Interpreting the Planning and Reasoning Ability of Imitation Learning in Autonomous Driving

Hyeon-Chang Jeon[1], Kyung-Beom Kim[2], Eugene Vinitsky[3], and Kyung-Joong Kim[2]

***Abstract*— The design of Autonomous Vehicles (AVs) have made significant progress through imitation of large scale open source datasets. Intuitively, AVs learn not only to act based on the current state, but also to anticipate future behavior—a process known as *planning*. Similarly, they predict the future actions of surrounding vehicles, which we refer to as *reasoning*. In this paper, we investigate the emergence of planning and reasoning abilities in transformer-based imitation learning (IL) where as a proxy we investigate the representations and predictive capabilities of the IL model with respect to other vehicles. In particular, we investigate whether it is a necessary precondition for internal layers to learn to predict the trajectories of other drivers in the scene. To evaluate this, we utilize linear probing for the future position of the ego vehicles and the future position of other vehicles. Our results show that planning converges at some period, but reasoning ability is gradually increasing as the dataset size increases. For propagation of planning and reasoning information, there is no difference in dataset size, but all models lose more information in the further future.**

I. INTRODUCTION

With the release of diverse open source datasets [1], [2] and GPU-accelerated simulation environments [3], [4], data-driven autonomous driving models have made significant strides in areas such as trajectory prediction [5], trajectory generation [6], and decision making [7]. In particular, data-driven imitation learning (IL) models aim to learn human driving behavior, including planning and reasoning abilities.

However, previous studies remain limited as to whether the IL truly possesses planning ability and the ability to reason with other vehicles. Basically, in a multi-agent autonomous driving scenario where multiple vehicles interact, the IL model with planning ability should be able to drive effectively regardless of the presence of other vehicles on the same road. It should efficiently reason about other vehicles based on their presence and, finally, modify its initial planning based on that reasoning.

Moreover, it remains underexplored whether planning and reasoning abilities follow data scaling laws with respect to dataset size and model capacity, even in large-scale open-source datasets. Understanding what abilities agents learn as the dataset grows is a critical question when training data-driven autonomous driving models.

[1]AI Graduate School, Gwangju Institute of Science and Technology, 123 chem-dan-gwa-gi-ro, Gwangju, Republic of Korea kevinjeon119@gm.gist.ac.kr

[2] School of Integrated Technology, Gwangju Institute of Science and Technology, 123 chem-dan-gwa-gi-ro, Gwangju, Republic of Korea kjkim@gist.ac.kr

[3] Urban and Civil Engineering, New York University, 6 MetroTech Center, New York, USA vinitsky.eugene@gmail.com

Corresponding Author: Kyung-Joong Kim kjkim@gist.ac.kr

Thus, a key question arises: Do current IL models possess these capabilities? Furthermore, understanding at what scale of the dataset these abilities emerge is crucial to comprehending the generalization gap of the model. In this paper, we validate the IL models' planning and reasoning abilities through a key research question: **How does the IL model learn planning ability and reasoning ability as the dataset size increases?** To evaluate this, we perform linear probing methods [8] and dataset scaling experiments [9]. First, we train models on datasets of varying sizes and evaluate their performance in simulation. We measure planning-related metrics (goal rate, goal progress ratio) and reasoning-related metric (collision rate). Next, we conduct a vehicle padding experiment, where surrounding vehicles are randomly added with a fixed probability. This evaluates how well the IL model generalizes its planning performance regardless of the presence of other vehicles. Finally, we apply linear probing, a common approach in interpretable deep learning. We extract representations from both early and later layers of the IL model and assess how well they capture planning and reasoning capabilities. For comparison, we also perform linear probing directly on the raw input. From our experimental results, we observe the following:

- With smaller datasets, the model primarily learns planning-related abilities; as the dataset size increases, it gradually acquires reasoning-related abilities.
- The accuracy of both linear probing of ego-autonomous vehicles (AVs) future positions and linear probing of other vehicles' future positions follow a similar trend with dataset scaling experiments.
- Regardless of the dataset size, the model has a representation of planning and reasoning up to the final layer, but the model loses further future information.

II. RELATED WORK

A. Planning in Autonomous Driving

Data-driven autonomous driving has advanced with datasets such as nuScenes [2], Waymo Open Motion Dataset (WOMD) [1], and Argoverse [10], along with simulators like Waymax [4], GPUDrive [3], and Nocturne [11]. Based on these resources, various models have emerged, ranging from simple behavior cloning models [11] to trajectory prediction algorithms like Wayformer [5] and MultiPath [12]. These models share the common characteristic of making accurate planning predictions while utilizing information about surrounding vehicles. **However, research analyzing how these models specifically perform planning and reasoning about other vehicles remains limited.**

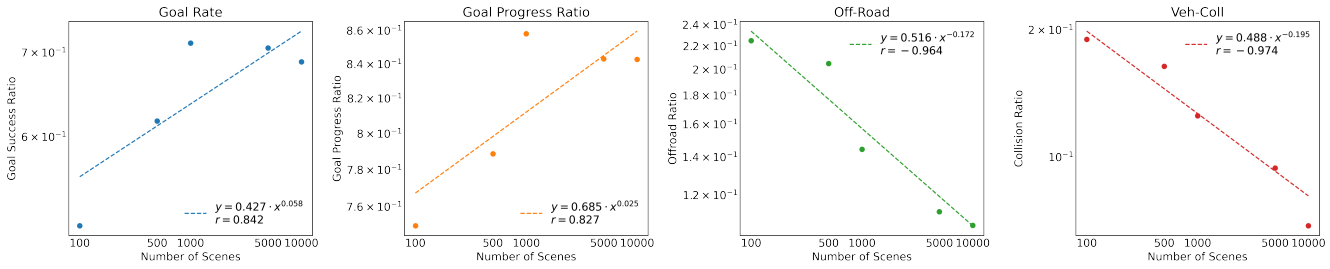

Fig. 1. **Power law relationships for simulation results**: Simulation results for IL with data scaling. We evaluated the goal and collision metrics. Dashed lines are power-law fits and r is correlation coefficient.

B. Interpretability in Deep Learning

Interpretable deep learning is an approach that transforms the representations of neural networks into an interpretable form to analyze their internal representations. Linear probing is a method used to verify how much information a trained neural network has learned from its original dataset by employing a probing dataset and a linear classifier [8]. Recently, the study [13] explored the planning ability of model-free reinforcement learning (RL) using both linear probing and concept-based interpretability [14]. This research analyzed whether the Deep Repeated ConvLSTM [15] agent possesses planning ability through linear probing and intervention methods. Our research follows a similar approach but differs in that we focus on multi-agent settings so that **our probing targets not only the ego agent's planning ability but also its reasoning ability regarding other vehicles**.

C. Data Scaling Laws

With the emergence of large foundation models, scaling laws have become a key research area in domains such as language [9], vision [16], and robotics [17]. In general, scaling laws analyze how model performance varies with dataset size, model capacity, and computing resources, especially for transformer-based architecture. This concept has also recently been extended to the field of autonomous driving [18]. Similarly, we investigate how the performance of models changes with varying dataset sizes in autonomous driving. However, unlike previous work that focuses primarily on overall performance metrics, we take a distinct approach by **analyzing how planning and reasoning capabilities evolve through the lens of interpretability.**

III. Approaches

In this section, we evaluate the reasoning and planning abilities of a transformer-based IL model. We begin with a data scaling experiment to examine how goal achievement and collision rates vary with dataset size (Section III-B). To assess planning ability, we introduce randomly padded trajectories of other active AVs and analyze how the goal related metrics change with varying dataset sizes (Section III-C). Lastly, we focus on two key questions: 1) Does the IL model possess representational capacity for future information (planning and reasoning)? and 2) Are these representations preserved and propagated to the final layer? To evaluate this, we perform linear probing on the ego AV's future positions to determine whether the model's internal representations capture planning behavior (Section III-D). For reasoning ability, we conduct linear probing on the future positions of surrounding AVs to assess whether the model understands and encodes other vehicles' behavior (Section III-E).

A. Experimental Setup

Dataset. We train our model using the WOMD. The dataset consists of over 100K driving scenes, each containing 9 seconds of trajectories sampled at 10 Hz. Each observation includes information about the ego vehicle, surrounding objects, and the map. As action labels are not available in the dataset, we derive them using a delta dynamics model. The data is structured such that each sample contains the past 5 steps as input and the action at the current timestep as the prediction target.

Imitation Learning. For evaluating our method, we employed the transformer-based IL model. Basically, we used the early fusion attention first for encoding layers and self-attention vehicles and road objects separately. Then, we used the cross-attention for using other vehicles' features and road object features related to ego AV. Lastly, we used the gaussian mixture model (GMM) for extracting actions (Δx, Δy, and Δyaw).

B. IL with Different Dataset Size

TABLE I

Driving performance metrics for different dataset scales.

Num Scenes	Dataset	Goal Rate	Goal Progress Ratio	Off-Road	Veh-Coll
100	Train	0.565 ± 0.116	0.785 ± 0.076	0.155 ± 0.048	0.120 ± 0.038
	Test	0.510 ± 0.039	0.750 ± 0.042	0.225 ± 0.053	0.190 ± 0.035
500	Train	0.648 ± 0.034	0.800 ± 0.044	0.189 ± 0.055	0.134 ± 0.029
	Test	0.617 ± 0.042	0.789 ± 0.042	0.205 ± 0.041	0.164 ± 0.025
1000	Train	0.707 ± 0.070	0.858 ± 0.036	0.132 ± 0.021	0.117 ± 0.014
	Test	**0.712 ± 0.057***	**0.858 ± 0.031***	0.144 ± 0.014	0.125 ± 0.006
5000	Train	0.697 ± 0.103	0.837 ± 0.052	0.127 ± 0.006	0.106 ± 0.014
	Test	0.706 ± 0.096	0.843 ± 0.048	0.112 ± 0.008	0.094 ± 0.012
10000	Train	0.684 ± 0.036	0.841 ± 0.020	0.116 ± 0.010	0.082 ± 0.008
	Test	0.688 ± 0.051	0.843 ± 0.022	**0.106 ± 0.010***	**0.069 ± 0.008***

We trained models using four different random seeds while gradually increasing the dataset size to $\{100, 500, 1000, 5000, 10000\}$ scenes. After training, we evaluated the models in the GPUDrive simulator [3] on both train and test scenes using four key metrics: goal rate, vehicle collision rate (Veh-Coll), off-road rate (Off-Road),

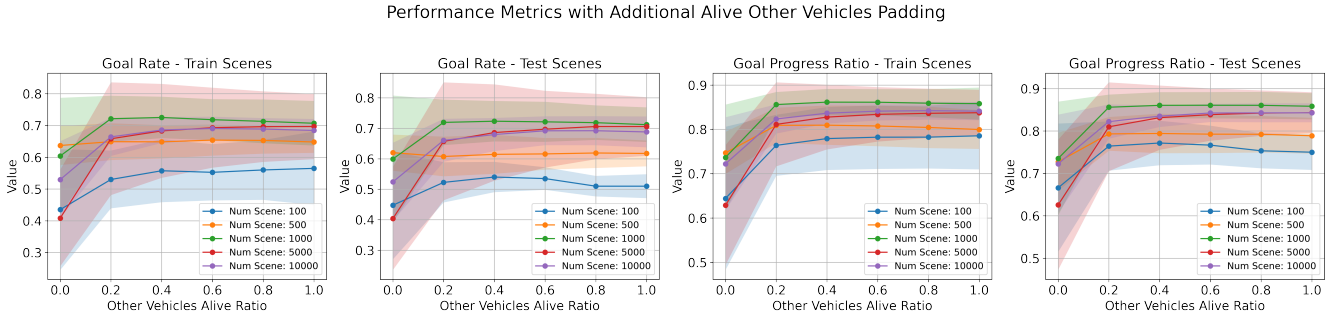

Fig. 2. **Performance metrics with changing partner alive ratio**: We tested the IL model in GPUDrive simulation by padding the other vehicles with ratio $\{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$. For the test set, we evaluated 1,000 scenes that are unseen in the training IL model.

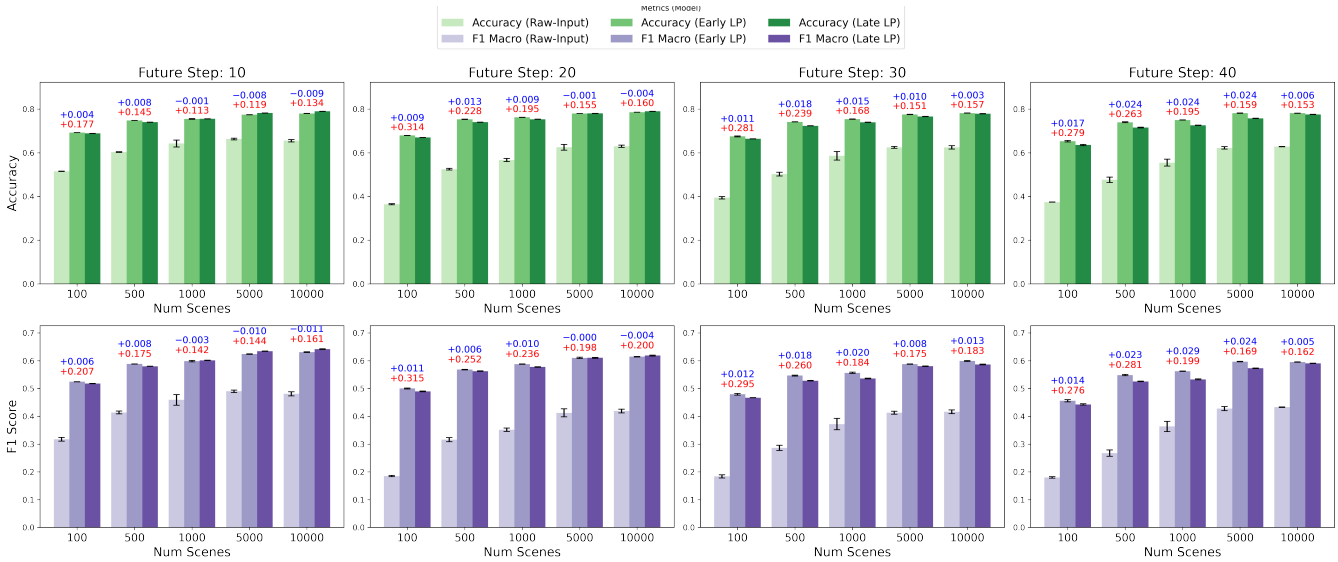

Fig. 3. **Performance metrics with ego-linear probing**: Light colors are raw-input probing and darker colors are IL probing (earlier layer and late layer) with error bars of standard deviation across different seeds. For convenience, we decided to call the linear probing of the IL model an *early LP*. Similarly, the linear probing of the IL model is called a *late LP*. Green color is future position accuracy and purple color is F1 score of future position. We chose the future steps for $\{10, 20, 30, 40\}$ which is 1 to 4 seconds ahead. **The red text** indicates the difference between the raw-input probing and the best IL probing model (early LP and late LP). **The blue text** indicates the difference between early LP and late LP (early LP - late LP).

and goal progress rate (the distance from goal at final timestep), as shown in Table I. The asterisk indicates statistically significant improvement over others. The results show a consistent decrease in collision rates up to 10,000 scenes in test scenes, indicating improved reasoning about surrounding agents. In contrast, the goal progress rate converges to around 85% with 1,000 scenes and remains stable thereafter.

We also performed linear model fitting on these metrics, as illustrated in Figure 1. The collision-related metrics (Veh-Coll and Off-Road) exhibit a strong correlation with dataset size, suggesting a potential power-law relationship. In comparison, goal-related metrics show weaker correlations, likely due to early convergence after 1,000 scenes.

C. Randomly Padding Other Vehicles

If the IL model possesses planning abilities, it should be able to generate a successful trajectory to the goal even in the absence of other vehicles in the scene. Furthermore, as its reasoning ability improves, it should consistently generate effective plans regardless of the number of surrounding vehicles. To evaluate this, we conducted experiments by randomly padding the number of active vehicles on the road with values from $\{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$, as shown in Figure 2.

The results revealed a notable performance gap between scenes with and without other vehicles. Although the IL model trained on 500 scenes achieved a relatively high goal rate, its goal progress ratio remained lower in the 0.0 setting compared to scenes with vehicles (0.2–1.0). Overall, these findings suggest that the IL model still struggles with planning when operating without the presence of other vehicles.

D. Linear Probing for Ego AV Planning

To gain a deeper understanding of the model's planning ability, we conducted a linear probing experiment on future position prediction. The ground truth was set as the ego

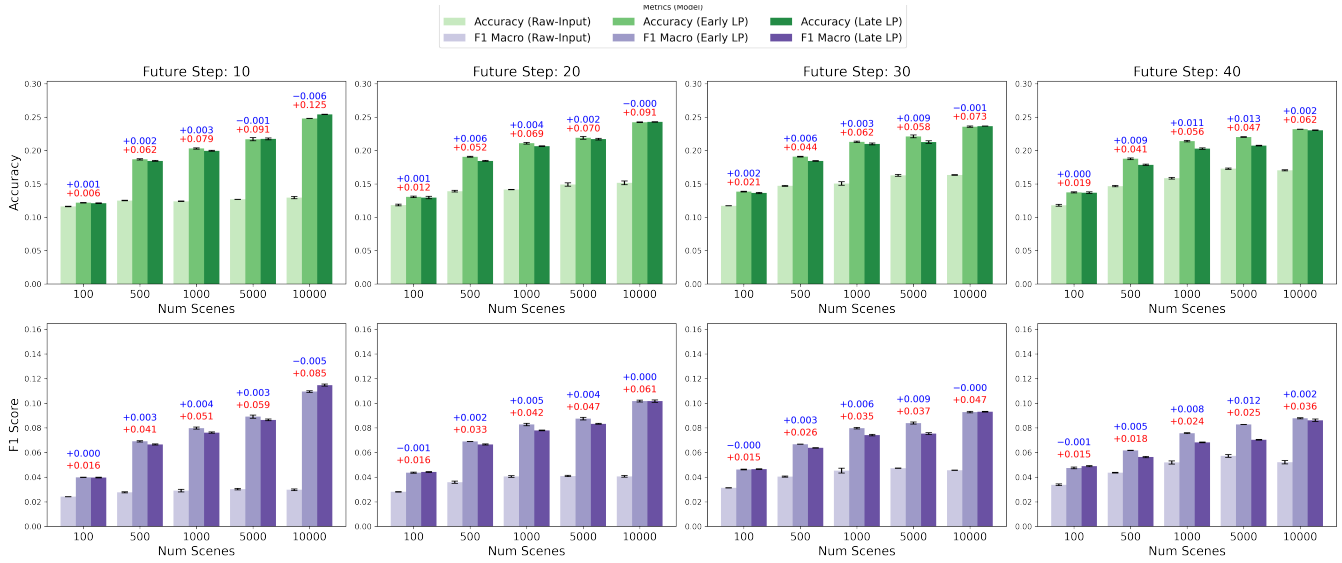

Fig. 4. **Performance metrics with other-linear probing**: Light colors are raw-input probing and darker colors are IL probing(earlier layer and late layer) with error bars of standard deviation across different seeds. Green color is future position accuracy and purple color is F1 score of future position. We chose the future steps for $\{10, 20, 30, 40\}$ which is 1 to 4 seconds ahead. **The red text** indicates the difference between the raw-input probing and the best IL probing model (early LP and late LP). **The blue text** indicates the difference between early LP and late LP (early LP - late LP).

vehicle's positions 1 to 4 seconds ahead (corresponding to 10 to 40 steps), and the positions were discretized into 64 labels (8 along the x-axis and 8 along the y-axis). We trained linear classifiers on the raw input, as well as on representations from both early and late attention layers, and evaluated performance using accuracy and F1 score to account for data imbalance. For hyperparameter, linear classifiers were trained for 20 epochs with a batch size of 256.

As shown in Figure 3, the IL model consistently demonstrated strong representational power for the ego's future trajectories, achieving at least 40% accuracy across all future time steps. Furthermore, the performance gap between raw input probing model and IL probing model has been reduced, suggesting that the IL model has converged in terms of planning capabilities. When comparing early and late attention layers, the overall performance with increasing data was similar, but information loss has occurred further into the future.

E. Linear Probing for Other Vehicles Reasoning

To test reasoning ability, we followed the same probing method and hyperparameter as for the ego vehicle, setting the ground truth as the other vehicles' position from 1 to 4 seconds ahead (10 to 40 steps). The results are shown in the Figure 4. To evaluate the model's reasoning ability, we applied the same linear probing method used for the ego vehicle, setting the ground truth as the future positions of other vehicles from 1 to 4 seconds ahead (10 to 40 steps). For raw-input model, we used the raw input features of other vehicles combined with ego vehicle information. The results are presented in Figure 4.

As shown in the figure, the linear probing performance improves as the dataset size increases, and the gap between the raw-input probing model and the IL probing model results becomes more pronounced. This suggests that the IL model gradually learns to reason about other agents, consistent with the trend observed in Table I. However, overall accuracy is still low, with F1 scores lower than 20%. Similar with ego-linear probing, other vehicle probing results have shown that the IL model throws up the reasoning information as the future step increases.

IV. Conclusion

In this work, we analyzed how the planning and reasoning abilities of transformer-based IL models in autonomous driving evolve with increasing dataset scale. Unlike prior studies on data scaling, we leveraged linear probing to gain deeper insights into how these abilities are represented across model layers. Our experiments revealed that with smaller datasets, the model first learns planning-related ability, and as the dataset increases, it gradually improves its reasoning ability. In the case of information propagation, planning and reasoning representation is keeping propagated, but losing as the predicting further future information.

For future work, although we used 4 million samples, this represents only 10% of the full WOMD dataset, and scaling to the full dataset is needed for comprehensive validation. Furthermore, effective planning should involve not only generating initial plans, but also adapting them based on the predicted trajectories of surrounding agents. We aim to explore this through intervention-based methods in future research. Lastly, reasoning ability is not simply evaluated with uniform accuracy about other vehicles since there are affective others and not.

ACKNOWLEDGMENT

This work was supported by GIST-IREF from Gwangju Institute of Science and Technology(GIST).

REFERENCES

[1] 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*, 2021, pp. 9710–9719.

[2] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 621–11 631.

[3] S. Kazemkhani, A. Pandya, D. Cornelisse, B. Shacklett, and E. Vinitsky, "Gpudrive: Data-driven, multi-agent driving simulation at 1 million fps," *arXiv preprint arXiv:2408.01584*, 2024.

[4] C. Gulino, J. Fu, W. Luo, G. Tucker, E. Bronstein, Y. Lu, J. Harb, X. Pan, Y. Wang, X. Chen *et al.*, "Waymax: An accelerated, data-driven simulator for large-scale autonomous driving research," *Advances in Neural Information Processing Systems*, vol. 36, pp. 7730–7742, 2023.

[5] N. Nayakanti, R. Al-Rfou, A. Zhou, K. Goel, K. S. Refaat, and B. Sapp, "Wayformer: Motion forecasting via simple & efficient attention networks," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 2980–2987.

[6] L. Feng, Q. Li, Z. Peng, S. Tan, and B. Zhou, "Trafficgen: Learning to generate diverse and realistic traffic scenarios," in *2023 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2023, pp. 3567–3575.

[7] A. Hu, G. Corrado, N. Griffiths, Z. Murez, C. Gurau, H. Yeo, A. Kendall, R. Cipolla, and J. Shotton, "Model-based imitation learning for urban driving," *Advances in Neural Information Processing Systems*, vol. 35, pp. 20 703–20 716, 2022.

[8] G. Alain and Y. Bengio, "Understanding intermediate layers using linear classifier probes," *arXiv preprint arXiv:1610.01644*, 2016.

[9] T. Henighan, J. Kaplan, M. Katz, M. Chen, C. Hesse, J. Jackson, H. Jun, T. B. Brown, P. Dhariwal, S. Gray *et al.*, "Scaling laws for autoregressive generative modeling," *arXiv preprint arXiv:2010.14701*, 2020.

[10] M.-F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan *et al.*, "Argoverse: 3d tracking and forecasting with rich maps," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 8748–8757.

[11] E. Vinitsky, N. Lichtlé, X. Yang, B. Amos, and J. Foerster, "Nocturne: a scalable driving benchmark for bringing multi-agent learning one step closer to the real world," *Advances in Neural Information Processing Systems*, vol. 35, pp. 3962–3974, 2022.

[12] Y. Chai, B. Sapp, M. Bansal, and D. Anguelov, "Multipath: Multiple probabilistic anchor trajectory hypotheses for behavior prediction," in *Conference on Robot Learning*. PMLR, 2020, pp. 86–99.

[13] T. Bush, S. Chung, U. Anwar, A. Garriga-Alonso, and D. Krueger, "Interpreting emergent planning in model-free reinforcement learning," in *The Thirteenth International Conference on Learning Representations*, 2025.

[14] B. Kim, M. Wattenberg, J. Gilmer, C. Cai, J. Wexler, F. Viegas *et al.*, "Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (tcav)," in *International conference on machine learning*. PMLR, 2018, pp. 2668–2677.

[15] A. Guez, M. Mirza, K. Gregor, R. Kabra, S. Racanière, T. Weber, D. Raposo, A. Santoro, L. Orseau, T. Eccles *et al.*, "An investigation of model-free planning," in *International conference on machine learning*. PMLR, 2019, pp. 2464–2473.

[16] K. Tian, Y. Jiang, Z. Yuan, B. Peng, and L. Wang, "Visual autoregressive modeling: Scalable image generation via next-scale prediction," *Advances in neural information processing systems*, vol. 37, pp. 84 839–84 865, 2024.

[17] H. Bharadhwaj, J. Vakil, M. Sharma, A. Gupta, S. Tulsiani, and V. Kumar, "Roboagent: Generalization and efficiency in robot manipulation via semantic augmentations and action chunking," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 4788–4795.

[18] Y. Zheng, Z. Xia, Q. Zhang, T. Zhang, B. Lu, X. Huo, C. Han, Y. Li, M. Yu, B. Jin *et al.*, "Preliminary investigation into data scaling laws for imitation learning-based end-to-end autonomous driving," *arXiv preprint arXiv:2412.02689*, 2024.