

Deep Active Inference with Diffusion Policy and Multiple Timescale World Model for Real-World Exploration and Navigation

Riko Yokozawa, Kentaro Fujii, Yuta Nomura, and Shingo Murata, *Member, IEEE*

Abstract—Autonomous robotic navigation in real-world environments requires exploration to acquire environmental information as well as goal-directed navigation in order to reach specified targets. Active inference (AIF) based on the free-energy principle provides a unified framework for these behaviors by minimizing the expected free energy (EFE), thereby combining epistemic and extrinsic values. To realize this practically, we propose a deep AIF framework that integrates a diffusion policy as the policy model and a multiple timescale recurrent state-space model (MTRSSM) as the world model. The diffusion policy generates diverse candidate actions while the MTRSSM predicts their long-horizon consequences through latent imagination, enabling action selection that minimizes EFE. Real-world navigation experiments demonstrated that our framework achieved higher success rates and fewer collisions compared with the baselines, particularly in exploration-demanding scenarios. These results highlight how AIF based on EFE minimization can unify exploration and goal-directed navigation in real-world robotic settings.

Index Terms—Active inference, autonomous navigation, diffusion policy, free-energy principle, mobile robot, world model

I. INTRODUCTION

AUTONOMOUS robotic navigation in real-world environments requires exploration to acquire environmental information as well as goal-directed navigation in order to reach designated targets efficiently. Achieving an adaptive balance between these two behaviors remains a fundamental challenge in machine learning and robotics [1]–[3]. In many real-world situations, a robot cannot determine its position or the surrounding structure from current observation alone. For example, in visually similar areas such as corridors or intersections, visually similar observations might correspond to multiple possible locations, creating uncertainty in self-localization [4], [5]. In such cases, exploration plays a crucial role by actively collecting additional information to resolve this uncertainty [4]–[6]. When sufficient knowledge has been acquired, the robot must shift its focus to goal-directed navigation in order to reach targets efficiently. Thus, both exploration and navigation are indispensable, and autonomous systems must be able to flexibly balance between them according to the situation.

Traditional approaches such as SLAM-based navigation and handcrafted planners can provide reliable goal-reaching strategies but tend not to generalize to unseen environments or adapt

This work was supported by the Japan Science and Technology Agency (PRESTO Grant Number JPMJPR22C9) and JSPS KAKENHI Grant Number JP24K03012. (Corresponding author: Shingo Murata)

The authors are with the School of Integrated Design Engineering, Keio University, Yokohama, Kanagawa 223-8522, Japan (e-mail: murata@elec.keio.ac.jp).

to unexpected situations [7]–[9]. Recent advances in learning-based methods, including transformer-based policies [10]–[13] and diffusion-based policies [14]–[19], have enabled diverse action generation and shown promising results in navigation tasks. However, these methods rely on explicit planners or extensive task-specific supervision to balance exploration and navigation, limiting their flexibility in real-world settings.

Meanwhile, active inference (AIF) based on the free-energy principle (FEP) [20] offers a unifying framework for exploration and goal-directed behavior by minimizing the expected free energy (EFE) [21]–[23]. EFE comprises two terms: epistemic value, which naturally encourages exploration; and extrinsic value, which accounts for goal-directed behavior. Prior work has demonstrated the potential of AIF-based navigation in simulation environments [24]–[26]; however, applications in real-world robotic systems remain relatively limited, due mainly to challenges in scaling AIF to complex and uncertain environments.

Building on this theoretical foundation, we aim to enhance the scalability of AIF for real-world robotic navigation. Achieving such scalability requires the ability to both generate diverse action sequences depending on the situation and to predict long-horizon state transitions under uncertainty. These capabilities can be realized by leveraging advances in deep generative models, which offer flexible policy representations and powerful predictive dynamics. In this work, we propose a deep AIF framework that integrates a diffusion policy [27] as the policy model and a multiple timescale recurrent state-space model (MTRSSM) [28] as the world model [29]–[31] (Fig. 1). The diffusion policy flexibly generates diverse candidate actions according to the situation [14], while the MTRSSM predicts their long-horizon consequences through latent imagination [28]. Together, these components enable principled action selection under EFE minimization, thereby balancing exploration and goal-directed navigation without relying on handcrafted planners.

The main contributions of this paper are summarized as follows:

- We advance the theoretical foundation of AIF by formalizing how deep generative models can extend the scalability of AIF for real-world navigation, highlighting the role of policy models in generating diverse candidate actions as well as that of world models in supporting long-horizon predictive dynamics.
- We realize this approach by using a deep AIF architecture that integrates a diffusion policy as the policy model and an MTRSSM as the world model, thereby enabling principled action selection via EFE minimization while balancing exploration and goal-directed behavior.

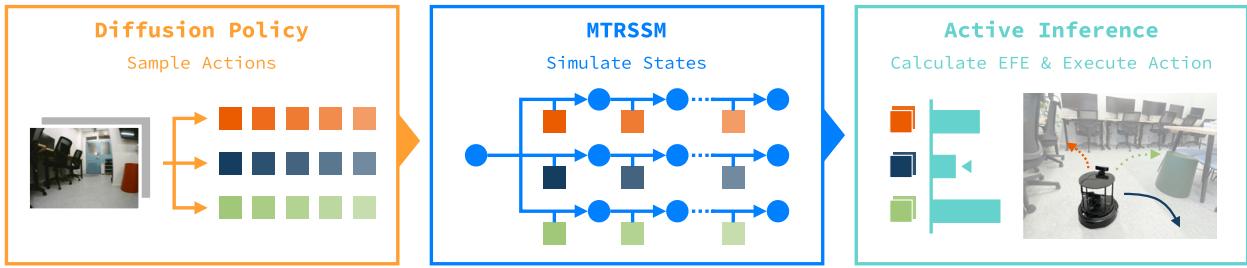


Fig. 1. Overview of the proposed deep active inference (AIF) framework. The framework integrates a diffusion policy and a multiple timescale recurrent state-space model (MTRSSM). The diffusion policy generates diverse candidate action sequences, and the MTRSSM predicts the resulting state transitions. The expected free energy (EFE) is evaluated for each candidate sequence, and the action with the lowest EFE is selected for execution in the real-world environment.

- We validate the proposed framework on a real mobile robot, showing that it achieves both exploration and goal-directed navigation in uncertain environments without relying on handcrafted planners.

The remainder of this paper is structured as follows. Section II reviews related work on learning-based navigation policies, world models, and AIF. Section III introduces the proposed methodology, including the formulation of EFE and its integration with a diffusion policy and an MTRSSM. Section IV describes the experimental setup. Section V reports the results. Section VI discusses the findings, and Section VII concludes the paper and outlines future directions.

II. RELATED WORK

Research on autonomous navigation has investigated numerous approaches, ranging from policy learning to world models and AIF frameworks. This section reviews the representative studies in these domains, with a particular focus on their applicability to real-world navigation as well as their limitations.

A. Policy Models

In recent years, policy learning approaches for navigation have increasingly leveraged powerful sequence models such as transformers [10]–[13], [32] and diffusion models [14]–[19], [27]. Transformer-based methods have demonstrated strong performance in both simulation [10], [11] and real-world environments [12], [13], often generating actions by conditioning on subgoals or return signals. Although these approaches can generalize to real-world scenarios with large-scale datasets, they typically rely on external planners for subgoal specification and are thus less suited for generating exploratory behaviors.

In contrast, diffusion-based policies have been applied to both exploration and navigation, demonstrating robust performance across diverse environments, including real-world settings [14]–[19]. A representative example is NoMaD, which employs diffusion policy to generate diverse action samples for both exploration and goal-directed navigation within a single policy model. Although this illustrates the flexibility of diffusion-based policies, NoMaD still requires additional components. Specifically, high-level planners are used to guide action selection during navigation, and the switching between

exploration and navigation is determined manually, depending on whether a goal image is provided. More broadly, balancing these two modes remains a fundamental challenge for diffusion-based methods.

In summary, policy learning for navigation has advanced through transformer- and diffusion-based approaches, but challenges remain in terms of achieving both exploratory and goal-directed behaviors without relying on manually designed modules. In the present work, we adopt a diffusion policy as the policy model, motivated by its ability to generate diverse behaviors that can support both exploration and goal-directed navigation within a single policy. By integrating this policy with AIF, we aim to address the challenge of balancing exploration and navigation in real-world environments.

B. World Models

World models have been studied as a means to capture environmental dynamics and predict future states without requiring direct interaction with the environment [29]–[31]. A representative example is the recurrent state-space model (RSSM), which combines deterministic and stochastic latent variables in order to learn compact dynamics.

World models can be leveraged in two main ways. First, they can facilitate policy learning, where imagined rollouts are used to train policies efficiently [33]–[35]. Second, they can support action planning, where imagined rollouts are utilized to evaluate candidate action sequences before execution [1], [36]–[38].

Regarding navigation tasks, world models have been studied for both policy learning [39]–[41] and action planning [42]–[44]. For policy learning, some methods incorporate semantic information [39] or contrastive representation learning [40]. For action planning, navigation world models employ conditional generative dynamics to imagine trajectories for multiple action candidates and evaluate them by comparing the imagined outcomes with goal observations [42]. Recent work has also developed generative models for autonomous driving, aiming to address large-scale real-world challenges [45], [46].

Nevertheless, world models face persistent limitations. Prediction errors accumulate over long horizons, making robust long-horizon imagination difficult, especially in partially observable real-world environments. To mitigate this, hierarchi-

cal extensions such as the MTRSSM [28] have been proposed, capturing dynamics at both fast and slow timescales in order to improve long-horizon prediction.

In summary, although world models provide versatility for both learning and planning, their deployment in real-world navigation remains challenged by error accumulation in long-horizon prediction. In the present work, we leverage a world model within the AIF framework, using it to provide the environmental state predictions required for EFE computation. To address long-horizon error accumulation, we adopt the MTRSSM, which captures temporal dependencies across multiple timescales, thereby improving long-horizon prediction.

C. Active Inference

AIF is a biologically inspired framework grounded in the free-energy principle, which explains learning, perception, and action in biological agents [20]–[23]. Under this framework, observations are assumed to be generated by hidden states of the environment, and the agent maintains a generative model to infer these hidden states from observations and select actions accordingly. While prediction errors on observations (termed “surprise”) are minimized through variational free energy (VFE), actions are selected to minimize EFE, which represents future uncertainty and goal-directed preferences.

By selecting actions that minimize EFE, the agent can effectively integrate both exploration and goal-directed behavior. In the context of mobile robot navigation, this property enables AIF to serve as a unified framework for handling both exploration and goal-directed navigation, enabling the robot to acquire additional information when necessary and to reach specified targets efficiently.

In machine learning and robotics, deep neural networks have been employed to implement probabilistic representations within AIF, including applications to mobile robot navigation [24]–[26], [47]–[49]. Previous research has tended to focus on either the epistemic (exploration-driven) value or the extrinsic (goal-directed) value in isolation. For instance, some studies have considered only the epistemic value needed to explore and build topological maps [24], [48], whereas others have emphasized the extrinsic value needed to achieve navigation objectives in simulation or with real robots [26], [47]. Additionally, hierarchical extensions of state-transition models have been incorporated into AIF [24]–[26], [48], [49].

In summary, despite these advances, most experiments remain confined to simulation, and few studies have simultaneously integrated exploration and navigation in real-world scenarios [48], [49]. Incorporating powerful policy and world models based on deep generative modeling into AIF might open the door to scaling AIF to more complex real-world environments. In the present study, we take a step in this direction by integrating policy and world models into AIF for real-world navigation tasks.

D. Summary

Recent advances in autonomous navigation span policy learning, world models, and AIF. Transformer- and diffusion-based policies have improved action generation, with diffusion

policies supporting both exploration and navigation. World models enable dynamics prediction for learning and planning, and hierarchical variants such as MTRSSM have improved long-horizon prediction. AIF provides a unified method of exploration and goal-directed behavior by minimizing EFE, but most studies remain limited to performing simulations and emphasize only one mode, limiting real-world applicability.

In contrast, our approach extends AIF to complex real-world navigation by (i) employing a diffusion policy to generate diverse candidate actions that support both exploration and goal-directed navigation, (ii) leveraging an MTRSSM to capture temporal dependencies and provide long-horizon state predictions for EFE computation, and (iii) integrating these components within the AIF framework for principled action selection. This combination enables efficient balancing of exploration and goal-directed navigation in uncertain environments.

III. METHODOLOGY

A. Overview

The proposed deep AIF framework enables autonomous navigation by integrating a diffusion policy as the policy model and the MTRSSM as the world model. The architecture of the framework is illustrated in Fig. 2. The diffusion policy generates diverse candidate action sequences conditioned on past observations, while the MTRSSM predicts the corresponding state transitions. For each candidate sequence, the EFE is calculated using latent imagination [36], and the action sequence with the lowest EFE is executed in the real-world environment.

B. Formulation of Active Inference

In the free-energy principle, perception and learning are formalized as the minimization of VFE. At time t , and given hidden states s_t and observations o_t under the generative model $p(o_t, s_t)$, the VFE \mathcal{F}_t serves as an upper bound on the surprise [20], [50]:

$$\begin{aligned} \mathcal{F}_t &= \mathbb{E}_{q(s_t)} [\log q(s_t) - \log p(o_t, s_t)] \\ &= D_{\text{KL}}[q(s_t) || p(s_t)] - \mathbb{E}_{q(s_t)} [\log p(o_t | s_t)] \\ &= D_{\text{KL}}[q(s_t) || p(s_t | o_t)] - \log p(o_t) \\ &\geq -\log p(o_t). \end{aligned} \quad (1)$$

In contrast, decision-making and action selection are driven by minimization of the EFE. For a future time step τ under policy π , the EFE is defined as follows [50]:

$$\mathcal{G}_\tau(\pi) = \mathbb{E}_{q(o_\tau, s_\tau | \pi)} [\log q(s_\tau | \pi) - \log p(o_\tau, s_\tau | \pi)]. \quad (2)$$

Note that although the VFE is computed after receiving an observation and therefore requires only the expectation over hidden states, the EFE considers future time steps before observations are available, and thus involves expectations over both hidden states and observations. Following standard

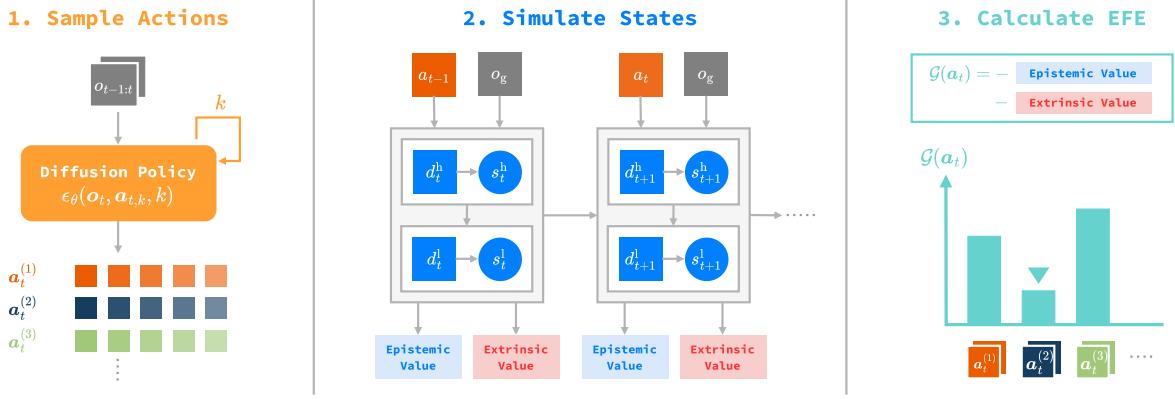


Fig. 2. Architecture of the proposed deep AIF framework. The process comprises three steps: (1) **Sample Actions**: the diffusion policy generates multiple candidate action sequences of length T_F , conditioned on past observations; (2) **Simulate States**: the MTRSSM performs latent imagination by simulating the state transitions for each candidate action sequence, sampling high- and low-level latent states in order to estimate epistemic and extrinsic values; (3) **Calculate EFE**: the EFE is calculated for each candidate action sequence, combining epistemic and extrinsic terms, and the action sequence with the lowest EFE is executed in the real-world environment.

derivations, and given preference C , the EFE can be decomposed into epistemic and extrinsic values as follows [50]:

$$\mathcal{G}_\tau(\pi) \approx - \underbrace{\mathbb{E}_{q(o_\tau|\pi)}[D_{\text{KL}}[q(s_\tau|o_\tau, \pi) || q(s_\tau|\pi)]]}_{\text{epistemic value}} - \underbrace{\mathbb{E}_{q(o_\tau|\pi)}[\log p(o_\tau|C)]}_{\text{extrinsic value}}. \quad (3)$$

C. Policy Model

To enable action selection under AIF, the policy model must represent a diverse set of candidate actions for a given situation. In this work, we adopt a diffusion policy [27] as the policy model. The diffusion policy models the conditional distribution of future action sequences \mathbf{a}_t based on past observation sequences \mathbf{o}_t , using a diffusion model [51]; that is, $p(\mathbf{a}_t|\mathbf{o}_t)$. Specifically, we define the action sequence as consisting of the two most recent actions followed by T_F future steps, and the observation sequence as the two most recent observations:

$$\begin{cases} \mathbf{a}_t = a_{t-1:t+T_F} \\ \mathbf{o}_t = o_{t-1:t} \end{cases}. \quad (4)$$

Diffusion models are generative models that are trained to iteratively denoise data corrupted by Gaussian noise. During training, clean data samples are perturbed with noise, and the model is optimized to predict this noise. After training, new data can be generated by starting from a noise sample and progressively denoising it through multiple steps.

During training, Gaussian noise is added to the ground-truth action sequence \mathbf{a}_t , and the network ϵ_θ is optimized to predict the added noise via the following objective:

$$\mathcal{L}_{\text{DP}}(\theta) = \text{MSE}(\epsilon_k, \epsilon_\theta(\mathbf{o}_t, \mathbf{a}_{t,k}, k)), \quad (5)$$

$$\mathbf{a}_{t,k} = \sqrt{\bar{\alpha}_k} \mathbf{a}_t + \sqrt{1 - \bar{\alpha}_k} \epsilon_k, \quad (6)$$

where k is the diffusion step, $\epsilon_k \sim \mathcal{N}(0, I)$ is Gaussian noise, and $\bar{\alpha}_k = \prod_{s=1}^k \alpha_s$ denotes the cumulative product of the noise-scheduling coefficients.

At the time of inference, candidate action sequences are generated by iteratively denoising a Gaussian noise sample \mathbf{a}_t^K through K reverse-diffusion steps, yielding a fully denoised sequence \mathbf{a}_t^0 . The reverse diffusion at step k is defined as:

$$\mathbf{a}_{t,k-1} = \frac{1}{\sqrt{\alpha_k}} \left(\mathbf{a}_{t,k} - \frac{1 - \alpha_k}{\sqrt{1 - \bar{\alpha}_k}} \epsilon_\theta(\mathbf{o}_t, \mathbf{a}_{t,k}, k) \right) + \epsilon_k, \quad (7)$$

where $\epsilon_k \sim \mathcal{N}(0, \sigma_k^2 I)$ is Gaussian noise and σ_k denotes the standard deviation at step k .

At deployment, only the first T_a steps of the generated sequence \mathbf{a}_t are executed, while $T_F > T_a$ facilitates planning further ahead.

D. World Model

To predict the action-conditioned state transitions, we used an MTRSSM [28], which extends the standard RSSM [30] by incorporating temporal hierarchies. This hierarchical design enables the model to capture both fast and slow dynamics in the environment.

Let d_t^h and d_t^l denote deterministic states at higher and lower levels, respectively, and s_t^h and s_t^l denote stochastic states at each level. The overall latent state becomes

$$z_t = \{z_t^h, z_t^l\}, \quad z_t^h = \{d_t^h, s_t^h\}, \quad z_t^l = \{d_t^l, s_t^l\}. \quad (8)$$

The deterministic states evolve according to separate recurrent functions at each timescale as follows:

$$d_t^h = f_\phi^h(d_{t-1}^h, s_{t-1}^h; \tau^h), \quad (9)$$

$$d_t^l = f_\phi^l(z_{t-1}^l, s_t^h, a_{t-1}; \tau^l), \quad (10)$$

where f_ϕ^h and f_ϕ^l are implemented by multiple timescale recurrent neural networks (MTRNNs) [52]–[54] for the higher and lower levels, respectively. The higher level d_t^h with a larger time constant τ^h updates slowly in order to capture long-horizon dependencies, while the lower level d_t^l with a smaller time constant τ^l updates rapidly in order to encode short-term transitions.

The stochastic states are sampled from either the prior distributions p_ϕ^h, p_ϕ^l or the approximate posterior distributions q_ϕ^h, q_ϕ^l , defined as follows:

$$\hat{s}_t^h \sim p_\phi^h(s_t^h | d_t^h), \quad \hat{s}_t^l \sim p_\phi^l(s_t^l | d_t^l), \quad (11)$$

$$s_t^h \sim q_\phi^h(s_t^h | d_t^h, d_t^l), \quad s_t^l \sim q_\phi^l(s_t^l | d_t^l, o_t). \quad (12)$$

Observations o_t are encoded into low-dimensional features via a convolutional neural network (CNN) encoder before being fed into the model. The MTRSSM is trained by minimizing the following VFE-based loss:

$$\begin{aligned} \mathcal{L}_{WM} = \sum_{t=1}^T \Big\{ & \beta D_{KL}[q_\phi^l(s_t^l | d_t^l, o_t) \| p_\phi^l(s_t^l | d_t^l)] \\ & + \beta D_{KL}[q_\phi^h(s_t^h | d_t^h, d_t^l) \| p_\phi^h(s_t^h | d_t^h)] \\ & - \mathbb{E}_{q_\phi(s_t^l | d_t^l, o_t)} [\log p_\phi^l(o_t | z_t^h, z_t^l)] \\ & - \mathbb{E}_{q_\phi(s_t^h | d_t^h, d_t^l)} [\log p_\phi^h(d_t^h | z_{t-1}^h)] \Big\}. \end{aligned} \quad (13)$$

This loss combines the KL regularization at both hierarchies with the reconstruction losses of observations and dynamics, making it possible for the model to learn both short- and long-horizon dependencies.

E. Active Inference with EFE Minimization

At the time of inference, the policy model generates multiple candidate action sequences \mathbf{a}_t , each of which is simulated using the MTRSSM through latent imagination [36]. Future states \hat{s}_τ^l are sampled from the prior distribution, and the corresponding predicted observations \hat{o}_τ are decoded. At each time step, M high-level latent states s_τ^h are sampled from the posterior distribution of the higher level, and for each one, N low-level latent states s_τ^l are sampled from the prior of the lower level. This results in $M \times N$ predicted observations $\hat{o}_\tau^{i,j}$. Using these predicted observations together with the candidate action and the current state, the posterior of the lower level is recomputed at each time step and propagated forward through the MTRSSM. The KL divergence between this posterior and the corresponding prior constitutes the epistemic value term. The EFE for a candidate action sequence is approximated as follows:

$$\begin{aligned} \mathcal{G}_\tau(\mathbf{a}_t) \approx -\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \Bigg\{ & D_{KL}\left[q_\phi^l(s_\tau^l | d_\tau^l, \hat{o}_\tau^{i,j}) \| p_\phi^l(s_\tau^l | d_\tau^l)\right] \\ & + \frac{1}{\sigma_\tau^2} \text{MSE}(f(\hat{o}_\tau^{i,j}), f(o_g)) \Bigg\}, \end{aligned} \quad (14)$$

where $f(\cdot)$ denotes a CNN encoder that maps observations onto a feature space. The first term corresponds to the epistemic value, and the second term corresponds to the extrinsic value in (3). Details of the EFE computation are illustrated in Fig. 3.

In this work, the epistemic value is computed only at the lower level of the MTRSSM, while the extrinsic value is defined as the temporal average of a feature-space distance between the predicted (imagined) observations \hat{o}_τ and the goal

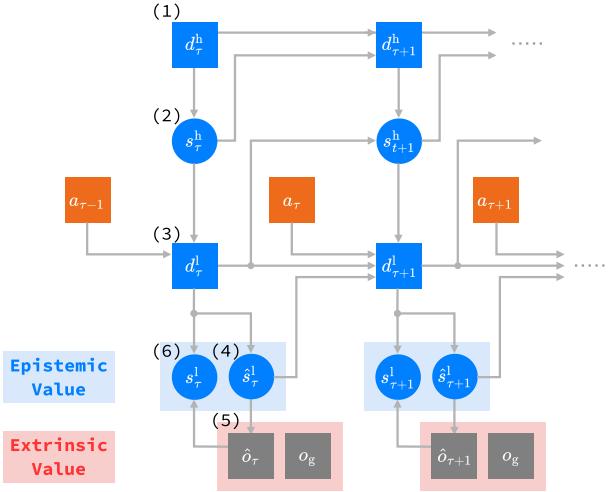


Fig. 3. Computation of EFE for candidate action sequences. Each candidate action sequence \mathbf{a}_t is simulated by the MTRSSM, using latent imagination. At each time step, the process unfolds as follows: (1) the higher-level deterministic state d_t^h is updated; (2) the higher-level stochastic state s_t^h is sampled; (3) the lower-level deterministic state d_t^l is updated; (4) the lower-level prior $q_\phi(s_t^l | d_t^l)$ is predicted and the stochastic state s_t^l is sampled; (5) a predicted observation \hat{o}_τ is generated; and (6) the lower-level stochastic posterior $q_\phi(s_t^l | d_t^l, \hat{o}_\tau)$ is inferred using \hat{o}_τ . At the lower level, the epistemic value is computed as the KL divergence between the posterior and the prior, while the extrinsic value is computed as the feature-space distance between the predicted observation \hat{o}_τ and the goal observation o_g . Combining these two terms yields the EFE $\mathcal{G}_\tau(\mathbf{a}_t)$, which is used to select the action sequence that balances exploration and goal-directed navigation.

observations o_g . To balance these two terms, the precision (inverse variance) $1/\sigma_\tau^2$ is designed as a time-decaying coefficient, where the epistemic value dominates in the earlier phase when self-localization uncertainty is high, while the extrinsic value gradually becomes dominant in the later phase once the robot has localized itself.

Finally, an action sequence \mathbf{a}_t^* is selected according to the following equation:

$$\mathbf{a}_t^* = \arg \min_{\mathbf{a}_t} \mathcal{G}_\tau(\mathbf{a}_t). \quad (15)$$

This action-selection mechanism based on the EFE minimization enables the robot to balance exploration driven by epistemic value with goal-directed navigation guided by extrinsic value.

IV. EXPERIMENTS

A. Hardware Setup

The proposed deep AIF framework was implemented with a TurtleBot 4 (Clearpath Robotics), as shown in Fig. 1. The robot is controlled via two velocity commands: linear velocity and angular velocity. A wide-angle RGB camera (CMS-V43BK, Sanwa Supply) was mounted on the top plate of the robot. The captured images were resized to 240×320 pixels and used as observations.

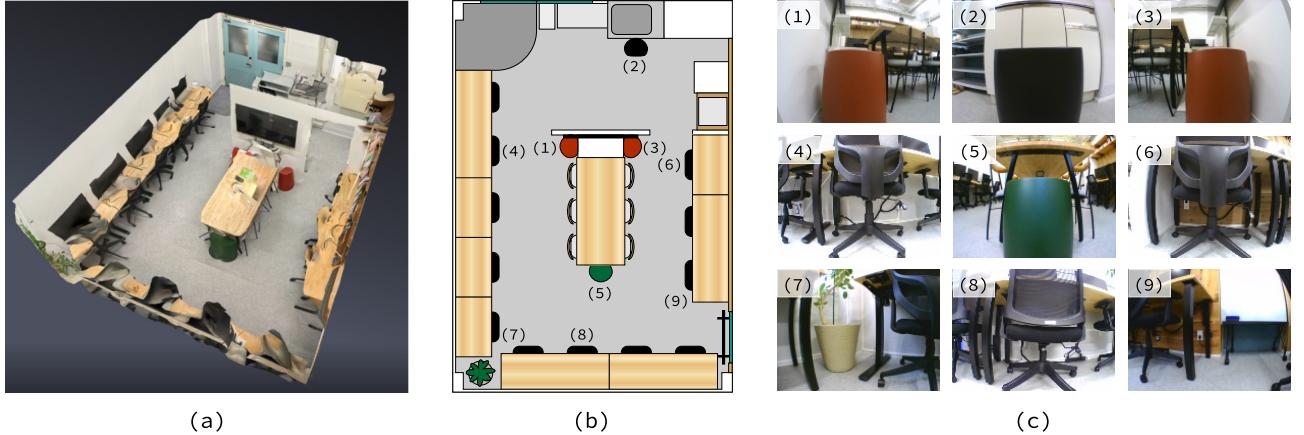


Fig. 4. Experimental environment. (a) Overhead view of the indoor room. (b) Top-down map of the environment. (c) Representative observations at nine designated location-orientation patterns (1)–(9) on the map. In the real-world navigation tasks, the initial position-orientation pairs and goal images were selected from among these patterns, resulting in 18 experimental cases.

B. Experimental Environment

The experiments were conducted in an indoor room (approximately $4.7 \text{ m} \times 6.3 \text{ m}$), the layout and representative observation images of which are shown in Fig. 4. Three sides of the room (excluding the entrance wall) were lined with desks and black desk chairs, and the center contained a meeting table surrounded by dining chairs. Additionally, four colored stools were placed as landmarks: two red, one black, and one green. A key challenge is that, because of the similarity of images across different locations, especially along the wall side with black desk chairs, localization based solely on the current observation can be ambiguous. Therefore, the robot must actively explore the environment to reduce uncertainty and realize goal-directed navigation.

C. Data Collection

Data were collected by manually teleoperating the robot within the environment. In total, 15 sequences of 2,000 time steps each were recorded at 5 Hz, for a total of 30,000 time steps. Each dataset contained velocity commands and RGB observation images.

D. Dataset for Policy Model Training

For policy model training, the observation images were resized to 120×160 pixels. Each sequence was segmented into 128-step windows every 32 steps, yielding 885 subsequences. From each, 64-step subsequences were randomly sampled to ensure that all observation images would appear as conditions during training.

E. Dataset for World Model Training

For world model training, the observation images were downsampled to 60×80 pixels. Each sequence was segmented into 600-step trajectories every 100 steps, resulting in 225 trajectories. From each trajectory, 500-step subsequences were randomly extracted.

F. Implementation Details

1) Training Details:

a) Diffusion policy: The observation encoder was composed of a 3-layer CNN followed by a spatial softmax, producing 32 keypoints that conditioned the diffusion model. The CNN used convolutional layers with channels (8, 16, 32), kernel sizes (3, 3, 3), strides (2, 2, 1), and paddings (1, 1, 1). The diffusion model was implemented as a 1D-UNet [55] with a DDIM sampler [56]. The action sequence length was set to $T_F = 64$, and the number of diffusion steps was set to $K = 100$.

b) MTRSSM: The observation encoder consisted of a 3-layer CNN with channel sizes (16, 32, 64), kernel sizes (3, 3, 3), strides (1, 2, 2), and paddings (1, 1, 1), producing a 128-dimensional embedding. The deterministic states of the MTRSSM were modeled using an MTRNN [52]–[54] with both higher-level deterministic states $d_t^h \in \mathbb{R}^{32}$ and lower-level deterministic states $d_t^l \in \mathbb{R}^{128}$. The time constants were set to $\tau^h = 64$ for the higher level and $\tau^l = 4$ for the lower level. The higher-level stochastic states s_t^h were represented by a categorical distribution over 4×4 classes, while the lower-level stochastic states s_t^l were represented by a categorical distribution over 8×8 classes. The image decoder was implemented as a CNN with eight residual blocks and two pixel-shuffle layers [57], [58]. Training employed truncated backpropagation through time [59] with a window size of 50 steps.

2) Experiment Details:

a) Diffusion Policy: At test time, the policy generated action sequences of $T_F = 64$ steps but executed only the first $T_a = 32$ steps. The number of diffusion steps was reduced to $K = 10$ for real-time execution, and eight candidate action sequences were sampled per inference.

b) MTRSSM: For EFE computation defined in (14), we sampled $M = 5$ higher-level latent states and $N = 5$ lower-level latent states at each time step. The precision for the extrinsic value in (14) was formulated as a smooth, sigmoidal function of the inference iteration n , which was incremented

once every T_a time steps. Specifically, it was defined as follows:

$$\frac{1}{\sigma_\tau^2} = 0.08 + \frac{3.0 - 0.08}{1 + \exp[-0.6(n - 10)]}. \quad (16)$$

This formulation gradually shifts the weighting between epistemic and extrinsic values, with the epistemic term dominating in the early phase, and the extrinsic term becoming increasingly influential as the inference iteration progresses. The feature encoder $f(\cdot)$ comprised a three-layer CNN with channel sizes (16, 32, 64), kernel sizes (3, 3, 3), strides (2, 2, 2), and paddings (1, 1, 1), followed by a three-layer fully connected network with 1,024 hidden units. The feature encoder was trained so that the feature-space distance between two observations reflects their spatial distance, as estimated from accumulated actions or odometry.

G. Navigation Task with Real-World Robot

The main evaluation was conducted using robot navigation experiments in the environment described above. Three initial positions were prepared, and each was tested under the following two facing directions: (1) facing toward the interior of the room, where the current location is visually distinct (2) facing the wall side, where the similarity of chairs along the three walls introduces perceptual aliasing, making localization highly uncertain. Accordingly, six distinct initial states were defined in total.

For each initial position, three different goal images were specified, yielding $6 \times 3 = 18$ task cases. Each case was executed twice, for a total of 36 trials. A trial was terminated after 1,000 time steps if the robot had not reached the goal. Success was defined as entering a predefined goal area (manually determined to correspond to the region where the goal image was observable). In addition to the success rate, the number of collisions with obstacles was also recorded.

H. Baseline Methods

We consider two baselines, which also served as ablations for our framework.

The first baseline replaces the MTRSSM with a standard RSSM. Although RSSM has been widely used in model-based reinforcement learning to capture short-term dynamics, it lacks hierarchical temporal structure, making long-horizon prediction more prone to error.

The second baseline, Only Extrinsic, minimizes only the extrinsic component of the EFE, ignoring the epistemic value. This corresponds to purely goal-directed navigation without exploration, similar to conventional planning- or reward-driven methods.

By including these ablations as baselines, we can directly assess the contributions of multiple timescale modeling and epistemic value to overall performance.

V. RESULTS

This section presents the experimental results of the proposed framework. We begin by evaluating the policy model, examining whether it is capable of generating diverse and

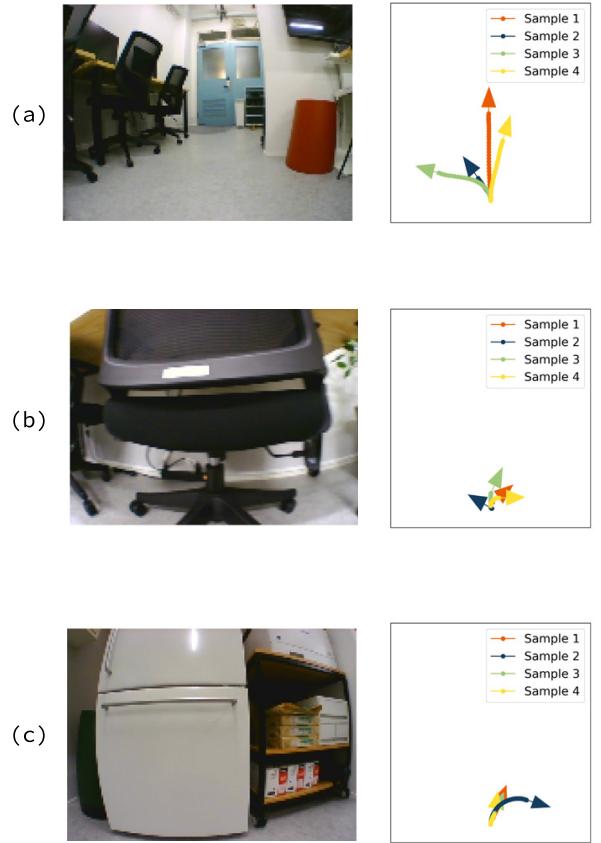


Fig. 5. Representative action sequences generated by the diffusion policy in three scenarios: (a) clear path, (b) obstacle ahead, and (c) approaching a corner. The policy adapts its action proposals to each situation and generates diverse behaviors such as forward movement and turns, illustrating its flexibility in handling different environmental contexts.

context-dependent action sequences. We then assess the predictive capability of the world model through imagination experiments. Finally, we demonstrate the overall navigation performance on a real mobile robot, comparing our framework with baseline methods.

A. Policy Model Evaluation

We first evaluated the diffusion policy to assess its ability to generate diverse and context-dependent action sequences. Fig. 5 shows representative cases illustrating how the policy responds under different environmental conditions.

Fig. 5(a) shows the case of a clear path ahead, where the corridor in front of the robot was unobstructed. The policy generated a wide variety of forward-directed actions, including driving straight ahead or turning left/right.

Fig. 5(b) illustrates a case in which a black desk chair was positioned directly in front of the robot. The generated action sequences included turns to both the left and right, as well as slight forward adjustments to navigate around the obstacle. This indicates that the policy successfully incorporated the perceived obstacle into its proposals.

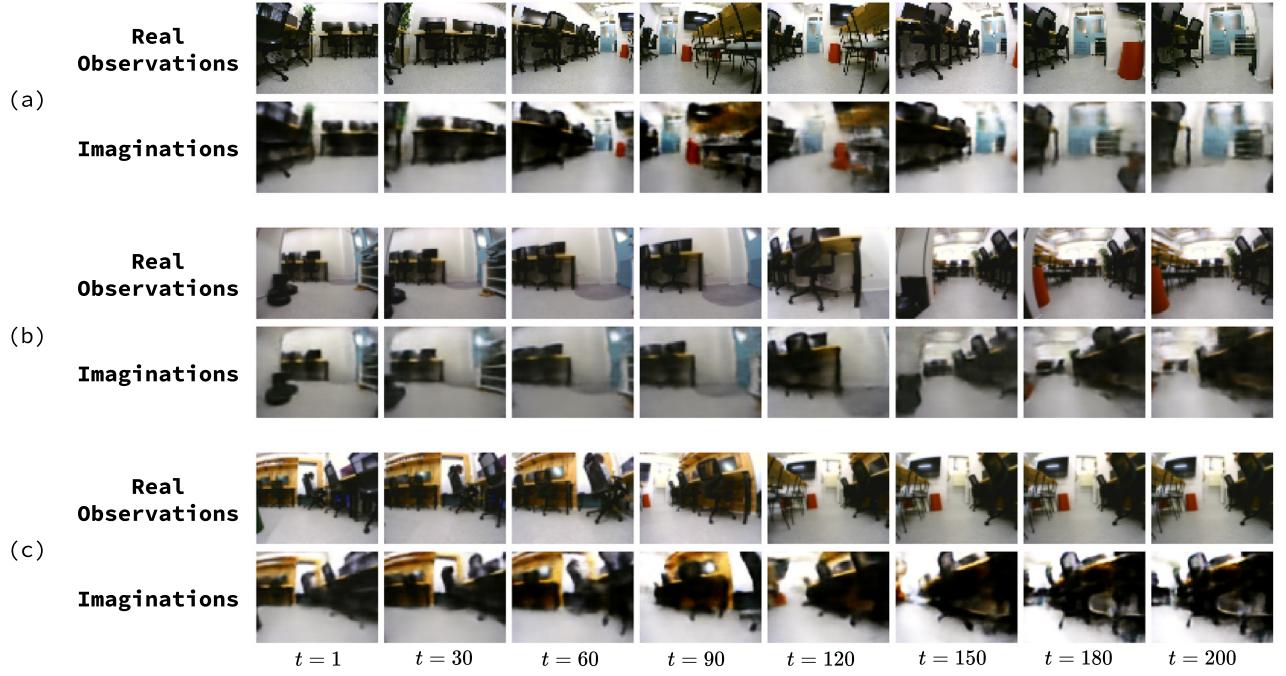


Fig. 6. Examples of real observations and imaginations generated by the MTRSSM. (a, b) Successful cases. (c) Failure case. In the successful cases, the imaginations captured the overall scene dynamics, although subtle temporal lags appeared after time step $t = 180$. In the failure case, temporal lags occurred after time step $t = 120$, followed by a spatial discontinuity at time step $t = 180$. Despite these deviations, the imagined sequence remained coherent for more than 100 steps, demonstrating the model's long-horizon prediction capability.

Fig. 5(c) presents the case of approaching a room corner. The policy produced turning actions that aligned with the room's geometry, suggesting that it can propose adaptive trajectories consistent with the layout of the environment.

Overall, these examples demonstrate that the diffusion policy is capable of flexibly generating diverse actions suited to the current situation. While this evaluation highlights adaptability to different situations, the role of exploratory and goal-directed behaviors will be further examined in the subsequent real-world experiments.

B. World Model Evaluation

Next, we evaluated the MTRSSM, examining its predictive capability in long-horizon imagination. The model received observation inputs for the first 20 time steps and then generated predictions for 200 steps without further observations.

Figs. 6(a) and 6(b) show successful imaginations in which the imaginations closely matched the real observations. Fig. 6(c) shows a case in which the imagination diverged from the actual observations, generating a wall-side black desk chair instead of a corridor with a red stool. Nevertheless, the imagined sequence remained coherent for more than 100 time steps, demonstrating the model's ability to maintain long-horizon predictions.

Further analyzing these results, Fig. 7 illustrates the internal state dynamics during a loop around the room. Fig. 7(a) shows the higher-level deterministic states d_t^h , Fig. 7(b) shows the lower-level deterministic states d_t^l , and Fig. 7(c) shows the image features. For each panel, the trajectories were obtained

by collecting the corresponding states or features across the entire training dataset and projecting them into three dimensions, using principal component analysis. In the visualization, all data are shown in gray, and the trajectory corresponding to the room loop is highlighted with a time-encoded colormap. Only the higher-level states d_t^h formed a smooth closed loop, which represents the global room structure and acts as an attractor in the state space. This indicates that the higher-level dynamics of the MTRSSM successfully captured the environmental geometry, enabling consistent imagination even in the absence of observation inputs.

Overall, these results demonstrate that the MTRSSM can leverage its hierarchical structure to generate coherent long-horizon predictions. This predictive capability is essential for evaluating candidate action sequences in AIF and will be combined with the diffusion policy in the subsequent real-world experiments.

C. Real-World Experiments

Finally, we evaluated the proposed deep AIF framework in real-world navigation tasks, using the setup described in Sec. IV-G. In total, 36 trials were conducted.

Table I summarizes the quantitative results in terms of success rate and collisions. Overall, our framework achieved a success rate of 75%, outperforming RSSM (64%) and Only Extrinsic (53%). The improvement was most pronounced in exploration-demanding scenarios, where our framework reached 78% compared with 61% for RSSM and only 28% for Only Extrinsic. In the no-exploration cases, all methods

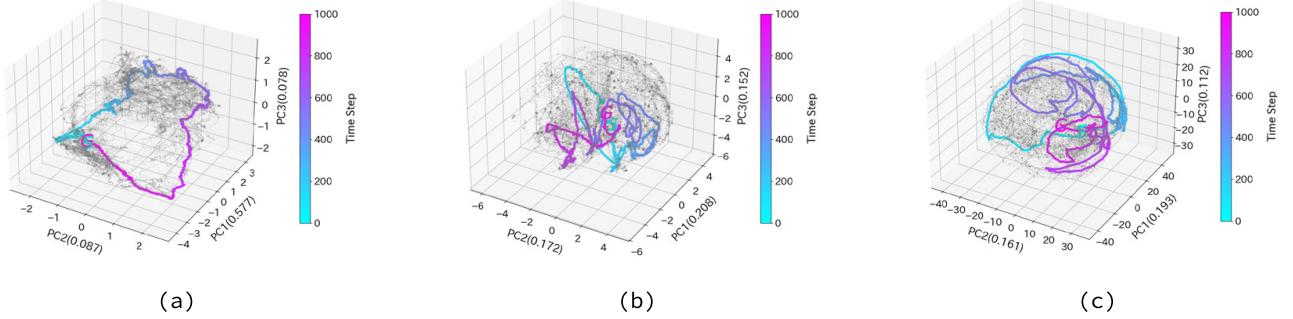


Fig. 7. Internal state dynamics during a loop around the room. (a) Higher-level deterministic states d_t^h . (b) Lower-level deterministic states d_t^l . (c) Image features. For each panel, trajectories were obtained by collecting the corresponding states or features across the entire training dataset and projecting them into three dimensions using principal component analysis. Gray points represent all data from the training set, and the trajectory corresponding to the room loop is highlighted with a color gradient from light blue to pink, indicating temporal progression. Only the higher-level states d_t^h formed a smooth closed loop, which reflected the global room structure and demonstrated that the higher-level dynamics of the MTRSSM successfully captured the environmental geometry, enabling consistent imagination even in the absence of observation inputs.

TABLE I
NAVIGATION SUCCESS RATES (%) AND AVERAGE COLLISIONS FOR
OVERALL, EXPLORATION (EXP), AND NON-EXPLORATION (NOEXP)
TRIALS.

Method	Success Rate (%)			Collisions		
	Overall	Exp	NoExp	Overall	Exp	NoExp
Ours	75	78	72	0.806	0.778	0.833
RSSM	64	61	67	0.806	1.056	0.556
Only Extrinsic	53	28	78	1.000	1.667	0.333

performed relatively well, with Only Extrinsic achieving the highest rate (78%) and our framework achieving 72%. For collisions, our framework matched RSSM overall (0.806) and clearly reduced collisions compared with Only Extrinsic (1.000), particularly under exploration-demanding scenarios (0.778 vs. 1.667).

These quantitative results highlight two key factors underlying the effectiveness of our framework. First, the use of the MTRSSM contributed to improved success rates by enabling more accurate long-horizon predictions compared with a single-layer RSSM. Second, the incorporation of epistemic value in the EFE formulation enabled the robot to actively explore and resolve uncertainty, which was especially beneficial in exploration-demanding scenarios. These contributions are further illustrated by the following qualitative examples, which show how epistemic and extrinsic values guide action selection in practice.

Fig. 8 and 9 provide qualitative examples of action selection. Here, we focus on a trial whose initial and goal observations correspond to images (4) and (5) in Fig. 4, respectively. Fig. 8 shows the early stage of the trial, whereas Fig. 9 shows the later stage near the goal. These examples illustrate how the robot selected actions at different phases of navigation in the trial.

In Fig. 8, the robot initially faced a black desk chair, and because similar chairs were placed throughout the environment, the current observation alone was insufficient for reliable self-

localization. Candidate actions included staying in place as well as turning in place. The imagined observations at the last time step revealed that the turning action (Sample 1 in the figure) would expose additional information beyond the line of chairs along the wall. Our framework selected this turning action because the EFE assigned it a high epistemic value, reflecting the potential information gain. This illustrates how epistemic considerations guide the robot toward exploratory behavior, enabling it to reduce uncertainty and improve subsequent localization.

In contrast to the early stage shown in Fig. 8, Fig. 9 illustrates the later stage of the same trial, when the robot was already near the goal. In the current observation, a green stool—also present in the goal observation—appears on the left side, providing a clear visual cue. At this point, some candidate actions would continue past the goal, whereas others would turn toward and approach it. The imagined observations for the goal-approaching actions (Samples 1 and 3) closely matched the goal observation. In this case, the EFE was dominated by the extrinsic value, favoring actions that brought the robot closer to the goal.

Finally, Fig. 10 compares our framework with the Only Extrinsic baseline in the same scenario that required exploration. Note that the first situation ($t = 10$) in our framework corresponds to the example shown in Fig. 8. The figure presents the sampled candidate actions together with their EFE values. The action with the lowest EFE was selected and executed, leading to a change in the robot's observations. The upper row (our framework) demonstrates active exploration, where the robot obtained increasingly diverse observations over time, whereas the lower row (Only Extrinsic) shows the robot largely remaining in place without gaining new information. These results indicate that incorporating epistemic value into the EFE formulation encouraged exploratory actions that produced new observations, thereby enabling the robot to resolve uncertainty and achieve self-localization, whereas the baseline relying solely on extrinsic value failed to do so.

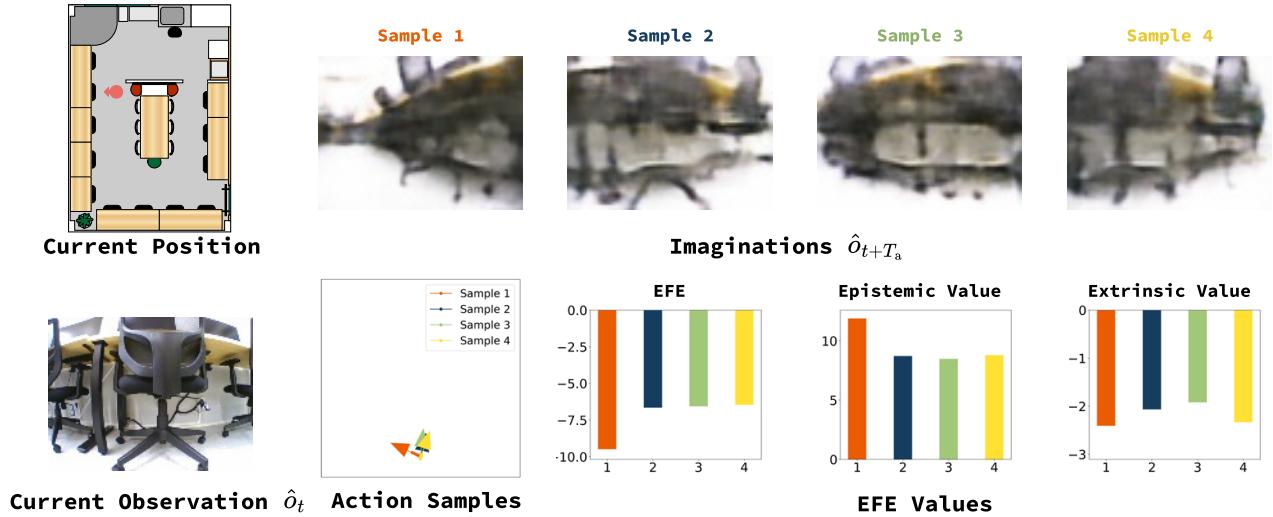


Fig. 8. Action selection during early-stage navigation. The robot evaluates candidate actions, such as staying in place or turning, while the current observation—showing a black desk chair that appears at multiple locations—is insufficient for precise localization. Imagined observations corresponding to each action suggest that turning reveals additional environmental information. In this situation, the EFE is dominated by the epistemic value, leading the robot to select the turning action that reduces uncertainty and improves localization.

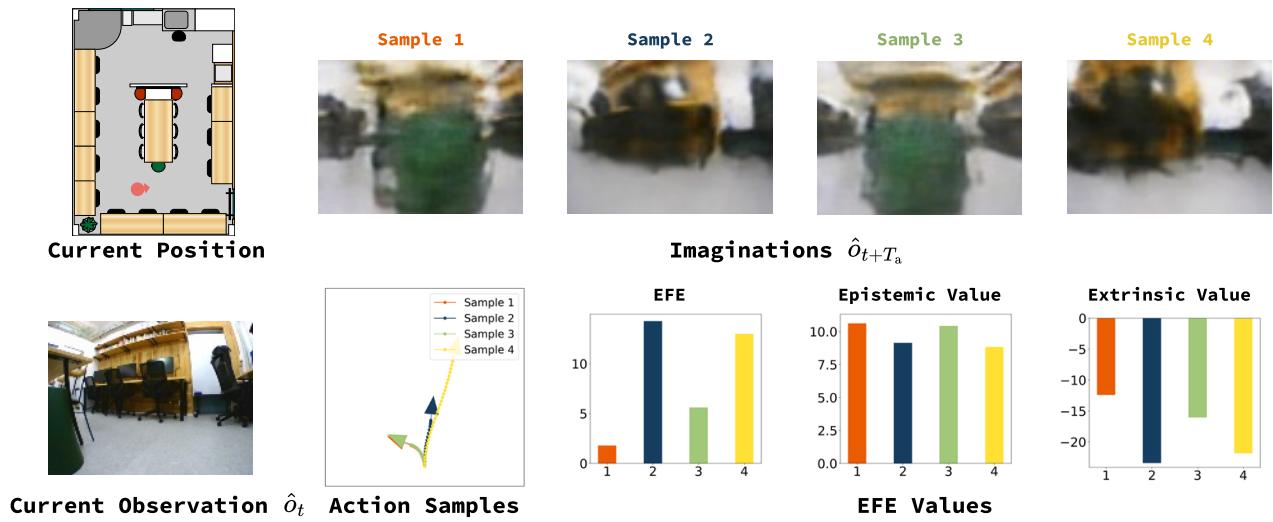


Fig. 9. Action selection near the goal. The robot evaluates candidate actions, some of which would pass the goal while others approach it. Imagined observations corresponding to goal-approaching actions closely match the goal image. In this situation, the EFE is dominated by the extrinsic value, leading the robot to select actions that move it closer to the goal.

VI. DISCUSSION

The central finding of this study is that AIF can effectively unify exploration and goal-directed navigation in a real-world setting by minimizing EFE. Although previous research has tended to emphasize either the epistemic value to drive exploration [24], [48] or the extrinsic value to pursue navigation objectives [26], [47], few studies have simultaneously integrated both aspects in physical robots [49]. As shown in Table I, our framework achieved higher success rates, particularly in exploration-demanding tasks, clearly indicating the benefit of

incorporating the epistemic value. This finding highlights the potential of AIF as a principled framework for navigation under partial observability and perceptual aliasing.

A key contribution of this work lies in its integration of a diffusion policy into the AIF framework. Prior studies have shown that diffusion policies are capable of generating diverse and adaptive action sequences in navigation tasks [14]–[19]. However, they often rely on additional modules such as high-level planners or manual switching mechanisms to balance exploration and navigation. For example, NoMaD [14] demonstrated the flexibility of diffusion-based action generation but

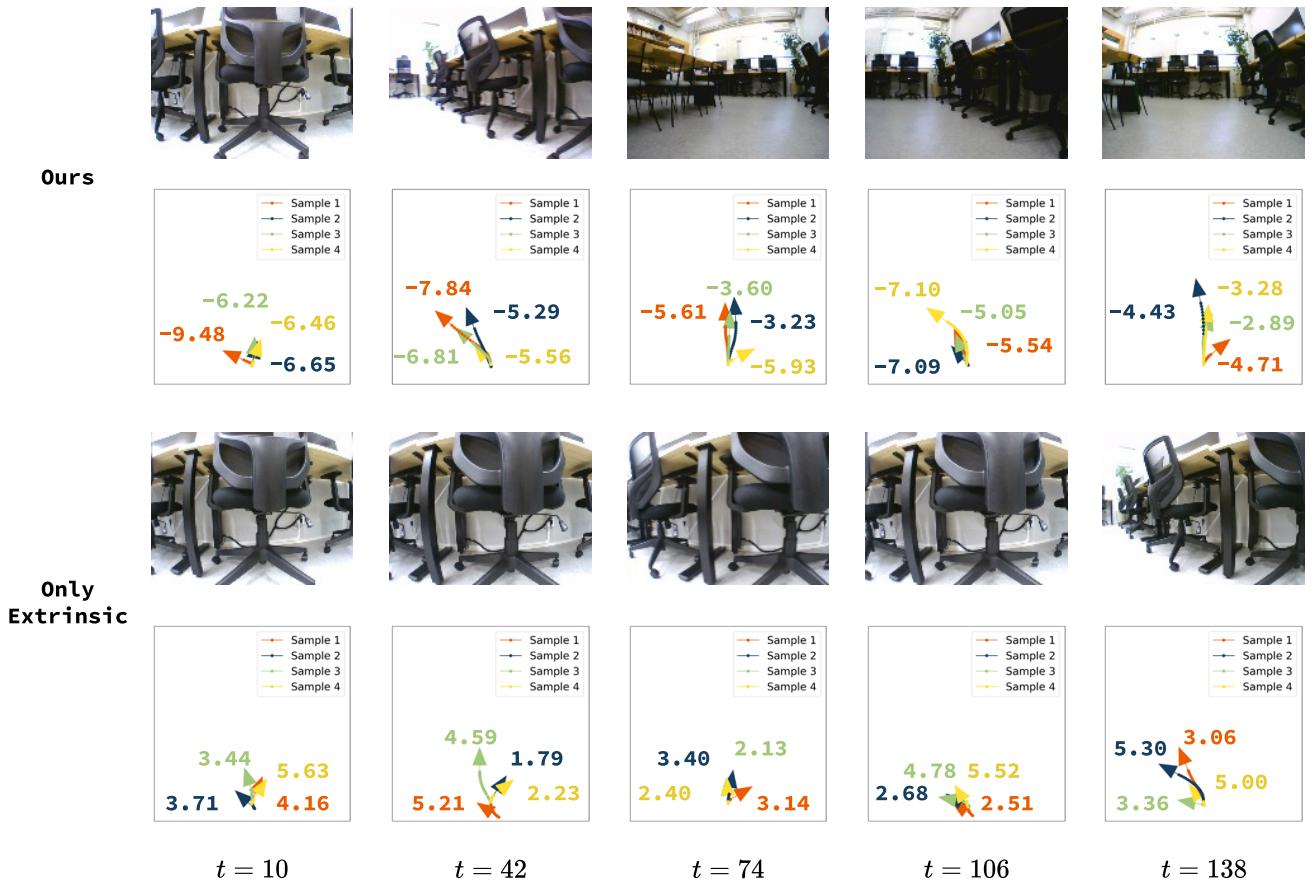


Fig. 10. Comparison of action selection between our deep AIF framework (upper row) and an Only Extrinsic baseline (lower row) in an exploration scenario. Our framework demonstrates active exploration, acquiring diverse observations over time, whereas the baseline largely remains stationary. The first action in our framework corresponds to the example in Fig. 8, highlighting the role of the epistemic value in driving exploratory behavior.

still required external guidance to select between exploratory and goal-directed behaviors. In contrast, our formulation embeds the diffusion policy within the AIF framework, enabling exploratory trajectories to be considered naturally alongside goal-directed ones. This effect is visible in Fig. 8, where the epistemic value drove the robot to turn and expose new observations, thereby reducing localization uncertainty.

Equally important is the role of the MTRSSM. Recurrent world models such as RSSM [30] have been widely adopted for imagination-based planning, yet they often suffer from error accumulation in long-horizon predictions. As shown in Fig. 7, only the higher-level deterministic states of the MTRSSM formed smooth closed loops, reflecting the global room structure. This indicates that MTRSSM successfully captured slow-varying dynamics, enabling robust long-horizon imagination. Such stability is essential for reliable estimation of epistemic and extrinsic values.

The integration of diffusion policy and MTRSSM within AIF reveals a synergistic effect. Although diffusion policy provides a broad set of candidate actions, MTRSSM supplies reliable long-horizon predictions that enable these candidates to be meaningfully evaluated. The EFE formulation then acts

as a unifying criterion selecting actions that balance epistemic and extrinsic considerations. This synergy was most apparent in uncertain initial states, as highlighted in Table I, where our framework outperformed the Only Extrinsic baseline by actively selecting exploratory actions. Similarly, Fig. 9 illustrates how extrinsic value dominates once the robot approaches the goal, ensuring efficient convergence.

From a broader perspective, the proposed framework contributes to extending the scalability of AIF in real-world robotics. Previous applications of AIF in navigation have largely been confined to simulation environments [24]–[26]. Even when deployed on real robots, these studies often addressed simplified tasks or considered exploration and navigation in isolation [48], [49]. By leveraging advances in generative modeling, the present work demonstrates that AIF can be scaled to more complex and uncertain real-world scenarios. With its ability to flexibly generate actions, perform stable long-horizon imagination, and balance epistemic and extrinsic values, AIF represents a competitive alternative to traditional methods. In particular, our framework provides a unified mechanism that does not require explicit mapping as in SLAM-based navigation [7]–[9] or handcrafted planners [12],

[14].

Nevertheless, several limitations must be acknowledged. First, the experimental environment was limited to a single indoor room, which limited the diversity of the conditions tested. Second, although MTRSSM reduced long-horizon prediction errors compared with RSSM, deviations such as spatial discontinuity still occurred (Fig. 6(c)). Third, the precision for the extrinsic term was heuristically scheduled as in (16) for numerical stability, limiting the theoretical completeness of the active inference formulation.

In summary, this discussion highlights how the integration of diffusion policy and MTRSSM within AIF enables principled action selection that balances exploration and goal-directed navigation. Our findings demonstrate that recent advances in deep generative modeling can be leveraged to address the long-standing challenge of scaling AIF to complex robotic tasks. By situating our contributions in relation to prior work, and by referencing the empirical evidence presented in Figs. 5, 6, 8, 9, and Table I, we emphasize both the theoretical significance of unifying epistemic and extrinsic values under EFE and the technical advances that made this integration feasible.

VII. CONCLUSION

In this paper, we proposed a deep AIF framework for real-world navigation that incorporates a diffusion policy as the policy model and an MTRSSM as the world model. The diffusion policy enabled the generation of diverse and context-dependent action candidates, while the MTRSSM provided stable long-horizon predictions that preserved the environmental structure. The integration of the policy and world models within the EFE formulation enabled the epistemic and extrinsic values to be exploited effectively, resulting in improved navigation performance compared with baseline methods. These findings demonstrate that recent advances in deep generative modeling can substantially enhance the scalability of AIF in real-world robotic systems.

Looking ahead, we identify three promising directions for future research. First, incorporating natural language instructions as goal specifications might enable more flexible and intuitive navigation. Second, leveraging pretrained foundation models might facilitate adaptation to entirely unseen environments. Finally, extending the framework to manipulation and integrating it with navigation might open the way for deep AIF-based mobile manipulation in real-world robotics.

REFERENCES

- [1] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, “Planning and acting in partially observable stochastic domains,” *Artificial Intelligence*, vol. 101, no. 1, pp. 99–134, 1998. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S000437029800023X>
- [2] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.
- [3] S. Levine and D. Shah, “Learning robotic navigation from experience: principles, methods and recent results,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 378, no. 1869, Dec. 2022. [Online]. Available: <http://dx.doi.org/10.1098/rstb.2021.0447>
- [4] M. Nowakowski, C. Joly, S. Dalibard, N. Garcia, and F. Moutarde, “Topological localization using wi-fi and vision merged into fabmap framework,” 09 2017, pp. 3339–3344.
- [5] A. Siddique, W. N. Browne, and G. M. Grimshaw, “Frames-of-reference-based learning: Overcoming perceptual aliasing in multistep decision-making tasks,” *IEEE Transactions on Evolutionary Computation*, vol. 26, no. 1, pp. 174–187, 2022.
- [6] D. Shah, B. Eysenbach, N. Rhinehart, and S. Levine, “Rapid Exploration for Open-World Navigation with Latent Goal Models,” in *5th Annual Conference on Robot Learning*, 2021. [Online]. Available: https://openreview.net/forum?id=d_SWJhyKFv
- [7] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, “Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age,” *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.
- [8] A. Kadian, J. Truong, A. Gokaslan, A. Clegg, E. Wijmans, S. Lee, M. Savva, S. Chernova, and D. Batra, “Sim2real predictivity: Does evaluation in simulation predict real-world performance?” *IEEE Robotics and Automation Letters*, vol. 5, no. 4, p. 6670–6677, Oct. 2020. [Online]. Available: <http://dx.doi.org/10.1109/LRA.2020.3013848>
- [9] T. Gervet, S. Chintala, D. Batra, J. Malik, and D. S. Chaplot, “Navigating to objects in the real world,” *Science Robotics*, vol. 8, no. 79, p. eadf6991, 2023. [Online]. Available: <https://www.science.org/doi/abs/10.1126/scirobotics.adf6991>
- [10] Q. Chen, R. Wang, M. Lyu, and J. Zhang, “Transformer-based reinforcement learning for multi-robot autonomous exploration,” *Sensors*, vol. 24, no. 16, 2024. [Online]. Available: <https://www.mdpi.com/1424-8220/24/16/5083>
- [11] H. Wang, A. H. Tan, and G. Nejat, “Navformer: A transformer architecture for robot target-driven navigation in unknown and dynamic environments,” *IEEE Robotics and Automation Letters*, vol. 9, no. 8, pp. 6808–6815, 2024.
- [12] D. Shah, A. Sridhar, N. Dashora, K. Stachowicz, K. Black, N. Hirose, and S. Levine, “ViNT: A foundation model for visual navigation” in *7th Annual Conference on Robot Learning*, 2023. [Online]. Available: <https://arxiv.org/abs/2306.14846>
- [13] D. Lawson and A. H. Qureshi, “Control transformer: Robot navigation in unknown environments through prm-guided return-conditioned sequence modeling,” in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 9324–9331.
- [14] A. Sridhar, D. Shah, C. Glossop, and S. Levine, “Nomad: Goal masked diffusion policies for navigation and exploration,” in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 63–70.
- [15] J. Liang, A. Payandeh, D. Song, X. Xiao, and D. Manocha, “Dtg : Diffusion-based trajectory generation for mapless global navigation,” 2024. [Online]. Available: <https://arxiv.org/abs/2403.09900>
- [16] Y. Cao, J. Lew, J. Liang, J. Cheng, and G. Sartoretti, “Dare: Diffusion policy for autonomous robot exploration,” in *Submission to IEEE International Conference on Robotics and Automation (ICRA)*, 2025.
- [17] W. Yu, J. Peng, H. Yang, J. Zhang, Y. Duan, J. Ji, and Y. Zhang, “Ldp: A local diffusion planner for efficient robot navigation and collision avoidance,” in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024, pp. 5466–5472.
- [18] B. Liao, S. Chen, H. Yin, B. Jiang, C. Wang, S. Yan, X. Zhang, X. Li, Y. Zhang, Q. Zhang, and X. Wang, “Diffusionondrive: Truncated diffusion model for end-to-end autonomous driving,” *arXiv preprint arXiv:2411.15139*, 2024.
- [19] W. Cai, J. Peng, Y. Yang, Y. Zhang, M. Wei, H. Wang, Y. Chen, T. Wang, and J. Pang, “Navdp: Learning sim-to-real navigation diffusion policy with privileged information guidance,” 2025. [Online]. Available: <https://arxiv.org/abs/2505.08712>
- [20] K. Friston, “Friston, k.j.: The free-energy principle: a unified brain theory? nat. rev. neurosci. 11, 127–138,” *Nature reviews. Neuroscience*, vol. 11, pp. 127–38, 02 2010.
- [21] K. Friston, FitzGerald, Thomas, Rigoli, Francesco, Schwartenbeck, Philipp, Pezzulo, and Giovanni, “Active inference: A process theory,” *Neural Computation*, vol. 29, no. 1, pp. 1–49, 01 2017. [Online]. Available: https://doi.org/10.1162/NECO_a_00912
- [22] T. Parr, G. Pezzulo, and K. J. Friston, *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. The MIT Press, 03 2022. [Online]. Available: <https://doi.org/10.7551/mitpress/12441.001.0001>
- [23] K. Friston, F. Rigoli, D. Ognibene, C. Mathys, T. Fitzgerald, and G. Pezzulo, “Active inference and epistemic value,” *Cognitive Neuroscience*, vol. 6, no. 4, pp. 187–214, 2015, PMID: 25689102. [Online]. Available: <https://doi.org/10.1080/17588928.2015.1020053>
- [24] de Tinguy, Daria and Verbelen, Tim and Dhoedt, Bart, “Exploring and learning structure : active inference approach in navigational agents,” in *Active inference : 5th international workshop, IWAI 2024, revised selected papers*, Buckley, Christopher L. and Cialfi,

- Daniela and Lanillos, Pablo and Pitliya, Riddhi J. and Sajid, Noor and Shimazaki, Hideaki and Verbelen, Tim and Wisse, Martijn, Ed., vol. 2193. Springer, 2025, pp. 105–118. [Online]. Available: http://doi.org/10.1007/978-3-031-77138-5_7
- [25] D. de Tinguy, T. Van de Maele, T. Verbelen, and B. Dhoedt, “Spatial and temporal hierarchy for autonomous navigation using active inference in minigrid environment,” *Entropy*, vol. 26, no. 1, p. 83, Jan. 2024. [Online]. Available: <http://dx.doi.org/10.3390/e26010083>
- [26] E. Delavari, J. Moore, J. Hong, and J. Kwon, “Perceptual motor learning with active inference framework for robust lateral control,” 2025. [Online]. Available: <https://arxiv.org/abs/2503.01676>
- [27] C. Chi, Z. Xu, S. Feng, E. Cousineau, Y. Du, B. Burchfiel, R. Tedrake, and S. Song, “Diffusion policy: Visuomotor policy learning via action diffusion,” *The International Journal of Robotics Research*, 2024.
- [28] K. Fujii and S. Murata, “Hierarchical latent dynamics model with multiple timescales for learning long-horizon tasks,” in *2023 IEEE International Conference on Development and Learning (ICDL)*, 2023, pp. 479–485.
- [29] D. Ha and J. Schmidhuber, “Recurrent world models facilitate policy evolution,” in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, ser. NIPS’18. Red Hook, NY, USA: Curran Associates Inc., 2018, p. 2455–2467.
- [30] D. Hafner, T. P. Lillicrap, M. Norouzi, and J. Ba, “Mastering atari with discrete world models,” in *International Conference on Learning Representations*, 2021. [Online]. Available: <https://openreview.net/forum?id=0oabwyZbOu>
- [31] T. Taniguchi, S. Murata, M. Suzuki, D. Ognibene, P. Lanillos, E. Uğur, L. Jamone, T. Nakamura, A. Ciria, B. Lara, and G. Pezzulo, “World models and predictive coding for cognitive and developmental robotics: frontiers and challenges,” *Advanced Robotics*, vol. 37, no. 13, pp. 780–806, 2023. [Online]. Available: <https://doi.org/10.1080/01691864.2023.2225232>
- [32] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: <https://proceedings.neurips.cc/paper/files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>
- [33] D. Ha and J. Schmidhuber, “World models,” 2018. [Online]. Available: <https://zenodo.org/record/1207631>
- [34] P. Wu, A. Escontrela, D. Hafner, K. Goldberg, and P. Abbeel, “Daydreamer: World models for physical robot learning,” 2022. [Online]. Available: <https://arxiv.org/abs/2206.14176>
- [35] R. Sekar, O. Rybkin, K. Daniilidis, P. Abbeel, D. Hafner, and D. Pathak, “Planning to explore via self-supervised world models,” in *ICML*, 2020.
- [36] D. Hafner, T. Lillicrap, I. Fischer, R. Villegas, D. Ha, H. Lee, and J. Davidson, “Learning latent dynamics for planning from pixels,” 2019. [Online]. Available: <https://arxiv.org/abs/1811.04551>
- [37] J. S. V, S. Jalagam, Y. LeCun, and V. Sobal, “Gradient-based planning with world models,” 2023. [Online]. Available: <https://arxiv.org/abs/2312.17227>
- [38] J. Chun, Y. Jeong, and T. Kim, “Sparse imagination for efficient visual world model planning,” 2025. [Online]. Available: <https://arxiv.org/abs/2506.01392>
- [39] W. Liu, H. Zhao, C. Li, J. Biswas, B. Okal, P. Goyal, Y. Chang, and S. Pouya, “X-mobility: End-to-end generalizable navigation via world modeling,” *arXiv preprint arXiv:2410.17491*, 2024. [Online]. Available: <https://arxiv.org/abs/2410.17491>
- [40] R. P. Poudel, H. Pandya, S. Liwicki, and R. Cipolla, “ReCoRe: Regularized Contrastive Representation Learning of World Model,” in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2024, pp. 22904–22913. [Online]. Available: <https://doi.ieeecomputersociety.org/10.1109/CVPR52733.2024.02161>
- [41] Y. Duan, W. Mao, and H. Zhu, “Learning world models for unconstrained goal navigation,” 2024. [Online]. Available: <https://arxiv.org/abs/2411.02446>
- [42] A. Bar, G. Zhou, D. Tran, T. Darrell, and Y. LeCun, “Navigation world models,” *arXiv preprint arXiv:2412.03572*, 2024. [Online]. Available: <https://arxiv.org/abs/2412.03572>
- [43] D. Nie, X. Guo, Y. Duan, R. Zhang, and L. Chen, “Wmnav: Integrating vision-language models into world models for object goal navigation,” 2025. [Online]. Available: <https://arxiv.org/abs/2503.02247>
- [44] B. Kayalibay, A. Mirchev, P. van der Smagt, and J. Bayer, “Tracking and planning with spatial world models,” 2022. [Online]. Available: <https://arxiv.org/abs/2201.10335>
- [45] A. Hu, L. Russell, H. Yeo, Z. Murez, G. Fedoseev, A. Kendall, J. Shotton, and G. Corrado, “Gaia-1: A generative world model for autonomous driving,” *arXiv preprint arXiv:2309.17080*, 2023. [Online]. Available: <https://arxiv.org/abs/2309.17080>
- [46] X. Wang, Z. Zhu, G. Huang, X. Chen, J. Zhu, and J. Lu, “Drivedreamer: Towards real-world-driven world models for autonomous driving,” *arXiv preprint arXiv:2309.09777*, 2023.
- [47] O. Çatal, S. Wauthier, T. Verbelen, C. D. Boom, and B. Dhoedt, “Deep active inference for autonomous robot navigation,” *arXiv preprint arXiv:2003.03220*, 2020. [Online]. Available: <https://arxiv.org/abs/2003.03220>
- [48] O. Çatal, T. Verbelen, T. Van de Maele, B. Dhoedt, and A. Safron, “Robot navigation as hierarchical active inference,” *Neural Networks*, vol. 142, pp. 192–204, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0893608021002021>
- [49] D. de Tinguy, T. Verbelen, E. Gamba, and B. Dhoedt, “Bio-inspired topological autonomous navigation with active inference in robotics,” 2025. [Online]. Available: <https://arxiv.org/abs/2508.07267>
- [50] R. Smith, K. J. Friston, and C. J. Whyte, “A step-by-step tutorial on active inference and its application to empirical data,” *Journal of Mathematical Psychology*, vol. 107, p. 102632, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0022249621000973>
- [51] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, ser. NIPS ’20. Red Hook, NY, USA: Curran Associates Inc., 2020.
- [52] Y. Yamashita and J. Tani, “Emergence of functional hierarchy in a multiple timescale neural network model: A humanoid robot experiment,” *PLOS Computational Biology*, vol. 4, no. 11, pp. 1–18, 11 2008. [Online]. Available: <https://doi.org/10.1371/journal.pcbi.1000220>
- [53] A. Ahmadi and J. Tani, “A novel predictive-coding-inspired variational rnm model for online prediction and recognition,” *Neural Computation*, vol. 31, no. 11, pp. 2025–2074, 11 2019. [Online]. Available: https://doi.org/10.1162/neco_a_01228
- [54] S. Murata, Y. Yamashita, H. Arie, T. Ogata, S. Sugano, and J. Tani, “Learning to perceive the world as probabilistic or deterministic via interaction with others: A neuro-robotics experiment,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 4, pp. 830–848, 2017.
- [55] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” 2015. [Online]. Available: <https://arxiv.org/abs/1505.04597>
- [56] J. Song, C. Meng, and S. Ermon, “Denoising diffusion implicit models,” 2022. [Online]. Available: <https://arxiv.org/abs/2010.02502>
- [57] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [58] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1874–1883.
- [59] C. Tallec and Y. Ollivier, “Unbiasing truncated backpropagation through time,” *arXiv preprint arXiv:1705.08209*, 2017. [Online]. Available: <https://arxiv.org/abs/1705.08209>