm is the Free Energy Principle, which posits that biological systems minimize variational free energy, leading to better approximations of the true posterior distribution of their environment [9]. This constitutes an alternative to the Reinforcement Learning (RL) paradigm [11] – where the agent's ultimate goal is to maximize reward – by framing RL itself as an inference problem [10]. Despite its tempting theoretical elegance and applicability across various research domains, the practical implementation of active inference for real-world problems presents computational challenges, particularly in action selection contexts where the exact computation of free energy can be intractable [14]. The goal of this paper is to improve the computational efficiency and performance of the active inference framework. Specifically, based on the work of [12], [13], we investigate various strategies for the analytical computation of free energy, highlighting its importance as a key challenge in the field. We propose a novel approach that employs diffusion models for this computation, offering new methodologies not only for integrating active inference with diffusion models but also for enhancing the overall performance and applicability of active inference frameworks. Although immediate effectiveness may be limited, this work represents a step forward toward efficient decision-making in complex, uncertain environments, making active inference potentially more broadly applicable to fields like robotics, healthcare, and adaptive systems.

II. Background: Free Energy and Active Inference

At the core of active inference theory [1], the variational free energy is defined as

$$F[q] = D_{KL}[q(z_{0:T} \mid o_{0:T}, a_{0:T}) \parallel p(z_{0:T} \mid o_{0:T})] - \log p(o_{0:T}) \quad (1)$$

Let us break down each component of this equation.

- $F[q]$ is the Variational Free Energy. Quantifies the difference between an organism's internal model of the world and the true state of the world.
- $q(z_{0:T}|o_{0:T}, a_{0:T})$ is the approximate posterior distribution. It represents the beliefs of the organism about the hidden states of the world ($z_{0:T}$), given its observations ($o_{0:T}$) and actions ($a_{0:T}$). This is the internal model that the organism uses to make sense of its sensory inputs and guide its actions.
- $p(z_{0:T}|o_{0:T})$ is the true posterior distribution. It represents the actual distribution of hidden states given the observations. In practice, this is usually unknown and is what the organism is trying to approximate.
- $D_{KL}[\cdot||\cdot]$ is the Kullback-Leibler divergence, a measure of the difference between two probability distributions. In this context, it quantifies how different the beliefs of the organism are from the true state of the world.
- $\log p(o_{0:T})$ is the logarithmic evidence or logarithmic marginal likelihood of the observations. It represents the log-probability of the sensory data under the organism's model, averaged over all possible hidden states.

The selection of a policy π is guided by its expected free energy $G(\pi)$, which influences decision-making through: The probability of selecting policy π is given by:

$$P(\pi) = \sigma(-\gamma G(\pi)) \quad (2)$$

where σ is the sigmoid function and γ is a scaling factor.

The expected free energy for policy π over future time steps τ is:

$$G(\pi) = \sum_{\tau} G_{\tau}(\pi) \quad (3)$$

Here, $G_{\tau}(\pi)$ (see Section IV-C) quantifies the divergence between the hidden states predicted under π and the prior, plus the expected entropy of observations, analogous to the terms used in the breakdown of variational free energy.

III. Bridging Active Inference and Diffusion Models

Although the intersection of reinforcement learning and diffusion models already constitutes a significant body of research [3], in this work, we propose a novel framework that integrates diffusion models with the active inference paradigm for reinforcement learning. Our approach combines the principles of diffusion processes and score-based guidance from diffusion models [4], [5] with the free energy minimization objective of active inference agents. The key challenges addressed by our proposed framework include *generative model approximation*, where diffusion models are used to learn expressive generative models that capture complex distributions over states, observations, and actions, and *efficient expected free energy computation*, by leveraging the conditioned sampling capabilities of diffusion models to efficiently evaluate the expected free energy of potential actions for more effective decision-making.

To understand the conceptual bridge between Active Inference and diffusion models, let us examine the key formulations of the latter. The objective of a diffusion model is to approximate a generative distribution p from which data o have been drawn. This is done by learning a model q that models how pure noise is denoised into real data and taking advantage of the formalization of the forward process in terms of the Gaussian diffusion process.

1) **Forward Process:** The forward process in a diffusion model progressively corrupts the observation o with Gaussian noise, transforming it into a series of increasingly noisy latent variables z_t at each time step t.

2) **Generative Model:** The objective of generative modeling is to minimize KL divergence: $D_{KL}[q(z_{0,\ldots,\tau})||p(z_{0,\ldots,\tau})] \approx 0$
3) **Score Network:** The approximation of the log density gradient by the score network: $s_\theta(z;\lambda) \approx \nabla_z \log q_t(z)$. The score network, denoted by $s_\theta(z;\lambda)$, is a neural network trained to approximate the gradient of the log-probability density $\nabla_z \log q_t(z)$ with respect to latent variables z. Here, $q_t(z)$ represents the distribution of the latent variable z at time t, modeled by q. It reflects the state of the corrupted data at each step of the diffusion process, progressively moving from pure noise towards real data through denoising.

In Active Inference, the expected free energy $G_\tau(\pi)$ that guides action selection is analogous to the minimization of KL divergence in diffusion models, where the data generation process is guided by the score network. This analogy, while the mathematical demonstration of this connection remains an open question, provides an intuitive bridge between the two frameworks. Furthermore, the probability of choosing an active inference policy π, based on the expected free energy, mirrors the optimization of the data generation paths in the diffusion models.

IV. Methods

We propose a framework that incorporates diffusion models into Active Inference.

A. Outline of the proposed method

Step 1: Approximating the generative model. In order to use a diffusion model to approximate the generative model $p(o_{0:t}, z_{0:t}, a_{0:t})$ we have to:

1) Set up a forward diffusion process that gradually adds noise to $(o_{0:t}, z_{0:t}, a_{0:t})$, transforming it into a series of increasingly noisy latent variables.
2) Learn the reverse process q using a score network s_θ to estimate gradients of log-probability density. This network is trained to predict the noise added at each step of the forward process.
3) Train the model by alternating between corrupting the data with noise and learning to reverse this corruption, effectively learning to reconstruct the original data from noise.

This approach reduces the free energy while enabling conditioned sampling for the estimation of expected free energy (EFE).

Step 2: Decision making through EFE minimization. The trained diffusion model is then used for decision-making to:

1) Infer future states and observations by running the reverse diffusion process, starting from noise and conditioning on new observations. This process reconstructs the likely internal states that led to these observations.
2) Select actions by evaluating those that lead to minimal EFE in future states. This involves simulating different action sequences, predicting their outcomes, and choosing the action that minimizes EFE.
3) Execute the chosen action in the environment, observe the outcome, and update beliefs by running the reverse diffusion process on the new observations.

This cycle of action selection, observation, and belief updating continues with the objective of continually reducing the agent's free energy and improving its understanding of the environment. These two steps are summarized in Algorithm 1.

B. Diffusion model for approximating the generative model

Our framework employs a diffusion model to approximate the generative model of the environment. The forward diffusion process is defined as:

$$q(y_t^k|y_t^{k-1}) = \mathcal{N}(y_t^k; \sqrt{1-\beta_t}y_t^{k-1}, \beta_t\mathbf{I}) \quad (4)$$

where y_t^k represents the noisy latent variables and β_t is the variance schedule. A score network learns to reverse this process by estimating the gradient of the log-likelihood [5].

C. Computing EFE for simulated actions

The Expected Free Energy (EFE) for a given policy π is computed as:

Algorithm 1 Active Inference with diffusion-based generative model

Initialize:
Define initial state z_0, observation o_0, action a_0
Initialize policy π with random actions
Initialize dataset $D = \{(z_0, o_0, a_0)\}$
repeat
  Train score network s_θ on dataset D ▷ Learning the generative model
  repeat
    Sample future states and observations $q(z_\tau, o_\tau | z_t, o_t, \pi)$ using conditioned sampling
    Sample a set of candidate policies $\{\pi_0, \ldots, \pi_K\}$
    for $k \leftarrow 0$ to K **do**
      Compute EFE $G(\pi_k)$ for policy π_k
    end for
    Select a_t from π_k minimizing EFE
    Perform a_t and collect z_{t+1} and o_{t+1}
    Add new data to dataset $D \leftarrow D \cup \{(z_{t+1}, o_{t+1}, a_t)\}$ ▷ Update dataset
  until Criteria for retraining the diffusion model
until Terminal state

$$G_t(\pi) = D_{KL}[q(o_t|\pi)||\tilde{p}(o_t)] + E_{q(z_t|\pi)}\left[H(p(o_t|z_t))\right] \quad (5)$$

This combines the epistemic value (first term) and the pragmatic value (second term). Monte Carlo sampling techniques approximate these values when direct computation is infeasible [8].

D. Conditioned sampling

Conditioned sampling, implemented by inpainting [6], generates future states and observations consistent with the current state and potential action sequences. This involves preparing a masked input and performing a conditioning-reverse diffusion process.

E. Action selection

Action selection involves simulating future states and observations, computing epistemic and pragmatic values, and selecting the action that minimizes the total EFE [7]. The process is repeated for sampled policies to balance exploration and goal alignment.

V. Simulations

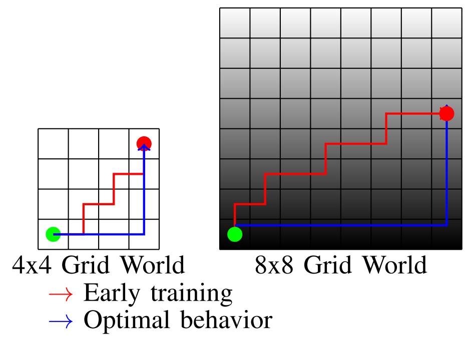


Fig. 1: Grid World Visualizations with Agent Paths: 4x4 (left) and 8x8 (right) environments, showing early training (red) and optimal (blue) paths. The shades of gray in the 8x8 grid represent varying levels of observability, with darker areas being less observable.

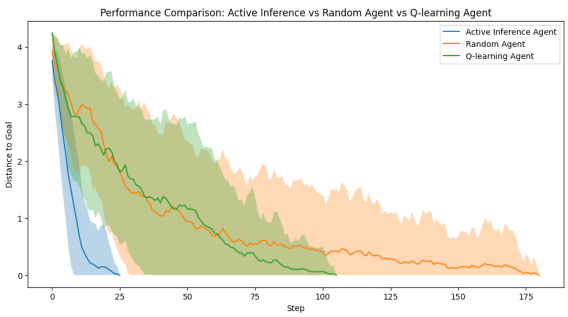
We implemented two toy simulations of active inference with diffusion models (Figure 1):

- A 4x4 grid world with full observability, where the agent navigates between discrete positions.
- An 8x8 grid world with partial observability and stochastic transitions.

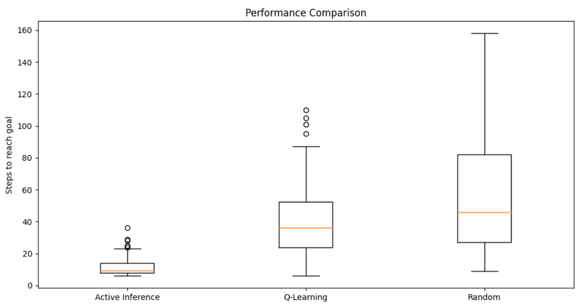
TABLE I: Statistical Analysis of Agent Performance: 4x4 grid

Metric	Active Inference	Q-Learning	Random
Mean steps	11.74	40.79	54.95
Median steps	9.50	36.00	46.00
Standard deviation	5.83	23.15	36.02
Statistical Tests (t-statistic, p-value)			
vs. Random	-11.78, ¡0.001	-3.29, 0.001	-
vs. Q-Learning	-12.11, ¡0.001	-	-
Effect Sizes (Cohen's d)			
vs. Random	-1.67	-0.47	-
vs. Q-Learning	-1.71	-	-

In both simulations, the agent's goal is to navigate to a target state while minimizing Expected Free Energy (EFE). These simulations compare



(a) Distance to Goal over Time



(b) Performance Comparison

Fig. 2: Simulation Results. (a) Average distance to goal versus the number of steps taken. (b) Performance comparison of Active Inference, Q-Learning, and Random action selection in the 4x4 grid world.

three approaches: Active Inference with Diffusion Models, Q-Learning, and a Random Agent. The Active Inference approach uses a ScoreNetwork, implemented as a neural network with transformer layers, to estimate the Expected Free Energy (EFE) for action selection. Incorporating both epistemic and pragmatic values in the EFE computation. The environment includes partial observability and stochastic transitions, modeled through uncertainty matrices. The Q-Learning agent uses a traditional table-based approach with epsilon-greedy exploration. All agents navigate from a start state to a goal state, with performance measured by the number of steps taken.

Figure 2 (a) shows the average distance to the goal state over time for different agents. The Active Inference agent (blue) converges to the goal more quickly than Q-Learning (orange) or Random (green) agents, demonstrating more efficient navigation. As shown in Figure 2, the Active Inference agent consistently achieves the goal in fewer steps in the 4x4 grid world with full observability, with a median of 9.50 steps compared to 36.00 for Q-Learning and 46.00 for the Random agent. Statistical analysis confirms this visual observation, with Active Inference significantly outperforming both Q-Learning ($t = -12.1084, p \leq 0.0001$) and random agents ($t = -11.7843, p \leq 0.0001$).

To further illustrate the decision-making of our Active Inference agent, we examine the evolution of action selection probabilities over time.

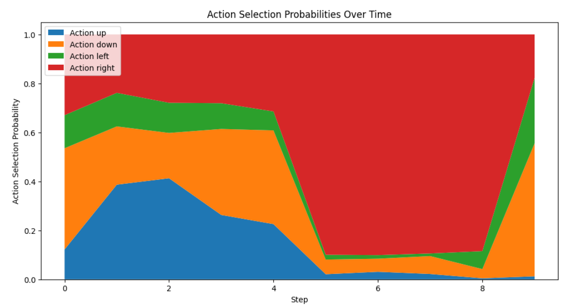


Fig. 3: Action Selection Probabilities Over Time for one Active Inference agent.

Figure 3 illustrates our Active Inference agent's decision-making process, showing how action selection probabilities evolve over time as the agent navigates the environment, with a notable shift toward favoring the "right" action in later steps, likely indicating the agent's increased certainty about the optimal path to the goal.

Figure 4 demonstrates the evolution of Expected Free Energy (EFE) for each possible action over time. Notably, the EFE values for different actions converge towards the end of the time series, suggesting that the agent's uncertainty about optimal actions decreases as it gains more information about the environment.

Although the model demonstrated promising results in the fully observable 4x4 environment, its performance in the 8x8 setting was notably diminished, highlighting the non-trivial nature of scaling

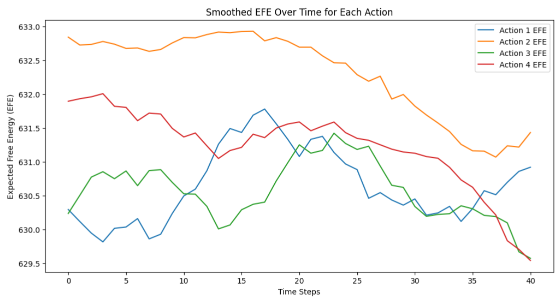


Fig. 4: Smoothed Expected Free Energy (EFE) Over Time for Each Action.

active inference approaches to larger, more realistic scenarios.

VI. Conclusion

We have presented a framework for integrating ideas of how the diffusion model paradigm could drastically reduce the computational complexity of active inference and, more broadly, of control as inference. This has been possible by first showing the mathematical equivalence between some similar operations of the active-inference and diffusion-model frameworks. Then we have presented a series of numerical simulations in increasingly complex grid world environments, demonstrating the new method's ability to handle partial observability and stochastic transitions. Despite this promising bridge, the simulations presented here are only simplistic versions of the diffusion model architecture, and further research remains to fully benefit from the ideas coming from machine learning to produce models accurate enough and not suffering from the curse of dimensionality. However, this work paves the way for future deployment of active inference methods for uncertainty management in real-world scenarios such as robotics, healthcare, and adaptive control systems.

Acknowledgments

This work was supported by the French Agence Nationale de la Recherche (ANR-18-CE28-0016-03) and the CAVAA Project of the European Commission (EIC 101071178).

References

[1] Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., and Pezzulo, G. (2015). Active inference and epistemic value. Cognitive neuroscience, 6(4), 187-214.

[2] Friston, K. (2009). The free-energy principle: a rough guide to the brain? Trends in cognitive sciences, 13(7), 293-301.

[3] Zhu, Z., Zhao, H., He, H., Zhong, Y., Zhang, S., Yu, Y., and Zhang, W. (2023). Diffusion models for reinforcement learning: A survey. arXiv preprint arXiv:2311.01223.

[4] Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, 6840-6851.

[5] Kingma, D., and Gao, R. (2024). Understanding diffusion objectives as the elbo with simple data augmentation. Advances in Neural Information Processing Systems, 36.

[6] Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., and Van Gool, L. (2022). RePaint: Inpainting using Denoising Diffusion Probabilistic Models. *CoRR*, abs/2201.09865.

[7] Millidge, B., Tschantz, A., and Buckley, C. L. (2020). Whence the Expected Free Energy?

[8] Alizadeh Noughabi, H. (2015). Entropy Estimation Using Numerical Methods. *Annals of Data Science*, 2(2), 231–241.

[9] Friston, K. (2010). The free-energy principle: a unified brain theory?. Nature reviews neuroscience, 11(2), 127-138.

[10] Levine, S. (2018). Reinforcement learning and control as probabilistic inference: Tutorial and review. arXiv preprint arXiv:1805.00909.

[11] Sutton, R. S. and Barto, A. (2018). Reinforcement learning: An Introduction A Bradford book.

[12] Gabrié, M. (2020). Mean-field inference methods for neural networks. Journal of Physics A: Mathematical and Theoretical, 53(22), 223002.

[13] Schwobel, S., Kiebel, S., and Markovic, D. (2018). Active inference, belief propagation, and the bethe approximation. Neural computation, 30(9), 2530-2567.

[14] Gershman, S. J. (2023). What does the free energy principle tell us about the brain? *Theory*. Department of Psychology and Center for Brain Science, Harvard University, Cambridge, MA, USA.