

Imagined pedestrian encounters: analyzing world-model rollouts in a reinforcement learning driving agent

Stefan Zlatinov.¹, Ilija Mizhimakoski.¹, Gorjan Nadzinski.¹, Mile Stankovski.¹

Faculty of Electrical Engineering and Information Technologies – Ss. Cyril and Methodius University in Skopje, Skopje, North Macedonia¹
zlatinov@feit.ukim.edu.mk

Abstract: *World-model reinforcement learning agents plan by imagining future trajectories through a learned latent dynamics model. In safety-critical driving tasks, the quality of these imagined rollouts — particularly their representation of vulnerable road users — directly determines whether the agent can anticipate and avoid dangerous situations. We analyze the imagined rollouts of a DreamerV3 agent trained to navigate urban environments in CARLA using semantic segmentation and depth observations. By decoding the agent's latent imagination into observation space, we qualitatively and quantitatively examine how pedestrians are represented in imagined futures, comparing episodes that result in collision against those where the agent successfully avoids. We investigate whether the world model accurately predicts pedestrian presence and motion in its imagined horizon, and whether reconstruction fidelity of pedestrian regions correlates with avoidance outcomes. Our analysis provides insight into the internal representations that underlie emergent safety behaviors in model-based driving agents and highlights limitations of finite imagination horizons for pedestrian safety.*

Keywords: WORLD MODEL, REINFORCEMENT LEARNING, AUTONOMOUS DRIVING, DREAMERV3, PEDESTRIAN SAFETY, IMAGINATION ROLLOUT, LATENT DYNAMICS, CARLA, SAFETY-CRITICAL AI

1. Introduction

Model-based reinforcement learning (MBRL) agents learn a predictive model of the environment and use it to plan by simulating future trajectories in a learned latent space [1]. In autonomous driving, this planning-by-imagination paradigm is appealing because it allows the agent to evaluate potential actions before executing them [2; 3]. DreamerV3 [4] is a recent world-model agent that has demonstrated strong performance across diverse domains, including simulated driving in CARLA [5]. Unlike model-free approaches that map observations directly to actions [6; 7], world-model agents rely on the quality of imagined rollouts to evaluate candidate action sequences. This dependence raises a natural concern: if the learned dynamics model fails to preserve a hazard in its imagined future, the agent may plan as if the road ahead is clear.

A central question for safety-critical deployment is therefore whether the learned world model faithfully represents vulnerable road users (in particular pedestrians) in its imagined futures. If the model fails to maintain an accurate representation of a pedestrian blocking the lane during imagination, the agent may select actions that lead to collision even when avoidance is physically possible. Pedestrian behavior prediction has been studied extensively in the context of conventional perception pipelines [8; 9], and intent-aware planning has been explored in model-predictive settings [10; 11]. However, relatively little work has examined how learned world models in MBRL agents internally represent pedestrians during latent imagination, the setting where planning decisions are actually made.

Recent work has applied world-model RL to driving in CARLA with varying observation modalities. Think2Drive [12] uses a latent world model for quasi-realistic autonomous driving, while CarDreamer [13] provides an open-source learning platform for world-model-based driving. A recent survey [14] covers the broader landscape of world models for autonomous driving. These works, along with broader RL-for-driving surveys [15; 16; 17], focus primarily on driving performance metrics (success rate, infraction count, route completion) rather than on the internal quality of imagined rollouts with respect to specific road users. On the evaluation side, adversarial testing frameworks such as ANTI-CARLA [18] stress-test driving agents but do not analyze the imagination content of world-model agents. Safe RL approaches have addressed uncertainty-aware planning [19] and latent-space trajectory optimization [20], yet the question of what the world model actually imagines about a pedestrian near the conflict point remains largely unexamined. Our work addresses this gap by directly decoding imagined latent states into observation space and

measuring how well the world model preserves pedestrian representations during planning.

In this work we study the imagined rollouts of a DreamerV3 agent in a controlled urban scenario where a pedestrian walks to the center of the lane and stands still, blocking the ego vehicle's path. We adopt an event-centered approach: we capture imagination at the moment the stationary pedestrian is visible ahead and examine the imagined future only over the short conflict-resolution period that follows. We decode the agent's latent imagination into observation space and compare imagined segmentation frames against ground truth, using a strict localized fidelity metric. Our analysis covers 565 pedestrian-encounter episodes and reveals that (1) the world model retains only weak pedestrian support under a strict exact-class metric, and (2) this weakness appears in both collision and avoidance episodes, suggesting that incorrect imagined lane clearance, rather than simple pedestrian disappearance, is the more informative failure mode.

2. Method

Agent and Environment

We use a DreamerV3 agent [4] trained in the CARLA simulator [5]. DreamerV3 learns a recurrent state-space model (RSSM) that encodes observations into a compact latent state and predicts future latent states conditioned on actions. An actor-critic algorithm optimizes the policy entirely within the learned latent space, using imagined trajectories rather than real environment interactions for policy improvement [21].

Perception. The agent receives two 128 x 128 image channels at each step: a single-channel semantic segmentation map where class labels are ordered by collision risk (pedestrians occupy the highest intensity, value 255; drivable road occupies the lowest), and a normalized inverse-depth image that concentrates representational precision in the near field. These visual inputs are complemented by a low-dimensional numerical vector containing the vehicle's speed, longitudinal acceleration, yaw rate, lateral and heading deviations from the planned route, the previous steering and throttle/brake command, and a sequence of ten ego-frame waypoints spaced 3 m apart along the route (30 m look-ahead). A shared convolutional encoder processes the image pair while a separate multilayer perceptron encodes the numerical channels; both streams are fused into the RSSM latent state.

Control outputs. The policy produces two continuous signals per step: a steering command and a signed longitudinal command, both in [-1, 1]. Positive longitudinal values map to throttle and negative values to braking. Actions are repeated for two consecutive simulation ticks (each 0.05 s), yielding an effective control frequency of 10 Hz.

Training reward. The agent was trained on 200 m routes in Town02 for 10 million environment steps. The reward signal consists of a dense route-progress component (virtual gates placed every 2 m along the route, each yielding a reward proportional to lateral proximity to the route center), combined with a quadratic overspeed penalty above 30 km/h and a set of terminal bonuses and penalties for reaching the goal, timing out, leaving the route, or colliding with a vehicle or pedestrian. Crucially, pedestrian safety is supervised only through a large terminal penalty upon collision, scaled quadratically with impact speed. No distance-based proximity costs, time-to-collision shaping, or explicit braking incentives are included; any anticipatory avoidance behavior is therefore emergent rather than directly rewarded. During training the environment contained non-adversarial crossing pedestrians but no deliberately adversarial scenarios; the protester archetype used in this study was never encountered during learning.

Scenario Design

Each episode uses a straight 60 m route in a CARLA urban map. A jaywalking pedestrian ("protester" archetype) is spawned on the sidewalk at approximately 45 m from the route start. When the ego vehicle approaches, the pedestrian walks from the sidewalk to a blocking point near the center of the ego's lane at 2.0 m/s, then stops and remains stationary for up to 10 s. The pedestrian does not react to the approaching vehicle and does not yield. By the time the agent's imagination is captured, the pedestrian has typically reached the blocking point and is standing still in the lane. The agent therefore faces a stationary obstacle that must be avoided by braking or steering. Because the pedestrian is not visible at route start, early episode time is irrelevant; the analysis begins only when the pedestrian enters the conflict zone. Out of 640 recorded episodes, 75 had no pedestrian encounter (the pedestrian was never visible). Of the remaining 565 encounter episodes, 475 resulted in successful avoidance (the agent slowed or steered around the pedestrian) and 90 resulted in collision.

Imagination Capture

At each episode step where the pedestrian is visible in the segmentation observation, the agent's current latent state is used to produce a 32-step imagined rollout. These imagined latent states are decoded back into observation space by the learned decoder, producing predicted segmentation and depth frames that can be compared against ground truth. We select one capture per episode, specifically the one whose actual pedestrian mean depth is closest to 10\,m, to ensure a consistent comparison distance across episodes. At this distance the pedestrian is typically already stationary at the blocking point.

Event-Centered Analysis

We restrict analysis to the conflict-resolution period: from the capture step until the episode terminates (collision) or the pedestrian leaves the camera field of view (avoidance). This yields a right-censored future for each capture. We do not analyze the full nominal 32-step imagination horizon, as most episodes terminate or resolve the conflict well before that point.

Strict Local Overlap Metric

For each valid future step k after capture, we define the ground-truth pedestrian mask G_k as the set of pixels with exact semantic class value 255. Let $K=|G_k|$ denote the number of pixels in this mask. We apply no dilation, so the support region is exactly the ground-truth mask. Within this region, we threshold imagined segmentation values at 250 and retain up to the top- K surviving pixels, forming the predicted set P_k . The strict local overlap is then the Intersection over Union (IoU) between P_k and G_k :

$$(1) \quad IoU_k = \frac{P_k \cap G_k}{P_k \cup G_k}$$

This metric is intentionally strict: it measures whether the imagined segmentation reproduces the pedestrian at the correct location with the correct class value.

3. Results

Qualitative Case Study

Fig.1 shows decoded segmentation frames from one representative collision episode (capture at 11.8 s) and one representative avoidance episode (capture at 10.8 s). In the actual future, the pedestrian (highlighted in green) remains present in the lane throughout the conflict period: it is stationary at its blocking point and grows larger as the ego vehicle approaches. In the imagined future, the pedestrian representation is faint and degrades rapidly. By $k=6$ (0.6 s into imagination), the pedestrian signal in the imagined frames is substantially weaker than in the corresponding actual frames. In several imagined frames the road region where the pedestrian should be standing is instead predicted as empty lane. The world model effectively imagines the obstacle clearing when in reality the pedestrian is still there. This pattern (the world model imagining premature lane clearance rather than maintaining the stationary pedestrian) is the central qualitative finding and is observed across multiple episodes beyond the two shown here.

Strict Local Fidelity Over Horizon

Fig.2 shows the mean strict local overlap as a function of future time after capture, separately for collision and avoidance episodes. Both curves start near 0.65-0.70 at $k=0$ (the capture step itself, where the imagined state is initialized from the encoder) and decline to approximately 0.25-0.28 within the first second. After 1.0 s, both curves continue a gradual decline but remain above zero, indicating that the pedestrian signal does not disappear entirely. The collision and avoidance curves overlap substantially throughout the observed horizon, indicating that strict pedestrian fidelity does not cleanly separate the two outcomes. Shaded bands indicate the standard error of the mean. The lower panel shows the number of valid episodes contributing to each point; collision episodes are censored earlier due to episode termination upon impact. Each point on the curve averages only over episodes that are still valid at that future step, so the curves are not affected by survivorship bias from early terminations.

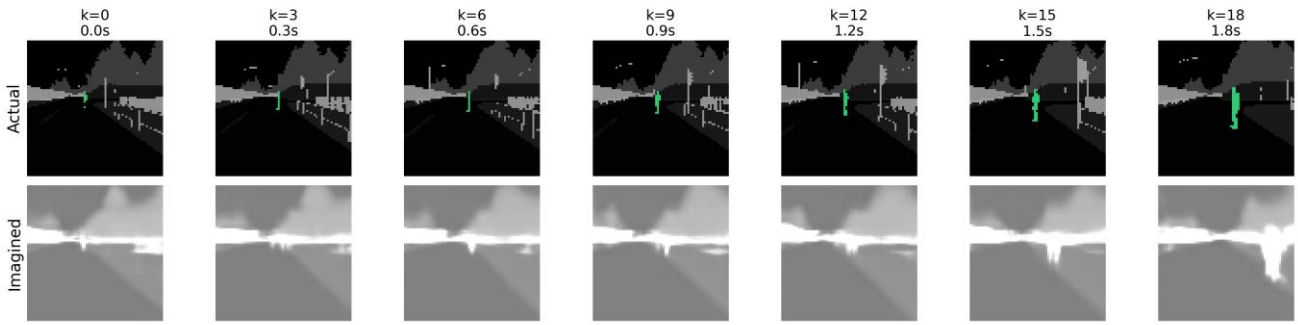
Episode-Horizon Heatmap

Fig.3 displays the per-episode strict local overlap over the first 1.0 s after capture. Rows represent individual episodes sorted by mean early overlap; columns represent future time steps at 0.1 s intervals. The heatmap reveals substantial heterogeneity: some episodes maintain moderate-to-high fidelity across the early horizon, while others show near-zero overlap from the first imagined step. Crucially, high-fidelity and low-fidelity episodes appear in both outcome groups, further supporting the observation that strict fidelity alone does not predict whether the agent avoids or collides. The grey cells in the heatmap indicate time steps where the pedestrian was no longer visible in the ground-truth frame, so no overlap could be computed.

Summary Table

Table 1 summarizes the imagination analysis across all 565 encounter episodes. The mean early strict overlap (first 1.0 s) is 0.40 overall, with similar values for avoidance (0.40) and collision (0.37) subgroups. The median valid horizon is approximately 2.2–2.4 s for both groups. On average, fewer than half of the ground-truth pedestrian pixels are reproduced by the imagination under the strict metric (mean predicted pixels: 6.3 out of 17.0 ground-truth pixels in the early window).

Collision ($t_0 = 11.8$ s)



Avoidance ($t_0 = 10.8$ s)

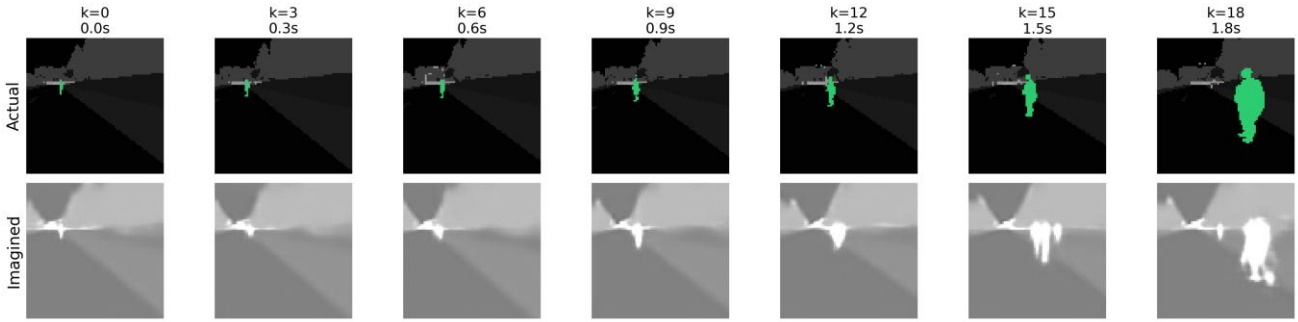


Fig. 1 Actual versus imagined segmentation in a collision episode (top two rows) and an avoidance episode (bottom two rows). The pedestrian (green) remains stationary in the lane in actual frames. In imagined frames, the pedestrian signal is weak and fades rapidly, as if the lane were clearing.

Table 1: Protester mid-crossing imagination summary. One capture is selected per episode near 10 m depth; no-encounter episodes are excluded. Early overlap averages the first 1.0 s after capture (imagination threshold 250).

	N	Med. Hor. [s]	Early IoU	Full IoU	GT px	Pred px
All	565	2.30	0.40	0.32	17.0	6.3
Avoidance	475	2.20	0.40	0.33	17.3	6.5
Collision	90	2.40	0.37	0.28	15.4	5.5

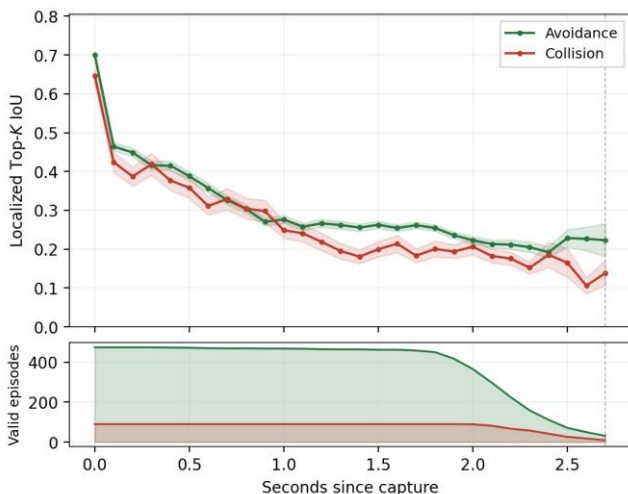


Fig. 2 Strict local overlap versus future time after capture. Top: mean localized Top-K IoU for collision and avoidance episodes (shaded: 1 standard error of the mean). Bottom: number of valid episodes at each time step. Both outcome groups show similar fidelity decay; collision episodes are censored earlier.

4. Discussion

The results show that the DreamerV3 world model retains some pedestrian signal in its imagined rollouts: the pedestrian does not vanish entirely at the capture step. However, under a strict exact-class metric with no dilation, this signal is weak: mean early strict overlap is at or below 0.40 in all subgroups, and fewer than half of the ground-truth pedestrian pixels survive the thresholded Top-K rule in imagination.

More importantly, the strict fidelity values for collision and avoidance episodes are similar (0.37 vs. 0.40). This means that pixel-level pedestrian reconstruction quality does not by itself explain why the agent avoids in some episodes and collides in

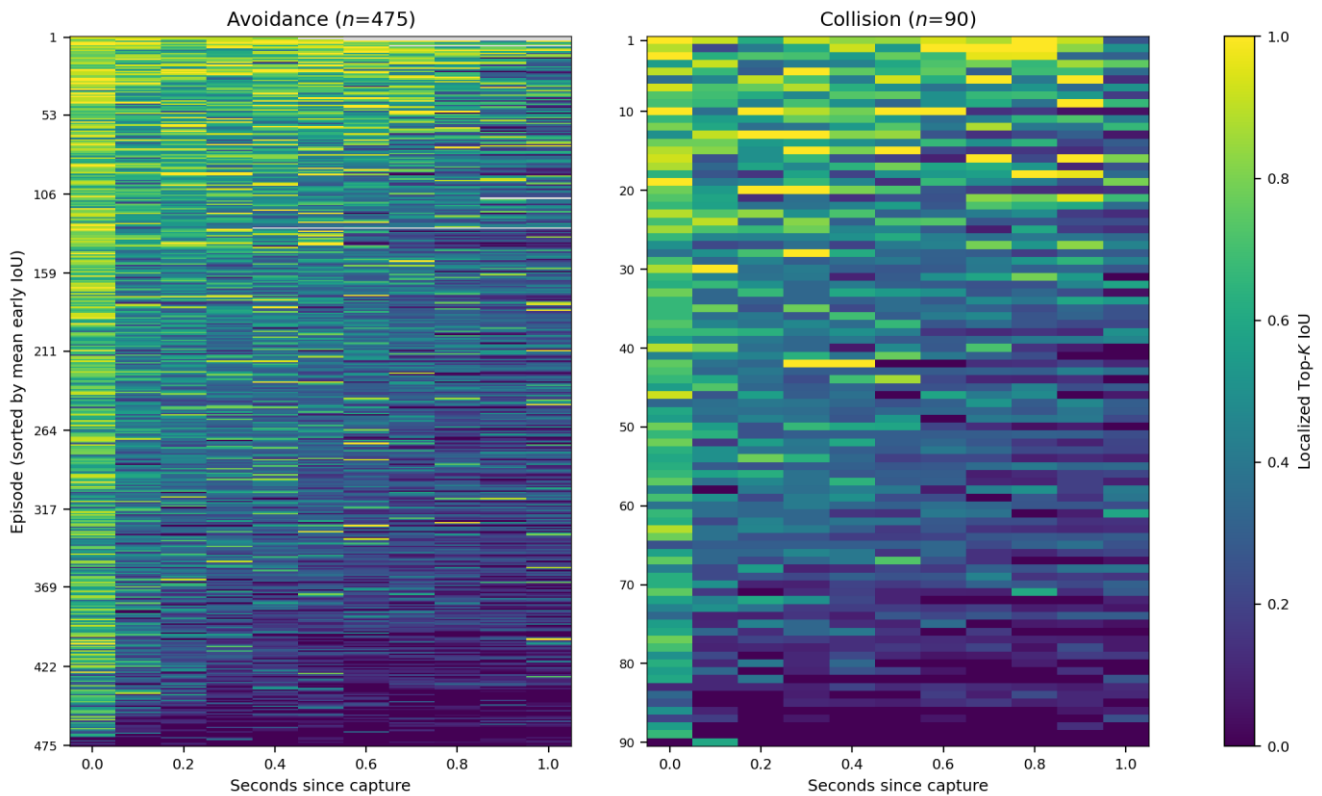


Fig. 3 Per-episode strict local overlap (IoU) of imagined vs. actual pedestrian segmentation over the first 1.0s after capture. Each row is one episode; each column is one time step (0.1s intervals). Color encodes IoU from 0 (dark, world model misses the pedestrian entirely) to 1 (bright, world model reproduces the pedestrian pixel-for-pixel). Grey cells indicate steps where the pedestrian was no longer visible in ground truth, so no comparison is possible. Left panel: 475 avoidance episodes. Right panel: 90 collision episodes. Within each panel, episodes are sorted top-to-bottom by descending mean early IoU, so the best-imagining episodes appear at the top. High-fidelity and low-fidelity episodes are present in both outcome groups, showing that imagination quality alone does not predict whether the agent avoids or collides.

others. The qualitative analysis (Fig.1) suggests a complementary explanation: the world model often imagines premature lane clearance. Although the pedestrian is stationary and should persist in the lane, the model's imagined future shows the pedestrian signal fading or disappearing, as if the obstacle were no longer there. Such incorrect predictions could lead the agent to select actions that do not account for the pedestrian's continued presence.

These findings are consistent with known limitations of learned world models in representing rare or out-of-distribution entities [22; 19]. The pedestrian occupies a small number of pixels and exhibits motion patterns that are uncommon in the training distribution, making faithful long-horizon prediction difficult. Our event-centered analysis complements full-episode evaluations by focusing on the critical conflict-resolution window.

From a practical perspective, these observations suggest that improving pedestrian safety in world-model agents may require more than increasing reconstruction fidelity. If the model can partially reconstruct the pedestrian at the capture step but imagines it disappearing shortly after, then the policy will plan around an incorrect future (one in which the lane is clear) regardless of initial pixel quality. Several directions could address this failure mode: (1) an *object-persistence prior* that biases latent dynamics toward maintaining detected static obstacles across the imagination horizon unless an explicit removal event is predicted; (2) *targeted data augmentation* during world-model training, such as oversampling pedestrian-blocking episodes, to better represent rare but safety-critical scenarios; (3) a *hazard-aware prediction head* trained alongside the decoder to explicitly forecast binary pedestrian occupancy at each imagined step, providing the actor with an auxiliary safety signal; and (4) *hybrid planning* that combines learned imagination with a lightweight rule-based safety layer (e.g.,

a time-to-collision check on the decoded depth channel) as a fallback when imagination is unreliable.

Several limitations apply. Our analysis uses a single scenario type (protester archetype) and one trained model. The sample size (565 encounters, 90 collisions) is substantial, but the strict metric is intentionally conservative. A larger-scale study with multiple pedestrian archetypes and relaxed metrics would provide a more complete picture. Additionally, the analysis is conducted in decoded observation space; the latent representations used for actual policy decisions may encode pedestrian information differently than what the decoder reveals.

5. Conclusion

We presented an event-centered analysis of imagination quality in a DreamerV3 driving agent facing a stationary pedestrian blocking the lane. Under a strict localized fidelity metric, the world model produces only weak pedestrian support in imagined futures, and this weakness does not clearly separate collision from avoidance outcomes. Qualitative inspection reveals that the more informative failure is imagined lane clearance: the model predicts that the pedestrian obstacle will disappear rather than persist. These findings highlight the importance of examining not only whether a world model retains object presence at the moment of capture, but whether it maintains that presence across the imagination horizon used for planning.

Acknowledgments

This work was funded by the COMET K2 - Competence Centers for Excellent Technologies Programme of the Austrian Federal Ministry for Transport, Innovation and Technology (bmvit), the Austrian Federal Ministry for Digital, Business and Enterprise

(bmdw), the Austrian Research Promotion Agency (FFG), the Province of Styria, and the Styrian Business Promotion Agency (SFG).

References

- [1] Danijar Hafner, Timothy P. Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. *ArXiv*, abs/1912.01603, 2020.
- [2] B Ravi Kiran, Ibrahim Sobh, Victor Talpaert, Patrick Mannion, Ahmad A. Al Sallab, Senthil Kumar Yogamani, and Patrick Pérez. Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 23:4909–4926, 2021.
- [3] Wilko Schwarting, Javier Alonso-Mora, and Daniela Rus. Planning and decision-making for autonomous vehicles. *Annual Review of Control, Robotics, and Autonomous Systems*, 1:187–210, 2018.
- [4] Danijar Hafner, Jurgis Pašukonis, Jimmy Ba, and Timothy P. Lillicrap. Mastering diverse domains through world models. *ArXiv*, abs/2301.04104, 2023.
- [5] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, pages 1–16, 2017.
- [6] Marin Toromanoff, Emilie Wirbel, and Fabien Moutarde. End-to-end model-free reinforcement learning for urban driving using implicit affordances. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7153–7162, 2020.
- [7] Dian Chen, Brady Zhou, Vladlen Koltun, and Philipp Krähenbühl. Learning by cheating. In *Proceedings of the Conference on Robot Learning*, volume 100 of *Proceedings of Machine Learning Research*, pages 66–75. PMLR, 2020.
- [8] Neha Sharma, Chhavi Dhiman, and S. Indu. Pedestrian intention prediction for autonomous vehicles: A comprehensive survey. *Neurocomputing*, 508:120–152, 2022.
- [9] Amir Rasouli and John K. Tsotsos. Autonomous vehicles that interact with pedestrians: A survey of theory and practice. *IEEE Transactions on Intelligent Transportation Systems*, 21:900–918, 2018.
- [10] Tirthankar Bandyopadhyay, Kok Sung Won, Emilio Frazzoli, David Hsu, Wee Sun Lee, and Daniela Rus. Intention-aware motion planning. In *Workshop on the Algorithmic Foundations of Robotics*, 2013.
- [11] Fanta Camara, Nicola Bellotto, et al. Pedestrian models for autonomous driving part ii: High-level models of human behavior. *IEEE Transactions on Intelligent Transportation Systems*, 22:5453–5472, 2020.
- [12] Qifeng Li, Xiaosong Jia, Shaobo Wang, and Junchi Yan. Think2drive: Efficient reinforcement learning by thinking in latent world model for quasi-realistic autonomous driving (in carla-v2). *ArXiv*, abs/2402.16720, 2024.
- [13] Dechen Gao, Shuangyu Cai, Hanchu Zhou, Hang Wang, Iman Soltani, and Junshan Zhang. Cardreamer: Open-source learning platform for world-model-based autonomous driving. *IEEE Internet of Things Journal*, 12:2866–2875, 2024.
- [14] Tuo Feng, Wenguan Wang, and Yi Yang. A survey of world models for autonomous driving. *ArXiv*, abs/2501.11260, 2025.
- [15] Zeyu Zhu and Huijing Zhao. A survey of deep rl and il for autonomous driving policy learning. *IEEE Transactions on Intelligent Transportation Systems*, 23:14043–14065, 2021.
- [16] Pranav Singh Chib and Pravendra Singh. Recent advancements in end-to-end autonomous driving using deep learning: A survey. *IEEE Transactions on Intelligent Vehicles*, 9:103–118, 2023.
- [17] Elahe Delavari, Feeza Khan Khanzada, and Jaerock Kwon. A comprehensive review of reinforcement learning for autonomous driving in the carla simulator. *ArXiv*, abs/2509.08221, 2025.
- [18] Shreyas Ramakrishna, Baiting Luo, Christopher B. Kuhn, Gabor Karsai, and Abhishek Dubey. Anti-carla: An adversarial testing framework for autonomous vehicles in carla. *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, pages 2620–2627, 2022.
- [19] Christopher Diehl, Timo Sievernich, Martin Krüger, Frank Hoffmann, and Torsten Bertram. Uncertainty-aware model-based offline reinforcement learning for automated driving. *IEEE Robotics and Automation Letters*, 8:1167–1174, 2023.
- [20] Oleh Rybkin, Chuning Zhu, Anusha Nagabandi, Kostas Daniilidis, Igor Mordatch, and Sergey Levine. Model-based reinforcement learning via latent-space collocation. *ArXiv*, abs/2106.13229, 2021.
- [21] Danijar Hafner, Timothy P. Lillicrap, Mohammad Norouzi, and Jimmy Ba. Mastering atari with discrete world models. *ArXiv*, abs/2010.02193, 2021.
- [22] Fan-Ming Luo, Tian Xu, Hang Lai, Xiong-Hui Chen, Weinan Zhang, and Yang Yu. A survey on model-based reinforcement learning. *Science China Information Sciences*, 67, 2022.