

From Imitation to Alignment: Human-Preference Flow Policies for Long-Horizon Sidewalk Navigation

Honglin He Zhizheng Liu Yukai Ma Bolei Zhou

University of California, Los Angeles

<https://vail-ucla.github.io/FlowPilot>

Abstract: Autonomous long-horizon sidewalk navigation is essential for micro-mobility applications such as robotic food delivery and assistive electronic wheelchairs. Unlike autonomous driving on the road, long-horizon sidewalk navigation requires precise maneuvering through unpredictable sidewalk terrains and pedestrians, with a lightweight perception stack as minimal as a single monocular RGB camera. While imitation learning (IL) from demonstrations offers a practical solution, the resulting autopilot policy often suffers from compounding errors, a lack of social compliance on sidewalks, and deficiencies in counterfactual reasoning to handle complex situations. To address these challenges, we introduce FlowPilot, a mapless navigation policy that achieves robust and efficient long-horizon navigation performance using only a monocular RGB camera. We propose to use anchored flow matching as an action representation for policy pre-training on large-scale robot fleet data and to capture the diverse, complex, multimodal distribution of sidewalk navigation behaviors. To bridge the gap between imitation and alignment, we further design a human-in-the-loop preference learning scheme to tune the policy on a small amount of human intervention data. It strengthens the model’s counterfactual reasoning and social compliance on sidewalks. We evaluate FlowPilot through simulation and real-world experiments in diverse sidewalk environments. FlowPilot-Base achieves 42% success rate and 66% route completion in simulation, while FlowPilot-HP further improves robustness and social compliance, reducing IR by 40.0% and NIR by 52.1% relative to the base model.

Keywords: Visual Navigation; Imitation Learning; AI Alignment

1 Introduction

An increasing number of mobile robots are being deployed on urban sidewalks to perform everyday tasks such as food delivery and personal mobility assistance [1, 2, 3]. As shown in Fig. 1, a food delivery robot must navigate a couple of miles along sidewalks to reach its destination. It needs to safely avoid collisions with obstacles like scooters and interact socially with pedestrians. This setting also imposes heavy onboard computing and battery constraints, requiring a lightweight perception stack as minimal as just a single monocular camera. Furthermore, sidewalk layouts can vary significantly across cities and even neighborhoods. Local conditions also change frequently due to construction, temporary obstacles such as abandoned furniture and parked scooters, and unpredictable pedestrian activity, creating significant challenges for autonomous sidewalk navigation [4].

Recent advances in end-to-end navigation policy learning [5, 6, 7, 8, 9] offer a promising way to directly map visual observations to actions, enabling urban navigation without explicit mapping or heavyweight sensors. Most policies are trained with imitation learning (IL) [10] on large-scale offline data, leveraging demonstrations from heterogeneous sources, including teleoperated robot fleets [11, 12], human demonstrations [13], and web videos [9]. It has enabled increasingly capable visual navigation policies and has shown potential for cross-embodiment generalization.

Despite these advances, the mapless long-horizon sidewalk navigation remains challenging. First, standard IL often lacks an expressive action representation. Sidewalk navigation is inherently multi-



Figure 1: **Long-horizon sidewalk navigation** for a food delivery robot (the pink bot shown in the upper-left corner). Equipped with only a monocular camera and a GPS, the wheeled robot navigates complex sidewalk terrains, avoid pedestrians and obstacles, and follow social norms to reach its destination safely. We show the first-person views of the robot and the trajectories from FlowPilot.

modal and control-sensitive, *e.g.*, the robot may pass an obstacle on either side, yield or proceed at an intersection, and continuously adjust its turning and speed. However, common action parameterizations are limited. Unimodal Gaussians are too restrictive, while diffusion or flow-matching policies [14, 8] may introduce excessive randomness and temporal inconsistency, and anchor-based regression methods [15] often oversmooth fine-grained variations. To address this, we introduce anchored flow matching, which combines discrete behavior priors with continuous flow refinement to preserve multimodal coverage while generating smooth, temporally consistent future trajectories.

On the other hand, scaling model training on more diverse data does not necessarily yield better, more precise maneuvering behaviors. We observe that imitation learning on web data or robot fleet data creates a *generalization-precision dilemma*: while training on diverse data improves the policy’s generalization across environments and robot instances, it may compromise the precision of the policy’s action output when deployed on a specific robot instance, where each instance might have subtle variations in camera parameters, mechanics, and physical attributes. Because sidewalk navigation requires high precision, it is essential to further align the navigation model with the specific robot instance on which it is deployed. In the literature, interactive imitation learning [16, 17, 18, 19] mitigates distribution shift by collecting on-policy corrective supervision, while preference learning shows that comparative human feedback can steer model behavior without dense expert demonstrations [20, 21, 22]. Inspired by these insights, we design a human-in-the-loop finetuning scheme after large-scale imitation pretraining, bridging large-scale imitation and robot-specific alignment.

In this work, we propose a learning framework for robust long-horizon sidewalk navigation that supports multi-modal action prediction and model alignment from human feedback. We first introduce an anchored flow-matching policy to capture complex multimodal action distributions under long-horizon intentions. To incorporate goal information without inducing shortcut behaviors, we further rank and select candidate actions via a gated conditioning mechanism, encouraging the policy to ground decisions in scene context rather than directly taking the shortest path. To further adapt a pretrained policy to the precise control requirements of specific robots in complex sidewalk environments, we develop an on-policy preference-learning algorithm using interventions collected during deployment. Specifically, we fine-tune the policy with interventions and preference supervision tailored to our anchored action representation, increasing the likelihood of corrected actions while suppressing undesirable executed actions. This enables a smooth transition from large-scale pretraining to test-time alignment for specific robots. We summarize our contributions:

- 1) We propose an anchored flow-matching policy to model action distributions for long-horizon sidewalk navigation with gated conditioning to mitigate goal-driven shortcuts.
- 2) We introduce a reward-free preference learning algorithm that aligns the policy with socially compliant behaviors from human feedback while preserving imitation priors.
- 3) We evaluate our method through extensive simulation and real-world experiments in diverse and challenging sidewalk environments. Our method achieves 42% success rate and 66% route comple-

tion in closed-loop simulation benchmark, and improves real-world robustness and social compliance, reducing IR by 40.0% and NIR by 52.1% after preference fine-tuning.

2 Related Work

Long-horizon sidewalk navigation Visual navigation has been extensively studied, but task difficulty varies significantly with environmental scale and dynamics, map access, robot perception, and computing constraints. Traditional approaches build maps for localization and planning [23], but they are difficult to maintain in sidewalks where layouts change, dynamic agents are pervasive, and reliable mapping is costly. Recent work increasingly adopts end-to-end visuomotor policies that map raw sensory observations directly to waypoints [24, 6, 8, 25, 26, 15, 27]. However, many works focus on indoor or small-scale outdoor navigation [24, 8, 28], where horizons are shorter and interactions are constrained. In contrast, sidewalk navigation requires sustained goal pursuit under dense obstacles, pedestrians, and implicit social norms, making compounding errors and social compliance central challenges [26]. Although VLMs can enhance reasoning [26, 29, 30], deploying VLMs on compact sidewalk robots remains difficult due to strict latency, compute, and battery constraints.

Navigation foundation models Recent navigation foundation models aim to improve generalization by scaling training data across environments, and robots [6, 31, 25, 15, 31, 30, 26]. By leveraging large-scale visual data, these models can reduce the need for task-specific training. These advances suggest that large-scale pretraining provides a strong initialization for sidewalk navigation. However, scaling passive demonstrations alone does not fully address the closed-loop nature of long-horizon navigation. Offline data often lacks interactive recovery behaviors [27], and human corrections in safety-critical states. As a result, policies trained primarily through imitation may still suffer from distributional shift, shortcut behaviors, and limited causal understanding [26].

Human alignment for navigation Safe and socially compliant sidewalk navigation requires more than imitating expert trajectories; the policy must also align with social norms and deployment-specific safety requirements. Early approaches such as inverse reinforcement learning and adversarial imitation learning [32, 33, 34] infer reward functions [35] from demonstrations, but they are computationally expensive and often limited by the coverage and quality of expert data. More recent human-in-the-loop methods use corrective interventions [18, 19, 36] or trajectory preferences [37, 38, 39, 19, 40] to refine policies beyond demonstrations.

3 Method

In this section, we introduce FlowPilot, a human-preference flow policy for long-horizon sidewalk navigation. We first present the preliminaries in Sec. 3.1 and then describe the model architecture for imitation pretraining and the preference learning algorithm for alignment in Sec. 3.2.

3.1 Preliminaries

Task setting We focus on **mapless** GPS-guided sidewalk navigation from a **monocular RGB camera**, one of the most challenging navigation settings. The objective of the robot is to reach a destination while maintaining safety and social compliance, *i.e.*, avoiding collisions with pedestrians and obstacles, staying on the sidewalk, and minimizing contact with the grass. At each timestep t , the robot receives continuous RGB frames $o_{t-T:t}$ from the past T steps, a GPS goal $g_t \in \mathcal{G}$, and its pose p_t . The policy π_{Θ} needs to plan a trajectory $\mathcal{P} = \{p_{t+1}, \dots, p_{t+T}\}$ to generate control commands. Both p_t and g_t are highly noisy, making goal conditioning and long-horizon planning particularly challenging and leading to brittle behaviors if the policy over-relies on them.

Conditional flow matching policy Given a data sample $(x^1, s_t) \sim \mathcal{D}$, where $x^1 \sim \mathcal{P}$ denotes the planned future trajectory and $s_t = \{o_{t-T:t}, g_t, p_t\}$ denotes the current navigation state, we sample an initial noisy trajectory $x^0 \sim \mathcal{N}(0, \mathbf{I})$, then adopt Rectified Flow [41] to generate

$$x^\tau = (1 - \tau) \cdot x^0 + \tau \cdot x^1, \quad \tau \sim \mathcal{U}(0, 1), \quad (1)$$

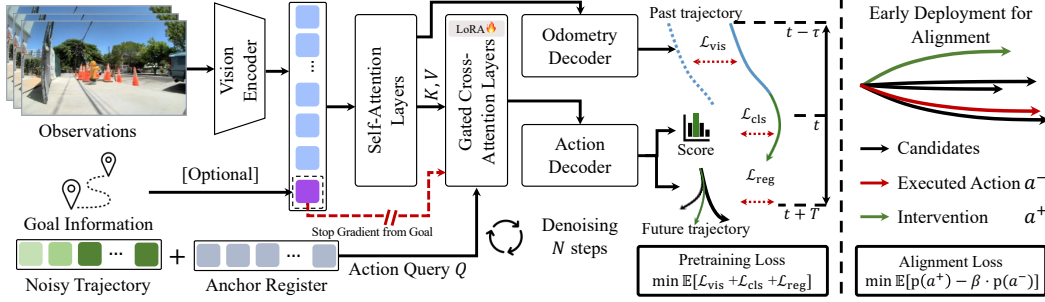


Figure 2: **Architecture and training pipeline of FlowPilot.** Given RGB observations and an optional goal, the encoder first extracts scene tokens, which are processed by self-attention layers and then used as keys and values in gated cross-attention. A set of noisy trajectory queries, together with learnable anchor registers, is denoised to generate multiple trajectory candidates. During pretraining, the model is supervised with visual odometry prediction, mode classification, and trajectory regression losses. During deployment, human interventions are collected to align the policy by increasing the likelihood of corrected actions a^+ and suppressing undesirable executed actions a^- .

We train a velocity model $v_{\Theta}(x^{\tau}, s_t, \tau)$ to regress the target velocity field:

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{(x^1, s_t) \sim \mathcal{D}, x^0 \sim \mathcal{N}(0, \mathbf{I}), \tau \sim \mathcal{U}(0, 1)} \left[\|v_{\Theta}(x^{\tau}, s_t, \tau) - v\|^2 \right], \quad (2)$$

where the target velocity is $v = x^1 - x^0$. During inference, we initialize $x^0 \sim \mathcal{N}(0, \mathbf{I})$ and solve the learned flow ODE to obtain the predicted trajectory $\hat{x}^1 = x^0 + \int_0^1 v_{\Theta}(x^{\tau}, s_t, \tau) d\tau$.

Preference learning Recent works of preference-based reinforcement learning have explored the use of preference datasets to align policy with human expectations. Given a dataset $\mathcal{D}_{\text{pref}}$ consisted of preference pairs $(o_{t-T:t}, g_t, p_t; x^+, x^-) \sim \mathcal{D}_{\text{pref}}$ means that the expert prefers the action x^+ over x^- given state $s_t = \{o_{t-T:t}, g_t, p_t\}$. Then we can train the policy π_{Θ} using the following objective [22]

$$\mathcal{L}_{\text{Pref}} = -\mathbb{E}[\log \sigma(\beta \cdot (\log \pi_{\Theta}(x^+ | s_t) - \log \pi_{\Theta}(x^- | s_t)))], \quad (3)$$

where $\sigma(\cdot)$ is the Sigmoid function, and $\beta > 0$ is a hyperparameter.

3.2 FlowPilot Framework

Model architecture As illustrated in Fig. 2, we first introduce the anchor design. Given state-action pair $(s_t, x^1) \sim \mathcal{D}$ sampled from demonstration dataset \mathcal{D} , we parameterize each trajectory x^1 in the local frame, then normalize the action space to $[-1, 1]^{T \times 2}$. Then we apply a constrained K -means clustering over the normalized action space to obtain a set of K anchor actions $\{\bar{x}_k\}_{k=1}^K$ as prototypical multimodal behaviors for subsequent mode scoring and trajectory refinement. During pretraining, each expert trajectory x^1 is assigned to its nearest anchor, $h = \arg \min_k \|x^1 - \bar{x}_k\|_2^2$. We sample noise $x^0 \sim \mathcal{N}(0, \mathbf{I})$ and a flow timestep $\tau \sim \text{Beta}(\alpha = 1.5, \beta = 1.0)$ [42]. The intermediate trajectory is constructed by linearly interpolating between noise and the expert trajectory, $\bar{x}_h^{\tau} = (1-\tau) \cdot x^0 + \tau \cdot \bar{x}_h$, with the target velocity field defined as $v = x^1 - x^0$. Conditioned on (s_t, τ, \bar{x}_h) , the model predicts a velocity field $v_{\Theta}(x^{\tau}, s_t, \tau, \bar{x}_h)$ and regresses it using an L2 loss. In parallel, a scoring head $h_{\Theta}(s_t)$ predicts the anchor assignment using cross-entropy loss $\text{CE}(h_{\Theta}(s_t), h)$.

To enable context-aware, goal-conditioned learning, we design the architecture shown in Fig. 2. We encode the observation history $o_{t-T:t}$ into features $V_{t-T:t} = [V_{t-T:t-1}; V_t]$ using an ImageNet-

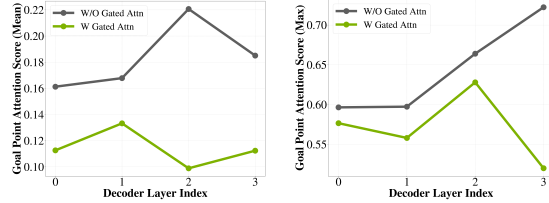


Figure 3: **Mitigating attention sink with gated attention.** The proportion of attention allocated to the goal / anchor across layers. The baseline exhibits attention-sink behavior, in which a large fraction of attention is concentrated on the goal, whereas gated attention substantially reduces this concentration and encourages better context utilization.



Figure 4: **Preference alignment with human intervention for a specific robot.** Human monitors policy predictions and intervenes when inappropriate behaviors occur. The **corrective action x^+** is treated as preferred over the **policy action x^-** , providing pairwise feedback for alignment.

pretrained[43] FastViT[44], and encode the goal g_t into G_t . Self-attention is applied over $V_{t-T:t}$ to obtain $\tilde{V}_{t-T:t} = \text{MSA}(V_{t-T:t})$. We also add two auxiliary heads that predict the frame-to-frame displacement from $\tilde{V}_{t-T:t-1}$ and the future trajectory, supervised with MSE losses \mathcal{L}_{vis} .

Given the context $C_t = [\tilde{V}_{t-T:t}; G_t]$, a noised query embedding $\mathbf{q}^\tau = \text{MLP}([\tilde{x}_h^\tau; \tau])$ attends to the context through gated cross-attention $\mathbf{z}^\tau = \text{softmax}\left(\frac{\mathbf{q}^\tau \cdot \sigma(C_t W_K)^\top}{\sqrt{d_h}}\right) \cdot \sigma(C_t W_V)$, where $\sigma(\cdot)$ is the sigmoid function and W_K, W_V are learnable projections. The latent \mathbf{z}^τ is then decoded to predict the velocity. As shown in Fig. 3, this gated attention mitigates attention sink on the goal and encourages the policy to use scene cues such as obstacles and sidewalk boundaries.

Preference alignment with human intervention Pure imitation learning remains limited by the coverage of demonstrations: the policy may fail in unseen safety-critical states, struggle to recover from distributional shift, or produce behaviors that are socially non-compliant. More importantly, although large-scale pretraining yields a broadly capable navigation policy, it does not necessarily provide the precise, robot-specific control. To bridge this gap, we further adapt the pretrained policy from broad demonstration-driven competence to robot-specific behavioral precision through human-in-the-loop preference alignment. As illustrated in Fig. 4, we deploy the policy π_Θ in the real world and allow a human to monitor the trajectory through the teleoperation platform. Whenever there is an unsafe or socially inappropriate prediction, the operator intervenes and provides corrective commands. Such interventions naturally focus on critical states s_π where pure imitation is insufficient. In each intervention, the human corrective action x^+ is treated as the preferred action, while the policy action $x^- = \pi_\Theta(s_\pi)$ is treated as the less preferred action. This yields a set of pairwise preference samples (s_π, x^+, x^-) that capture specific requirements of safety and social compliance.

Based on Eq. 3, we optimize the policy with preference objectives and regularize [45, 46] it toward a reference policy, *i.e.*, the pretrained policy $\pi_{\Theta_0} = (v_{\Theta_0}, h_{\Theta_0})$. Given a preference pair (x^+, x^-) , where x^+ denotes the human-preferred corrective action and x^- denotes the less-preferred policy prediction, we assign their corresponding anchor modes as h^+ and h^- . The scoring head is then encouraged to assign higher probability to the preferred anchor than to the rejected anchor. We use the score-level preference margin $\Delta_\Theta^{\text{score}} = \log p_\Theta(h^+ | s_t) - \log p_\Theta(h^- | s_t)$, and $\mathcal{L}_{\text{score}} = -\log(\sigma(\beta \cdot \Delta_\Theta^{\text{score}}))$ for finetuning. To preserve the prior, we regularize the scoring head toward the original distribution from the reference policy. Specifically, we define $p_\Theta(k | s_t) = \text{softmax}(h_\Theta(s_t))_k$ and $p_{\Theta_0}(k | s_t) = \text{softmax}(h_{\Theta_0}(s_t))_k$. We then use a KL regularization term:

$$\mathcal{L}_{\text{score}}^{\text{reg}} = D_{\text{KL}}(p_\Theta(\cdot | s_t) \| p_{\Theta_0}(\cdot | s_t)). \quad (4)$$

For the velocity estimation network, unlike previous works that explicitly reduce the likelihood of rejected actions [45, 46], we do not directly decrease the probability of the rejected action x^- . Since the velocity field models a continuous action distribution, directly pushing away x^- may introduce unstable gradients and damage nearby feasible actions. Instead, we use an asymmetric trajectory-level objective: the preferred corrective action is used as a positive regression target, while the rejected mode is regularized toward its anchor prior rather than toward the rejected policy prediction:

$$\mathcal{L}_{\text{vel}} = \mathbb{E}_{\tau, \epsilon} \left[\left\| \mathbf{v}_\Theta(s_t, \bar{x}_{h^+}^\tau, \tau, h^+) - v^+ \right\|^2 \right] + \lambda_{\text{anc}} \mathbb{E}_{\tau, \epsilon} \left[\left\| \mathbf{v}_\Theta(s_t, \bar{x}_{h^-}^\tau, \tau, h^-) - v^- \right\|^2 \right], \quad (5)$$

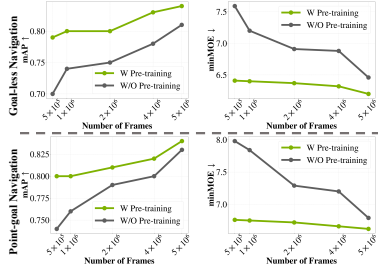


Figure 5: **Effectiveness of pre-training on mixed dataset.**

Method	minMOE ↓	minADE ↓	L2 ↓	mAP ↑
GNM [5]	7.31	0.63	1.32	0.63
ViNT [6]	8.51	0.74	1.54	0.74
NoMaD [8]	13.77	1.34	2.71	0.63
CityWalker [9]	8.94	0.71	1.48	0.77
MIMIC [27]	9.31	0.52	1.43	0.69
S2E [15]	6.77	0.46	1.73	0.81
DiffusionDrive [47]	6.96	0.63	1.71	0.77
FlowPilot-Base	6.63	0.49	1.04	0.87

Table 1: **Open-loop benchmark results.** All models are retrained on the same dataset for a fair comparison.

where $v^+ = x^+ - x^0$, $v^- = \bar{x}_{h^-} - x^0$. And to avoid overfitting to sparse corrective feedback and preserve the prior, we further regularize the finetuned policy toward the reference policy π_{Θ_0} using:

$$\mathcal{L}_{\text{vel}}^{\text{reg}} = \mathbb{E}_{\tau, \epsilon, h=(h^+, h^-)} \left[\left\| \mathbf{v}_{\Theta}(s_t, \bar{x}_h^{\tau}, \tau, h) - \text{sg} \left[\mathbf{v}_{\Theta_0}(s_t, \bar{h}_t^h, \tau, h) \right] \right\|^2 \right], \quad (6)$$

where $\text{sg}[\cdot]$ denotes the stop-gradient operation. Together, the preference alignment objective is

$$\mathcal{L}_{\text{Pref}} = \mathcal{L}_{\text{score}} + \lambda_{\text{reg},s} \mathcal{L}_{\text{score}}^{\text{reg}} + \lambda_v \mathcal{L}_{\text{vel}} + \lambda_{\text{reg},v} \mathcal{L}_{\text{vel}}^{\text{reg}} + \lambda_{\text{FM}} \mathcal{L}_{\text{FM}}, \quad (7)$$

where \mathcal{L}_{FM} is the original flow matching loss.

4 Experiments

4.1 Implementation Details

Dataset For pretraining, we collect **SideWalks-300**, a large-scale 300-hour video teleoperation dataset with 48,508 trajectory segments and about 19.4 million frames. The data are collected by a fleet of 281 robot instances navigating sidewalks across different cities, teleoperated by humans. Each segment contains 1920×1080 fisheye RGB video at 20 Hz, synchronized with robot states. For finetuning, we collect about 13 K frames of real-world intervention data and simulation intervention data generated by an A* [48] planner. More details are provided in the [Appendix](#).

Training details The model takes 11 consecutive frames as input and predicts a 4-second trajectory at 20 Hz. For real-time deployment and fair comparison, all methods use fixed 352×128 inputs and the same frame rate. We train the model in three stages. First, we pretrain the backbone policy on a mixture of large-scale video-action datasets [49, 50, 51, 52, 53, 11, 12, 54, 55] collected from different robots and environments. To handle heterogeneous distributions, we append an embodiment ID token [42] to the visual-temporal tokens and apply a dataset-specific normalization factor to the target trajectories. After pretraining, we perform supervised finetuning (SFT) on our dataset to adapt the policy to sidewalk navigation. This stage uses the same observation and trajectory prediction format as pretraining, but focuses on real-world sidewalk behaviors such as lane following, obstacle avoidance, pedestrian interaction, curb handling, and long-horizon goal-conditioned navigation. Finally, for preference alignment, we freeze all weights except the LoRA [56] modules applied to gated cross attention layers. More details are in the [Appendix](#).

Baselines We compare with state-of-the-art visual navigation models, including GNM [24], ViNT [6], NoMaD [8], MBRA [11], CityWalker [9], MIMIC [27], and S2E [15]. We also compare different policy adaptation strategies [57, 58, 35] using the same intervention data on the simulator [59, 60]. All baselines are **retrained** with the same protocol and data splits for a fair comparison.

4.2 Open-loop Evaluation

Table 1 summarizes the open-loop performance of different visual navigation models on the held-out evaluation set. We report minMOE [9, 26] to evaluate the minimum error between the generated trajectories and the ground-truth trajectory, while minADE and L2 measure the trajectory prediction accuracy in terms of average and endpoint distance errors. We further use mean average precision (mAP) to evaluate whether the model assigns higher confidence to accurate trajectory modes [61].

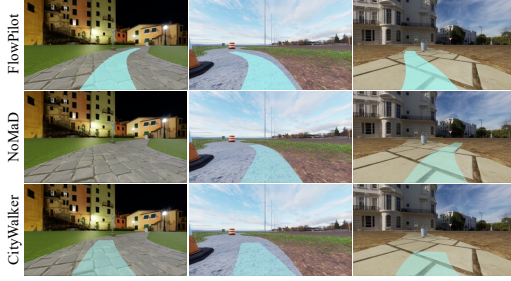


Figure 6: **Qualitative comparison in simulated sidewalk environments.**

Method	SR \uparrow	RC \uparrow	CR \downarrow	OBR \downarrow
GNM [24]	0.19	0.39	0.68	0.01
ViNT [6]	0.27	0.51	0.60	0.04
NoMaD [8]	0.21	0.45	0.74	0.04
MBRA [11]	0.31	0.50	0.54	0.15
CityWalker [25]	0.28	0.48	0.70	0.02
MIMIC [27]	0.25	0.44	0.62	0.03
S2E [15]	0.35	0.55	0.51	0.13
FlowPilot w/o GCA	0.29	0.59	0.61	0.09
FlowPilot-Base	0.42	0.66	0.43	0.01

Table 2: **Closed-loop evaluation with robot R0 on the simulation benchmark.**

Method	Basic Robot Configuration R0					Novel Robot Configuration R1					Novel Robot Configuration R2				
	\mathcal{D}_{Int}	SR \uparrow	RC \uparrow	CR \downarrow	OBR \downarrow	\mathcal{D}_{Int}	SR \uparrow	RC \uparrow	CR \downarrow	OBR \downarrow	\mathcal{D}_{Int}	SR \uparrow	RC \uparrow	CR \downarrow	OBR \downarrow
FlowPilot-Base	–	0.42	0.66	0.43	0.01	–	0.29	0.52	0.64	0.07	–	0.24	0.42	0.21	0.01
BC	265 K	0.22	0.44	0.28	0.46	265 K	0.27	0.51	0.71	0.02	265 K	0.28	0.49	0.35	0.37
HG-DAGger [57]	46 K	0.31	0.51	0.24	0.30	50 K	0.27	0.48	0.43	0.30	91 K	0.23	0.43	0.29	0.48
Ensemble-DAGger [58]	60 K	0.14	0.38	0.13	0.73	77 K	0.25	0.46	0.30	0.45	80 K	0.22	0.42	0.33	0.45
FlowPilot w/o Ref.	46 K	0.29	0.47	0.50	0.18	50 K	0.21	0.43	0.36	0.39	91 K	0.26	0.50	0.20	0.47
FlowPilot-HP	46 K	0.55	0.74	0.31	0.14	50 K	0.44	0.66	0.43	0.13	91 K	0.35	0.63	0.49	0.16

Table 3: **Closed-loop evaluation on the simulation benchmark with different robots.**

Compared with prior methods, FlowPilot achieves the best performance on minMOE, L2, and mAP, while maintaining competitive minADE. These results show that the proposed architecture not only predicts more accurate future trajectories, but also ranks its multi-modal predictions more reliably.

As shown in Fig. 5, pretraining on mixed dataset consistently improves performance across different data scales and task settings. With more training frames, both the pretrained and non-pretrained models improve, but the pretrained model achieves better mAP and lower minMOE throughout the scaling curve. This suggests that unified pretraining provides a strong robot-agnostic navigation prior, enabling the model to leverage downstream sidewalk data more effectively.

4.3 Closed-loop Evaluation

Unit-test benchmarks We evaluate different navigation models and conduct ablation studies on a closed-loop simulation benchmark built upon Urban-Sim [60, 59]. We first compare navigation methods on the basic robot configuration R0, which matches the average dynamics and camera of the wheeled robots in our training data. We then evaluate human-in-the-loop learning on R0 and two novel configurations, R1 and R2, with shifted dynamics and cameras such as a 20% change in focal length. This setting tests adaptation to deployment-specific embodiment shifts. Each scenario contains a 20-meter sidewalk segment with obstacles, and each method is evaluated over 300 trials. We report success rate (SR), route completion (RC), collision rate (CR), and off-boundary rate (OBR), where collisions, off-boundary behaviors, and timeouts are counted as failures. Intervention data \mathcal{D}_{Int} are collected from the same scenario distribution but are disjoint from the test scenarios.

Long-horizon benchmarks We further evaluate long-horizon navigation performance to evaluate deployment-scale robustness over extended routes. Each episode uses start and goal locations separated by more than 100 m. When a failure occurs, we do not terminate the episode; instead, we reset the robot to the closest valid waypoint along the route, continue the route, and record the event as an intervention. The benchmark contains 100 procedurally generated scenes and covers 36 km of total route distance. We report Normalized Intervention Rate (NIR) [62], Normalized Collision Rate (NCR), and Normalized Out-of-Boundary Rate (NOBR), which measure intervention events, collision events, and sidewalk-boundary violations normalized by traveled distance, respectively.

Results As shown in Tab. 2 and Fig. 6, FlowPilot-Base performs the best across all metrics, demonstrating the effectiveness of the proposed method. The ablation further shows the benefit

Metric	ViNT [6]	NoMaD [8]	MBRA [11]	CityWalker [25]	MIMIC [27]	S2E [15]	FlowPilot
NIR ↓	0.212	0.142	0.103	0.055	0.061	0.054	0.034
NCR ↓	0.043	0.040	0.044	0.035	0.030	0.021	0.023
NOBR ↓	0.169	0.101	0.059	0.019	0.009	0.033	0.011

Table 4: Closed-loop long-horizon evaluation on the simulation benchmark.



Figure 7: Qualitative comparison in real-world sidewalk environments.

Method	IR ↓	NIR ↓	DCR ↑	TCR ↑
ViNT [6]	0.081	0.304	0.867	0.858
NoMaD [8]	0.093	0.423	0.885	0.864
CityWalker [9]	0.146	0.483	0.809	0.804
FlowPilot-Base	0.020	0.073	0.904	0.916
FlowPilot-HP	0.012	0.035	0.928	0.939

Table 5: Closed-loop evaluation in long-horizon real-world sidewalk environments.



Figure 8: Qualitative results in diverse real-world sidewalk environments. We visualize the trajectories predicted by FlowPilot and executed by the robot during real-world deployment.

of Gated Cross Attention (GCA); compared with the one without GCA, FlowPilot-Base improves SR from 0.29 to 0.42 and reduces CR from 0.61 to 0.43, highlighting the importance of GCA for robust context understanding for navigation. For cross-robot adaptation, Tab. 3 shows that the model trained with human preference FlowPilot-HP achieves the highest SR and RC across R0, R1, and R2, demonstrating effective adaptation to shifted robot dynamics and camera configurations. Compared with the one without reference policy regularization (FlowPilot w/o Ref.), FlowPilot-HP consistently improves both SR and RC, indicating that reference regularization helps preserve the pretrained navigation prior during finetuning. And for the long-horizon task, Tab. 4 shows that FlowPilot achieves the lowest NIR, outperforming all baselines in large-scale scenarios.

4.4 Real-world Deployment

Robot, scenarios and metrics To validate real-world navigation performance, we conduct experiments on a wheeled robot and a robot dog shown in Fig. 1 and Fig. 8 in 8 scenarios covering 16 routes, with a total route length of approximately 2.8 km. In real-world environments, a human operator takes over when the robot exhibits unsafe behavior, such as approaching obstacles or pedestrians, or drifting toward non-walkable regions. Since these interventions ensure safety, collision rate and success rate are less informative. We therefore report Intervention Rate (IR) and



Figure 9: **Kilometer-scale long-horizon sidewalk navigation.** We evaluate FlowPilot in a long-horizon, complex real-world scenario, demonstrating its robust progress over long distances.



Figure 10: **Long-horizon sidewalk navigation under challenging nighttime illumination.** We evaluate FlowPilot on a long-horizon route at night with low-light regions and complex illumination changes, demonstrating its robustness to degraded visual conditions.

Normalized Intervention Rate (NIR) [62]. Specifically, IR is the fraction of distance driven under human takeover, *i.e.*, $IR = d_{intervention} / d_{total}$, while NIR measures the number of intervention events normalized by the traveled distance. We also report social-compliance metrics [26], including Distance Compliance Rate (DCR) and Time Compliance Rate (TCR), which measure the fraction of distance and time during which the robot remains socially compliant.

Results As shown in Tab. 5, both FlowPilot-Base and FlowPilot-HP substantially outperform prior methods in real-world deployment, especially on intervention-related metrics. FlowPilot-HP further reduces IR by 40.0% and NIR by 52.1% relative to FlowPilot-Base, validating the effectiveness of human preference fine-tuning. Qualitatively, Fig. 8, Fig. 9 and Fig. 10 show that our model produces smooth and stable trajectories across diverse sidewalk scenarios, including sidewalk lane keeping, obstacle avoidance, pedestrian awareness, and cross-embodiment deployment. And as shown in Fig. 7, our method also exhibits more socially compliant behavior than baselines where it maintains safer clearance and avoids collision.

5 Conclusion

We present FlowPilot, a novel long-horizon sidewalk navigation framework. We introduce an anchored flow-matching policy with a gated transformer for long-horizon, GPS-guided mapless sidewalk navigation. It provides a powerful representation that enables the model to capture the complex features of sidewalk environments. Built on large-scale imitation learning, we then fine-tune the policy to transition from imitation to alignment, using paired preference data collected in real-world environments. Simulation and real-world experiments show that FlowPilot outperforms baselines: FlowPilot-Base achieves 42% SR and 66% RC in simulation, while FlowPilot-HP improves robot-specific adaptation and real-world social compliance. In particular, human preference fine-tuning reduces IR by 40.0% and NIR by 52.1% relative to the base model.

6 Limitations

Our current policy, based on a monocular RGB camera, is susceptible to scale ambiguity and occasionally struggles to understand fine-grained 3D structural cues. This can lead to performance degradation in scenarios involving textureless surfaces or reflective obstacles. Future iterations would benefit from integrating explicit depth estimation or occupancy prediction to provide the necessary geometric priors for safer navigation in dense, cluttered environments.

References

- [1] V. Engesser, E. Rombaut, L. Vanhaverbeke, and P. Lebeau. Autonomous delivery solutions for last-mile logistics operations: A literature review and research agenda. *Sustainability*, 15(3): 2774, 2023.
- [2] X. Liu, L. Zhang, and T. Zhu. Service robots in my workplace: Effects of employee-service robot co-work experiences on psychological empowerment. *Journal of Hospitality Marketing & Management*, 34(2):175–203, 2025.
- [3] A. Tuomi, I. P. Tussyadiah, and J. Stienmetz. Applications and implications of service robots in hospitality. *Cornell Hospitality Quarterly*, 62(2):232–247, 2021.
- [4] E. Arntz, J. Van Duin, A. Van Binsbergen, L. Tavasszy, and T. Klein. Assessment of readiness of a traffic environment for autonomous delivery robots. *Frontiers in Future Transportation*, 4:1102302, 2023.
- [5] D. Shah, A. Sridhar, A. Bhorkar, N. Hirose, and S. Levine. Gnm: A general navigation model to drive any robot. *arXiv preprint:2210.03370*, 2022.
- [6] D. Shah, A. Sridhar, N. Dashora, K. Stachowicz, K. Black, N. Hirose, and S. Levine. Vint: A foundation model for visual navigation. *arXiv preprint arXiv:2306.14846*, 2023.
- [7] Y. Hu, J. Yang, L. Chen, K. Li, C. Sima, X. Zhu, S. Chai, S. Du, T. Lin, W. Wang, et al. Planning-oriented autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17853–17862, 2023.
- [8] 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)*, pages 63–70. IEEE, 2024.
- [9] X. Liu, J. Li, Y. Jiang, N. Sujay, Z. Yang, J. Zhang, J. Abanes, J. Zhang, and C. Feng. Citywalker: Learning embodied urban navigation from web-scale videos. *arXiv preprint arXiv:2411.17820*, 2024.
- [10] B. D. Argall, S. Chernova, M. Veloso, and B. Browning. A survey of robot learning from demonstration. *Robotics and autonomous systems*, 57(5):469–483, 2009.
- [11] N. Hirose, L. Ignatova, K. Stachowicz, C. Glossop, S. Levine, and D. Shah. Learning to drive anywhere with model-based reannotation. *IEEE Robotics and Automation Letters*, 11(2): 1242–1249, 2025.
- [12] H. Karnan, A. Nair, X. Xiao, G. Warnell, S. Pirk, A. Toshev, J. Hart, J. Biswas, and P. Stone. Socially compliant navigation dataset (scand): A large-scale dataset of demonstrations for social navigation. *IEEE Robotics and Automation Letters*, 7(4):11807–11814, 2022.
- [13] X. Pan, T. Zhang, B. Ichter, A. Faust, J. Tan, and S. Ha. Zero-shot imitation learning from demonstrations for legged robot visual navigation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 679–685. IEEE, 2020.
- [14] 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*, 44(10-11):1684–1704, 2025.
- [15] H. He, Y. Ma, W. Wu, and B. Zhou. From seeing to experiencing: Scaling navigation foundation models with reinforcement learning. *arXiv preprint arXiv:2507.22028*, 2025.
- [16] S. Ross, G. Gordon, and D. Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 627–635. JMLR Workshop and Conference Proceedings, 2011.

- [17] C. Celemin, R. Pérez-Dattari, E. Chisari, G. Franzese, L. de Souza Rosa, R. Prakash, Z. Ajanović, M. Ferraz, A. Valada, J. Kober, et al. Interactive imitation learning in robotics: A survey. *Foundations and Trends® in Robotics*, 10(1-2):1–197, 2022.
- [18] Z. M. Peng, W. Mo, C. Duan, Q. Li, and B. Zhou. Learning from active human involvement through proxy value propagation. *Advances in neural information processing systems*, 36: 77969–77992, 2023.
- [19] H. Cai, Z. Peng, and B. Zhou. Predictive preference learning from human interventions. In *Advances in Neural Information Processing Systems*, 2025.
- [20] D. M. Ziegler, N. Stiennon, J. Wu, T. B. Brown, A. Radford, D. Amodei, P. Christiano, and G. Irving. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*, 2019.
- [21] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al. Training language models to follow instructions with human feedback. *arXiv preprint arXiv:2203.02155*, 2022.
- [22] R. Rafailov, A. Sharma, E. Mitchell, C. D. Manning, S. Ermon, and C. Finn. Direct preference optimization: Your language model is secretly a reward model. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL <https://arxiv.org/abs/2305.18290>.
- [23] S. Thrun. Probabilistic robotics. *Communications of the ACM*, 45(3):52–57, 2002.
- [24] D. Shah, A. Sridhar, A. Bhorkar, N. Hirose, and S. Levine. Gnm: A general navigation model to drive any robot. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7226–7233. IEEE, 2023.
- [25] X. Liu, J. Li, Y. Jiang, N. Sujay, Z. Yang, J. Zhang, J. Abanes, J. Zhang, and C. Feng. City-walker: Learning embodied urban navigation from web-scale videos. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 6875–6885, 2025.
- [26] Z. Chen, Y. Guo, Z. Chu, M. Luo, Y. Shen, M. Sun, J. Hu, S. Xie, K. Yang, P. Shi, et al. Socialnav: Training human-inspired foundation model for socially-aware embodied navigation. *arXiv preprint arXiv:2511.21135*, 2025.
- [27] H. He, Y. Ma, B. Squicciarini, W. Wu, and B. Zhou. Learning sidewalk autopilot from multi-scale imitation with corrective behavior expansion. *arXiv preprint arXiv:2603.22527*, 2026.
- [28] 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. *arXiv preprint arXiv:2505.08712*, 2025.
- [29] A.-C. Cheng, Y. Ji, Z. Yang, Z. Gongye, X. Zou, J. Kautz, E. Bıyık, H. Yin, S. Liu, and X. Wang. Navila: Legged robot vision-language-action model for navigation. *arXiv preprint arXiv:2412.04453*, 2024.
- [30] M. Wei, C. Wan, J. Peng, X. Yu, Y. Yang, D. Feng, W. Cai, C. Zhu, T. Wang, J. Pang, et al. Ground slow, move fast: A dual-system foundation model for generalizable vision-and-language navigation. *arXiv preprint arXiv:2512.08186*, 2025.
- [31] A. Bar, G. Zhou, D. Tran, T. Darrell, and Y. LeCun. Navigation world models. *arXiv preprint arXiv:2412.03572*, 2024.
- [32] H. Kretzschmar, M. Spies, C. Sprunk, and W. Burgard. Socially compliant mobile robot navigation via inverse reinforcement learning. *The International Journal of Robotics Research*, 35(11):1289–1307, 2016.

- [33] J. Ho and S. Ermon. Generative adversarial imitation learning. *Advances in neural information processing systems*, 29, 2016.
- [34] B. D. Ziebart, A. L. Maas, J. A. Bagnell, A. K. Dey, et al. Maximum entropy inverse reinforcement learning. In *Aaai*, volume 8, pages 1433–1438. Chicago, IL, USA, 2008.
- [35] G. Seneviratne, J. An, S. Ellahy, K. Weerakoon, M. B. Elnoor, J. D. Kannan, A. T. Sunil, and D. Manocha. Halo: Human preference aligned offline reward learning for robot navigation. *arXiv preprint arXiv:2508.01539*, 2025.
- [36] H. Cai, Z. Peng, and B. Zhou. Robot-gated interactive imitation learning with adaptive intervention mechanism. In *International Conference on Machine Learning*, 2025.
- [37] D. Sadigh, A. D. Dragan, S. S. Sastry, and S. A. Seshia. Active preference-based learning of reward functions. In *Robotics: Science and Systems*, 2017.
- [38] J. Choi, C. Dance, J.-e. Kim, K.-s. Park, J. Han, J. Seo, and M. Kim. Fast adaptation of deep reinforcement learning-based navigation skills to human preference. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3363–3370. IEEE, 2020.
- [39] W. B. Knox, S. Hatgis-Kessell, S. Booth, S. Niekum, P. Stone, and A. Allievi. Models of human preference for learning reward functions. *arXiv preprint arXiv:2206.02231*, 2022.
- [40] R. Wang, W. Wang, and B.-C. Min. Feedback-efficient active preference learning for socially aware robot navigation. In *2022 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 11336–11343. IEEE, 2022.
- [41] X. Liu, C. Gong, and Q. Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.
- [42] NVIDIA, J. Bjorck, F. Castañeda, N. Cherniadev, X. Da, R. Ding, L. J. Fan, Y. Fang, D. Fox, F. Hu, S. Huang, J. Jang, Z. Jiang, J. Kautz, K. Kundalia, L. Lao, Z. Li, Z. Lin, K. Lin, G. Liu, E. Llontop, L. Magne, A. Mandlekar, A. Narayan, S. Nasiriany, S. Reed, Y. L. Tan, G. Wang, Z. Wang, J. Wang, Q. Wang, J. Xiang, Y. Xie, Y. Xu, Z. Xu, S. Ye, Z. Yu, A. Zhang, H. Zhang, Y. Zhao, R. Zheng, and Y. Zhu. Gr00t n1: An open foundation model for generalist humanoid robots. *arXiv preprint arXiv:2503.14734*, 2025.
- [43] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [44] P. K. A. Vasu, J. Gabriel, J. Zhu, O. Tuzel, and A. Ranjan. Fastvit: A fast hybrid vision transformer using structural reparameterization. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5785–5795, 2023.
- [45] J. Liu, G. Liu, J. Liang, Y. Li, J. Liu, X. Wang, P. Wan, D. Zhang, and W. Ouyang. Flow-grpo: Training flow matching models via online rl. *arXiv preprint arXiv:2505.05470*, 2025.
- [46] J. Liu, G. Liu, J. Liang, Z. Yuan, X. Liu, M. Zheng, X. Wu, Q. Wang, M. Xia, X. Wang, et al. Improving video generation with human feedback. *arXiv preprint arXiv:2501.13918*, 2025.
- [47] B. Liao, S. Chen, H. Yin, B. Jiang, C. Wang, S. Yan, X. Zhang, X. Li, Y. Zhang, Q. Zhang, et al. Diffusiondrive: Truncated diffusion model for end-to-end autonomous driving. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 12037–12047, 2025.
- [48] P. E. Hart, N. J. Nilsson, and B. Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968.

- [49] A. Zhang, C. Eranki, C. Zhang, J.-H. Park, R. Hong, P. Kalyani, L. Kalyanaraman, A. Gamare, A. Bagad, M. Esteve, et al. Toward robust robot 3-d perception in urban environments: The ut campus object dataset. *IEEE Transactions on Robotics*, 40:3322–3340, 2024.
- [50] T. Akhtyamov, M. A. Mdfaa, J. A. R. Benavides, A. Nigmatzyanov, S. Bakulin, G. Devchich, D. Fatykhov, D. R. Salinas, A. Mazurov, K. Zipa, et al. Egowalk: A multimodal dataset for robot navigation in the wild. *arXiv preprint arXiv:2505.21282*, 2025.
- [51] J. Liang, D. Das, D. Song, M. N. H. Shuvo, M. Durrani, K. Taranath, I. Penskiy, D. Manocha, and X. Xiao. Gnd: Global navigation dataset with multi-modal perception and multi-category traversability in outdoor campus environments. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2383–2390. IEEE, 2025.
- [52] N. Hirose, D. Shah, A. Sridhar, and S. Levine. Sacson: Scalable autonomous control for social navigation. *IEEE Robotics and Automation Letters*, 9(1):49–56, 2023.
- [53] D. M. Nguyen, M. Nazeri, A. Payandeh, A. Datar, and X. Xiao. Toward human-like social robot navigation: A large-scale, multi-modal, social human navigation dataset. In *2023 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 7442–7447. IEEE, 2023.
- [54] D. Dauner, M. Hallgarten, T. Li, X. Weng, Z. Huang, Z. Yang, H. Li, I. Gilitschenski, B. Ivanovic, M. Pavone, et al. Navsim: Data-driven non-reactive autonomous vehicle simulation and benchmarking. *Advances in Neural Information Processing Systems*, 37:28706–28719, 2024.
- [55] NVIDIA Corporation. Physicalai autonomous vehicles dataset. <https://huggingface.co/datasets/nvidia/PhysicalAI-Autonomous-Vehicles>, 2025. Hugging Face dataset. Accessed: 2026-06-10.
- [56] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, et al. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3, 2022.
- [57] M. Kelly, C. Sidrane, K. Driggs-Campbell, and M. J. Kochenderfer. Hg-dagger: Interactive imitation learning with human experts. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8077–8083. IEEE, 2019.
- [58] K. Menda, K. Driggs-Campbell, and M. J. Kochenderfer. Ensembledagger: A bayesian approach to safe imitation learning. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5041–5048. IEEE, 2019.
- [59] M. Mittal, P. Roth, J. Tigue, A. Richard, O. Zhang, P. Du, A. Serrano-Munoz, X. Yao, R. Zurbrugg, N. Rudin, et al. Isaac lab: A gpu-accelerated simulation framework for multi-modal robot learning. *arXiv preprint arXiv:2511.04831*, 2025.
- [60] W. Wu, H. He, C. Zhang, J. He, S. Z. Zhao, R. Gong, Q. Li, and B. Zhou. Towards autonomous micromobility through scalable urban simulation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 27553–27563, 2025.
- [61] S. Ettinger, S. Cheng, B. Caine, C. Liu, H. Zhao, S. Pradhan, Y. Chai, B. Sapp, C. R. Qi, Y. Zhou, et al. Large scale interactive motion forecasting for autonomous driving: The waymo open motion dataset. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9710–9719, 2021.
- [62] A. Zhang, H. Sikchi, A. Zhang, and J. Biswas. Creste: Scalable mapless navigation with internet scale priors and counterfactual guidance. In *Robotics: Science and Systems (RSS)*, 2025.
- [63] J. Van Den Berg, S. J. Guy, M. Lin, and D. Manocha. Reciprocal n-body collision avoidance. In *Robotics research: the 14th international symposium ISRR*, pages 3–19. Springer, 2011.

Appendix

We organize the appendix as follows. We first introduce the demonstration video in Sec. A, followed by additional real-world results in Sec. B. We then provide additional simulation results in Sec. C, open-loop evaluation results in Sec. D, and the implementation details in Sec. E. We also provide an ethics statement in Sec. F, where we describe the safety protocol used during real-world deployment.

A Demonstration Video

We highly recommend watching our supplementary video for detailed demonstrations. The video presents a broad set of real-world experiments that thoroughly evaluate the performance of FlowPilot across diverse sidewalk environments. It consists of four sections:

1. FlowPilot capabilities demonstration FlowPilot demonstrates robust navigation in complex real-world sidewalk environments. It successfully negotiates narrow passages, cluttered layouts, and broken curbs while maintaining safe and socially compliant behaviors, including effective obstacle avoidance and pedestrian awareness.

2. Long-horizon sidewalk navigation FlowPilot completes GPS-guided long-horizon navigation with only a few human interventions. It maintains stable sidewalk lane keeping and consistent goal progress over time, while remaining robust to lighting changes and transient disturbances in challenging sidewalk environments.

3. Comparison with SOTA methods We present side-by-side real-world evaluations against representative state-of-the-art methods under the same setting. The comparisons highlight the advantages of FlowPilot in trajectory smoothness, navigation stability, and safety.

4. Cross-embodiment generality FlowPilot transfers effectively across different robot platforms both without finetuning and with only a few embodiment-specific examples. It preserves reasonable navigation behaviors under changes in platform dynamics and sensing configurations, demonstrating strong generalization and rapid adaptation across embodiments.

B Real-World Closed-Loop Results

B.1 Robot hardware setup

As illustrated in Fig. 11, we evaluate FlowPilot on two robot platforms for real-world experiments. We use a wheeled robot for the majority of real-world experiments, as it represents the primary deployment platform in our sidewalk navigation setting. In addition, we deploy FlowPilot on a legged robot to assess cross-embodiment generalization, as discussed in Sec. B.6. This setup allows us to evaluate whether the learned navigation behaviors remain effective when transferred to platforms with different dynamics, actuation constraints, and sensing configurations. For both platforms, we use onboard sensors to capture visual observations during navigation. The sensor streams are transmitted over a network to an external compute unit, where the policy performs inference and predicts navigation actions. These actions are then sent back to the robot for execution at 10-20 FPS. This deployment pipeline enables us to use the same policy interface across different platforms while keeping the onboard robot setup lightweight.

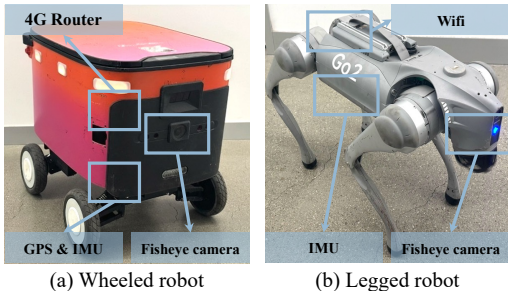


Figure 11: **Robot platforms used in real-world experiments.** We use a wheeled robot for the majority of experiments, and deploy a legged robot to assess cross-embodiment generalization.

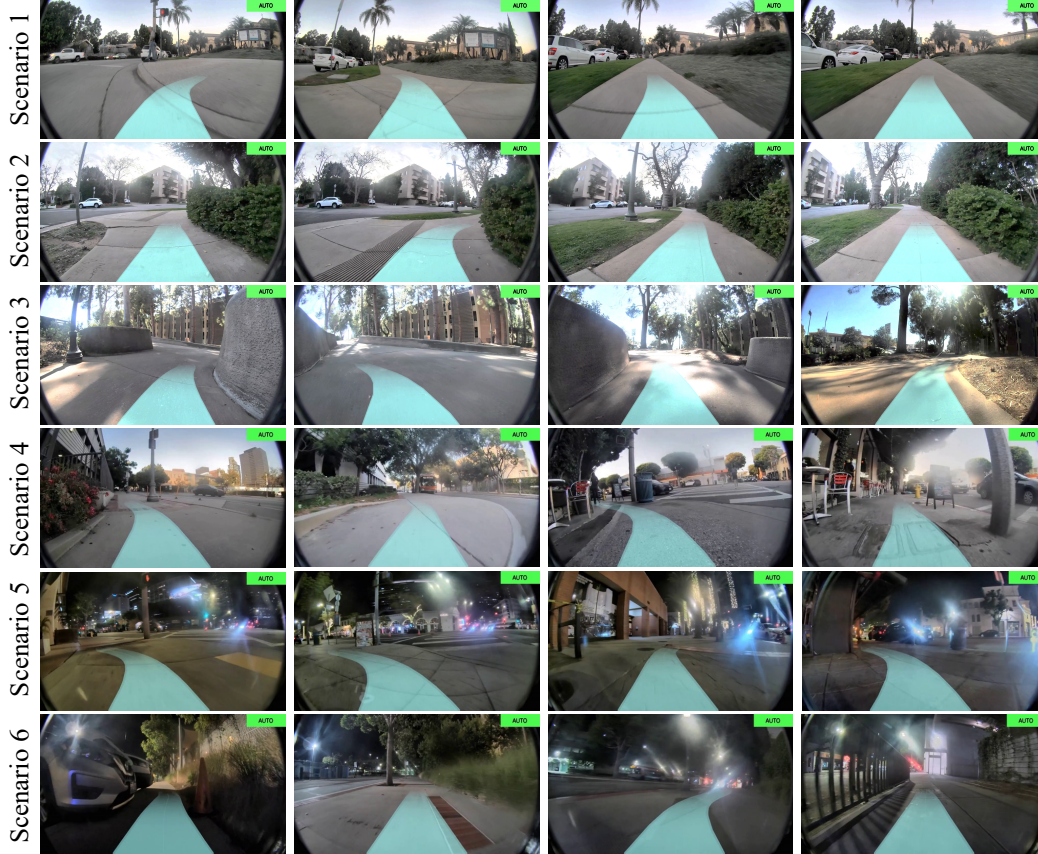


Figure 12: **Sidewalk lane keeping in diverse real-world scenes.** FlowPilot maintains a stable position within the sidewalk under varying appearances and layouts, producing smooth, safe trajectories.

During deployment, we use a laptop as the operator interface to visualize the live camera stream, predicted trajectory, robot state, and system status. A joystick controller is connected to the deployment laptop and is used by the safety operator to take over control whenever intervention is required. Each intervention is automatically recorded by monitoring the control-status topic, which indicates whether the robot is under autonomous policy control or manual joystick control. The corresponding timestamps are synchronized with the logged sensor streams and robot states, allowing us to label intervention segments and compute intervention-based evaluation metrics.

We conduct an additional real-world closed-loop evaluation across more diverse scenarios to further validate the robustness of our policy. We deploy the robot in 5 urban blocks during dusk and nighttime, resulting in 10 total trials and covering approximately 17.8 km of real-world sidewalk navigation. Due to the unstable deployment behavior of several baseline methods in long-horizon real-world settings, we do not repeatedly deploy them in this expanded evaluation for safety reasons. As given in Tab. 6, the results show that FlowPilot-HP reduces both IR and NIR while improving DCR and TCR, demonstrating better real-world robustness and social compliance.

Method	IR ↓	NIR ↓	DCR ↑	TCR ↑
FlowPilot-Base	0.055	0.106	0.922	0.937
FlowPilot-HP	0.017	0.049	0.949	0.951

Table 6: **Additional closed-loop evaluation in real-world sidewalk environments.**

B.2 Long-horizon sidewalk navigation

As illustrated in Fig. 9, Fig. 10, as well as Sec. 2 of the supplementary video, FlowPilot demonstrates robust long-horizon sidewalk navigation across diverse real-world conditions. Fig. 9 presents



Figure 13: **Reactive obstacle avoidance in cluttered sidewalks.** FlowPilot selects safe bypass maneuvers around static obstacles and returns to the nominal path after passing them.

a kilometer-scale daytime route, where the robot maintains consistent goal progress over an extended distance while performing stable sidewalk lane keeping, smooth turning, and obstacle avoidance. Fig. 10 further evaluates FlowPilot in a nighttime route with low-light regions and complex illumination changes, showing that the policy remains robust under degraded visual conditions. Across both cases, FlowPilot generates stable and socially compliant behaviors, avoids obstacles, and recovers from minor deviations without human intervention. Sec. 2 of the supplementary video provides the full video demonstrations of these two long-horizon cases, and we highly recommend readers to watch the video for a more detailed visualization of the navigation behaviors and qualitative results.

B.3 Sidewalk lane keeping

As illustrated in Fig. 12, FlowPilot reliably tracks the sidewalk corridor across a wide range of layouts, including narrow passages, curb-constrained segments, texture changes, etc. The policy keeps a stable lateral offset from boundaries (*e.g.*, curbs, grass edges, sidewalk boundaries), resulting in smooth behaviors. When the visual appearance or sidewalk width changes, the robot quickly re-centers and continues forward progress without requiring teleoperation or explicit re-initialization.

B.4 Obstacle avoidance

As illustrated in Fig. 13, FlowPilot performs robust obstacle avoidance in cluttered sidewalk environments with common static obstacles (*e.g.*, trash bins, cones, parked micromobility devices, etc.). The robot proactively slows down when clearance becomes tight, chooses a conservative side to pass, and maintains a collision-free margin while preserving forward progress. Once the obstacle is bypassed, the policy executes a smooth transition back to the sidewalk centerlines, rather than oscillating near the boundaries.



Figure 14: **Pedestrian-aware navigation without interventions.** FlowPilot keeps comfortable clearance, yields in crowded situations, and re-centers once the path is clear.

B.5 Pedestrian awareness

As illustrated in Fig. 14, FlowPilot exhibits pedestrian-aware behaviors that are consistent with socially compliant navigation. In crowded scenes, the robot reduces speed, maintains a comfortable clearance, and yields when the available space is insufficient for a safe pass. When pedestrians move away or the corridor opens up, the policy accelerates smoothly and re-centers, avoiding abrupt turns or any aggressive behaviors.

B.6 Cross-embodiment generalization

Since our policy outputs *waypoints* (rather than platform-specific low-level controls), it can be deployed across robot embodiments with minimal interface changes. Concretely, the same FlowPilot predicts short-horizon waypoints in a robot-centric frame, which are then tracked by a platform-specific controller (*e.g.*, a PD controller for the wheeled base and a velocity tracker for the legged robot). This modeling enables transferring the learned navigation behavior to other robot platforms without additional fine-tuning or few-shot demonstrations. Specifically, we collect 30 minutes of preference data on the Go2 platform and finetune the policy with the same preference-learning objective. More quantitative results are provided in Sec. D.2. As illustrated in the final section of the demonstration video, we provide two deployment examples on the Unitree Go2 legged robot, showing that FlowPilot can preserve stable sidewalk-following and obstacle-avoidance behaviors across different robot embodiments.

B.7 Typical failure cases

As shown in Fig. 15, while FlowPilot generally exhibits pedestrian-aware and socially compliant behaviors, we also observe suboptimal behaviors in challenging edge cases. **(1) Narrow passage with pedestrians:** in a narrow space, the policy fails to generate a feasible passing behavior between the other robots and nearby pedestrians and instead becomes stuck in an overly conservative state, repeatedly stopping until an intervention changes the scene or the robot state. **(2) Crosswalk / traffic-light scenario with a pedestrian present:** when a pedestrian remains within the interaction zone, the policy defaults to stopping and does not execute an alternative safe behavior (*e.g.*, edging forward to im-



Figure 15: **Representative failure cases in crowded or constrained scenes.** When faced with ambiguous right-of-way or insufficient clearance, FlowPilot may be overly conservative and stop, requiring human intervention to continue.

prove visibility or re-positioning), leading to deadlock. These cases highlight the need for stronger intent reasoning and explicit deadlock resolution strategies in dense human-robot interaction.

C Simulation Closed-Loop Results

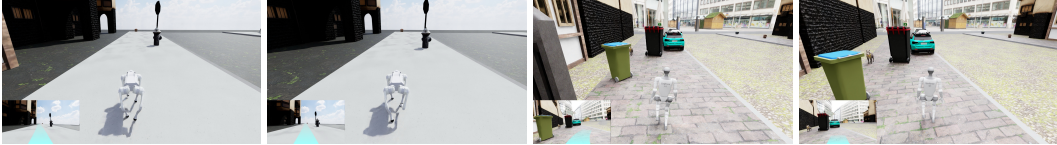


Figure 16: **Cross-embodiment generalization in simulated environments.**

We further demonstrate the cross-embodiment generalization ability of FlowPilot in closed-loop simulation, as demonstrated in Fig. 16. Since FlowPilot predicts platform-agnostic waypoints in the robot-centric frame, the same high-level policy can be deployed on different robot configurations by changing only the low-level tracking controller. We deploy the policy on two legged embodiments, Unitree Go2 and Unitree G1, to demonstrate cross-embodiment generalization in closed-loop simulation. Since FlowPilot outputs short-horizon waypoints in the robot-centric frame, we use an embodiment-specific tracking module to convert the predicted waypoints into low-level locomotion commands. For each robot, we train a separate locomotion policy in IsaacLab [59] to track desired base velocities and local waypoint targets. This setup allows us to test whether the learned navigation behavior can transfer across substantially different robot dynamics.

D Open-Loop Results

In this section, we provide additional open-loop evaluation results to complement the results in the main paper. Open-loop evaluation measures whether the predicted future trajectories match human teleoperation trajectories under the same visual observations. We analyze the effectiveness of each key component in the following sections.

D.1 Effectiveness of action representation

Tab. 7 summarizes an ablation study on different trajectory parameterizations for open-loop prediction on SideWalks-300.

Discrete (K-means + Scoring): we first cluster expert trajectories into K anchors and train a scoring network to select one anchor at inference time. While this provides a multi-modal hypothesis set, the selected prototype can be overly coarse, yielding suboptimal tracking quality when converted into controls.

Regression (S2E [15]): direct regression predicts a single trajectory, which tends to average over multiple plausible futures and can underperform in ambiguous scenes.

Diffusion (DiffusionDrive [47]): diffusion improves multi-modality but may introduce high-frequency noise, leading to less stable predicted plans. **Ours (anchored flow matching):** by combining anchor-based multi-modality with continuous refinement, our method achieves the best overall performance across metrics.

D.2 Effectiveness of preference learning

Tab. 8 and Tab. 9 quantify the impact of preference learning on robot-specific open-loop benchmarks for the wheeled and legged platforms, respectively. Specifically, we collect approximately 2 hours

Method	minMOE ↓	minADE ↓	L2 ↓	mAP ↑
Discrete	7.68	0.58	1.59	0.82
Regression	6.77	0.46	1.73	0.81
Diffusion	6.96	0.63	1.71	0.77
FlowPilot-Base	6.63	0.49	1.04	0.87

Table 7: **Action representation comparison on open-loop benchmark.**

Method	minMOE ↓	minADE ↓	L2 ↓	mAP ↑
GNM	7.31	0.63	1.32	-
ViNT	8.31	2.61	4.47	-
NoMaD	11.10	1.72	3.17	-
CityWalker	10.83	2.04	3.59	0.37
CityWalker-HP	9.71	1.55	2.93	0.46
FlowPilot-Base	4.63	1.17	3.11	0.53
FlowPilot-HP	3.64	0.61	1.67	0.79

Table 8: **Open-loop benchmark results on robot-specific dataset (Wheeled robot).**

Method	minMOE ↓	minADE ↓	L2 ↓	mAP ↑
GNM	31.99	2.13	3.79	-
ViNT	11.68	1.34	2.48	-
NoMaD	22.81	1.34	2.59	-
CityWalker	12.49	1.01	1.99	0.19
CityWalker-HP	17.17	0.75	1.70	0.23
FlowPilot-Base	3.59	0.53	2.99	0.27
FlowPilot-HF	3.24	0.21	0.91	0.74

Table 9: **Open-loop benchmark results on robot-specific dataset (Legged robot).**

of data for each robot and split it into two splits for training and testing. We first pretrain FlowPilot on the large-scale dataset and then fine-tune it on the robot-specific training split; preference learning is performed using human intervention pairs collected on the corresponding platform. As shown in the tables, preference fine-tuning consistently improves trajectory quality and action-level alignment (lower minMOE / minADE and L2, higher mAP), indicating better adaptation to safety-critical interactions and ambiguous situations in real deployments. We also fine-tune CityWalker [9] to get CityWalker-HP on the same robot-specific training split. Notably, FlowPilot achieves substantially stronger performance, suggesting that our model provides richer priors and more transferable navigation knowledge, which preference learning further refines to better match human preferences.

E Implementation Details

E.1 Pretraining dataset



Figure 17: **Overview of the large-scale pretraining dataset.** We pretrain FlowPilot on ~ 300 hours of diverse real-world navigation data spanning varied lighting, weather, and scene layouts, including frequent interactions with obstacles and pedestrians.

As illustrated in Fig. 17, our pretraining dataset contains approximately 300 hours of real-world robot navigation collected across diverse environments and conditions. It covers substantial variation in visual appearance (*e.g.*, time of day, strong sunlight/glare, shadows, and different camera exposures), sidewalk layouts (*e.g.*, curved segments, ramps, curbs, and intersections), and dynamic



Figure 18: **Qualitative comparison between FlowPilot predictions and human interventions.** We visualize representative real-world frames with the model-predicted path (cyan) and the trajectory executed during human intervention (green). Across diverse sidewalk scenes, the intervention traces highlight where humans deviate from the model to ensure safety and progress.

complexity. The data includes frequent interactions with static obstacles and moving agents such as pedestrians, micromobility robots, and animals, providing rich information for learning.

E.2 Expert intervention data collection in real-world sidewalk environments

After pretraining, we deploy FlowPilot in real-world navigation with a human-in-the-loop safety protocol to collect intervention data. At each timestep, the policy predicts a short-horizon trajectory, which is visualized and treated as the *candidate* control plan. A human operator monitors the robot and intervenes whenever the predicted plan is unsafe or likely to fail (*e.g.*, insufficient clearance, imminent collision, or deadlock). During an intervention, we record a paired sample consisting of the model prediction and the human-corrected execution, forming a supervision signal that captures the human preference. As illustrated in Fig. 18, these intervention pairs provide targeted corrections in challenging situations and serve as training data for preference-based fine-tuning.

E.3 Expert intervention data collection in simulated sidewalk environments

We collect expert intervention data in simulation using an ORCA [63]-based expert policy. For each episode, we generate a reference route τ_{ref} from the start location to the goal and sample intermediate waypoints along the route. The learned policy is deployed in closed loop, and its predicted trajectory is continuously compared with the local reference trajectory. When the average displacement error exceeds a threshold, *i.e.*, $\text{ADE}(\pi_{\theta}, \tau_{\text{ref}}) > 1.0$ m, or when the robot approaches unsafe states such as collision or off-sidewalk regions, the ORCA expert takes over control. During takeover, we record the model trajectory as the negative action and the expert-corrected trajectory as the preferred action. The resulting paired trajectories constitute the intervention dataset \mathcal{D}_{Int} for finetuning.

E.4 Details of model architecture

In this section, we provide details of the model architecture used in FlowPilot. We adopt a FastViT-based visual encoder [44]. Specifically, we encode the past 10 frames sampled at 20 Hz into frame-level tokens (*i.e.*, one token per frame) via global average pooling, and encode the current frame into patch-level tokens by flattening the backbone feature map. Each token is projected to a d_{model} -dimensional embedding using a lightweight MLP, and the past-frame tokens are further augmented with a sinusoidal time embedding. We additionally use a learnable position embedding over the visual tokens, followed by a 4-layer Transformer with 8 attention heads to extract scene context.

To incorporate goal information, the goal point is embedded by an MLP into a d_{model} -dimensional token and appended to the visual tokens as a conditioning token. To mitigate shortcut learning and improve robustness, we apply stochastic masking over the goal point during training, randomly dropping the goal-point token with a very high probability 0.9.

For action generation, we represent candidate plans using a fixed set of anchor trajectories obtained by K-Means clustering over the training dataset. During training, we apply the anchored flow-matching process to these anchors by sampling a noise level τ from a Beta distribution and linearly mixing anchor trajectories with Gaussian noise. The noised anchors are encoded by a lightweight action encoder and then refined by an anchored flow-matching Transformer consisting of 4 layers and 8 attention heads, which predicts the flow (velocity field) conditioned on the visual tokens and the goal token via gated cross-attention layers. The refined anchor features are decoded by the MLP heads to predict i) per-step normalized velocities and headings for each anchor mode and ii) a mode score for anchor selection. As shown in Tab. 10, we provide detailed parameters of the model.

Parameter	Value
RGB Resolution	352×128
Observation Sequence	11 frames @ 20Hz
Backbone	FastViT-MA36 [44]
Backbone Input Channels	6 ($\text{RGB}_t + \text{RGB}_{t-1}$)
Backbone Feature Dimension	1216
Token Dimension (d_{model})	1280
Gated Self-Attn Layers	4
Gated Self-Attn Heads	8
Dropout	0.1
Goal Dropout	0.9
Anchors	Fixed (From K-Means)
Anchor Modes K	64
Anchor Horizon	20 steps @ 5Hz
Anchor Dimension	3D ($\Delta x, \Delta y, \Delta \theta$)
Noise Schedule	$\tau \sim \text{Beta}(1.5, 1.0)$
Time Buckets	1000
Inference Steps	4
Gated Cross-Attn Layers	4
Gated Cross-Attn Heads	8
Dropout	0.1

Table 10: **Details of FlowPilot model.**

E.5 Details of pretraining

As illustrated in Tab. 11, we give the details of parameters used in the pretraining stage.

E.6 Details of finetuning

During finetuning, we apply LoRA [56] to the attention projection layers (query/key/value projections). We give the details of parameters used in the finetuning stage as illustrated in Tab. 12.

Parameter	Value
# Epochs n_{ep}	50
Batch Size	32 ($\times 8$)
weight Decay	1×10^{-6}
Learning Rate	1×10^{-4}
Optimizer	AdamW
LR Schedule	Cosine
Scheduler Period	100
Compute Resources	$8 \times$ NVIDIA PRO 6000

Table 11: **FlowPilot pretraining.**

Parameter	Value
Optimization	
# Epochs n_{ep}	30
Batch Size	256
Learning Rate	1×10^{-5}
Optimizer	AdamW
Weight Decay	0
LR Schedule	Cosine
Scheduler Period	80
Compute Resources	$8 \times$ NVIDIA PRO 6000
LoRA	
LoRA Rank r	16
LoRA Scaling α	16
LoRA Dropout	0
LoRA Target Modules	Attention projection layers
Trainable Parameters	LoRA only (others frozen)

Table 12: **FlowPilot finetuning hyperparameters.**

E.7 Details of controller

In practice, we find that this simple PD controller is sufficient for stable deployment since the policy already outputs smooth, goal-directed short-horizon waypoints, and the controller mainly serves as a robust mapping from waypoint space to (v, w) under actuation limits. We convert the predicted short-horizon trajectory into low-level control commands using a lightweight trajectory-based PD controller. Given a predicted waypoint $(\Delta x, \Delta y)$ in the robot local frame over a fixed control interval $\Delta t = 1s$, we compute the linear and angular velocities as $v = \frac{\Delta x}{\Delta t}$ and $w = \frac{\arctan(\Delta y/\Delta x)}{\Delta t}$, with additional safeguards for degenerate cases (*e.g.*, $\Delta x \approx 0$). We further apply standard clipping on velocity ($v_{\max} = 1.5 \text{ m/s}$) and angular rate ($w_{\max} = 0.35 \text{ rad/s}$) to satisfy platform constraints, and attenuate the commanded speed based on an estimated local curvature to improve safety in sharp turns ($v = \frac{\Delta x}{\Delta t} \cdot \exp(-0.05 \cdot |w|)$ when $|w| \geq 0.2$).

F Ethics Statement

For real-world deployment, the robots operated under strict safety constraints: maximum velocity and acceleration were limited by both hardware and software. Researchers monitored all experiments and were able to immediately intervene, stop the robot, or take over control to ensure safety. These interventions are explicitly marked as **Manual** in the figures and video. This protocol ensured that the robot operated under strict safety constraints throughout the experiments.