

Self-Supervised Representation Learning for Large-Scale Vehicle Motion Prediction

Oswin Hentschel

Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV,
USA.

hentscheloswin@unr.edu

Abstract

Large-scale vehicle motion prediction has become a central problem for autonomous driving, intelligent transportation, and safety-critical urban mobility. The field increasingly depends on representation learning that can exploit massive unlabeled or weakly labeled trajectory corpora without requiring exhaustive manual annotation. This paper examines self-supervised representation learning as a system-level strategy for vehicle motion prediction at scale. It argues that the principal challenges are not only model architecture but also data infrastructure, curation, evaluation, deployment, governance, robustness, fairness, and long-term sustainability. Beginning from conceptual foundations in transformer architectures, masked modeling, contrastive learning, and redundancy reduction, the discussion develops an integrated view of how self-supervised objectives can structure multi-agent interaction, map context, temporal dependency, and uncertainty. The paper then analyzes architectural trade-offs among centralized, hierarchical, and scene-centric representations, and it considers how data pipelines, benchmark design, and density-balanced evaluation shape model behavior. Particular attention is given to the tension between generalization and operational reliability, because motion prediction systems must perform under distribution shift, sensor degradation, edge cases, and heterogeneous traffic norms. The paper further assesses deployment constraints involving latency, memory, fleet-scale inference, and continuous learning, as well as sustainability concerns related to compute, data storage, and lifecycle maintenance. Fairness and policy implications are treated as first-order engineering concerns rather than external constraints, especially where predictive systems allocate risk across pedestrians, cyclists, and underserved communities. Through case illustrations from autonomous driving datasets and cross-domain comparisons with language and vision pretraining, the paper outlines a research agenda for self-supervised motion prediction that is technically rigorous, institutionally accountable, and socially legible. The conclusion emphasizes that scalable representation learning must be embedded within broader socio-technical infrastructures if it is to support dependable vehicle motion prediction.

Keywords

self-supervised learning; vehicle motion prediction; large-scale systems; representation learning; autonomous driving; data curation; robustness; governance; sustainability.

1. Introduction

Vehicle motion prediction sits at the intersection of machine learning, robotics, transportation engineering, and public policy. In large-scale settings, the task requires anticipating the future trajectories of vehicles, pedestrians, cyclists, and other agents over horizons that may range from a few seconds to several tens of seconds. The difficulty is not merely one of sequence modeling. It involves reasoning about intention, social interaction, road geometry, traffic rules,

vehicle dynamics, occlusion, and uncertainty. Recent advances in attention-based architectures have demonstrated the value of flexible sequence representation [1], while pretraining paradigms in natural language processing have shown that large corpora can support transferable representations without exhaustive task-specific labels [2]. These developments motivate self-supervised learning as a strategy for vehicle motion prediction, where unlabeled trajectory data are abundant but high-quality semantic annotations are expensive, uneven, and often difficult to validate.

The promise of self-supervised representation learning is that models can learn useful structure from raw or lightly processed observations before being fine-tuned for prediction, planning, or risk assessment. Contrastive learning has shown that discriminative representations can emerge from augmented views without manual labels [3], and simpler frameworks have further improved stability and efficiency [4]. Other approaches avoid negative sampling altogether by bootstrapping latent targets [5], while redundancy-reduction methods regularize representations directly to prevent collapse [6]. For motion prediction, these ideas suggest that models might learn invariances to sensor noise, viewpoint changes, sampling rates, and local traffic composition, while preserving information about intent, map structure, and interaction dynamics.

Yet self-supervision in this domain cannot be treated as a direct transfer from vision or language. Vehicle motion is embodied, multi-agent, physically constrained, and socially regulated. A representation that ignores lane topology, right-of-way, or pedestrian crossing behavior may achieve low average error while remaining unsafe in rare but consequential settings. Moreover, large-scale deployment introduces system-level questions that are not resolved by better pretext tasks alone. Data collection, curation, versioning, evaluation, model updating, and accountability mechanisms determine whether representation learning produces dependable infrastructure or fragile experimental artifacts. This paper therefore approaches self-supervised representation learning for vehicle motion prediction as a socio-technical systems problem. It examines architecture, data infrastructure, deployment trade-offs, robustness, fairness, governance, and sustainability, with the goal of articulating both the technical opportunities and the institutional responsibilities that accompany scale.

2. Conceptual Foundations of Self-Supervised Representation Learning

Self-supervised learning refers to a family of methods that generate supervisory signals from the structure of data itself. In language modeling, masked prediction and next-token objectives allow models to acquire syntactic, semantic, and pragmatic regularities from vast corpora [2]. In vision, contrastive objectives encourage representations that bring augmented views of the same instance together while separating different instances [3], [4]. Bootstrapping methods remove the need for explicit negative pairs by predicting latent representations of one view from another [5]. Regularization-based approaches maintain information content through variance, invariance, and covariance terms [6]. Across these traditions, the central premise is that useful abstractions can be learned before labels are introduced, and that downstream tasks benefit from representations that capture stable structure rather than superficial correlations.

For vehicle motion prediction, the relevant structures are relational and temporal. Agents do not move independently. Their futures depend on neighbors, lane configurations, traffic signals, crosswalks, intersections, and cultural norms of negotiation. A self-supervised objective might mask future segments and ask the model to reconstruct them, but such an objective can reward physical plausibility without capturing social contingency. It might

contrast temporally separated windows of the same trajectory, but this may conflate routine following behavior with lane changes. It might learn predictive embeddings of map tiles, but map semantics vary across cities and jurisdictions. The design of pretext tasks therefore becomes a value-laden modeling choice because it determines which invariances are learned and which distinctions are erased.

A robust conceptual foundation must combine invariance with sensitivity. Representations should be stable under irrelevant perturbations such as minor localization noise or temporary occlusion, but sensitive to safety-relevant changes such as pedestrian intention, brake-light activation, or aggressive merging. This distinction is difficult to encode in a single objective. It requires multi-task self-supervision, careful augmentation design, and evaluation that measures not only average displacement error but also calibration, mode coverage, interaction compliance, and tail behavior. In this sense, self-supervised motion prediction is not a single algorithm but an architectural commitment to learning structured priors from large-scale interaction data while preserving the distinctions that operational safety demands.

3. System Architecture for Large-Scale Vehicle Motion Prediction

Architectural choices in motion prediction shape how self-supervised representations are learned, stored, and deployed. Early social pooling methods represented nearby agents through recurrent networks and handcrafted interaction neighborhoods [7]. Generative adversarial approaches introduced multimodal futures and social acceptability constraints [8]. Dynamically feasible forecasting incorporated heterogeneous agent types, map information, and vehicle kinematics [9]. Graph-based lane representations then demonstrated that road topology can be encoded explicitly rather than inferred only from trajectories [10]. Target-driven prediction reframed the problem as selecting and refining candidate goals [11], while scene-centric transformers treated multiple agents and map elements as a unified tokenized scene [12]. Agent-aware attention models further emphasized the interdependence of social and temporal reasoning [13], and hierarchical vector transformers improved efficiency by structuring local and global interactions [14]. Simpler attention-based forecasting architectures have also shown that scale and data quality can outweigh architectural complexity [15].

Self-supervised learning interacts with these architectures in several ways. A scene-centric transformer may be pretrained by masking agent tokens, map tokens, or future time steps. A graph-based model may learn edge representations that predict whether two lanes will interact. A hierarchical model may learn local encodings that remain stable across cities and global encodings that capture intersection-level dynamics. The trade-offs are substantial. Centralized scene models can capture rich interactions but scale poorly with the number of agents and map elements. Hierarchical models improve scalability but may lose long-range dependencies. Agent-centric models simplify batching but can duplicate context and complicate global consistency. For fleet deployment, the architecture must also support low-latency inference, efficient memory use, and graceful degradation when sensor inputs are incomplete.

Large-scale systems therefore require architectural pluralism. A self-supervised backbone may be pretrained centrally on a massive corpus, then distilled into smaller models for vehicle-edge inference. Some representations may be cached at the map level, while dynamic interaction embeddings are computed online. The system must distinguish between representations that are stable over months, such as lane geometry and regional driving norms, and representations that change rapidly, such as weather, congestion, or incident conditions. This temporal stratification is an architectural concern because it determines update frequency,

version compatibility, and rollback safety. Self-supervision can help learn these strata, but only if the system design preserves metadata about where, when, and under what conditions each representation was learned.

4. Data Infrastructure, Curation, and Evaluation

The success of self-supervised motion prediction depends heavily on data infrastructure. Large-scale driving datasets have provided multimodal sensor streams, high-definition maps, tracked agents, and forecasting benchmarks [17], [18], [19]. These datasets enable pretraining and evaluation, but they also embody decisions about geography, sensor configuration, annotation policy, and scenario selection. A dataset collected in a small number of cities may underrepresent rural roads, informal intersections, extreme weather, or vulnerable road users. A benchmark that emphasizes average displacement may reward smooth but socially indifferent predictions. A corpus dominated by uneventful highway driving may produce representations that fail in dense urban negotiation.

Recent work on density-balanced data curation and evaluation for trajectory prediction highlights that data distribution and metric design are not neutral technical details [16]. If rare but safety-critical interactions are drowned out by common following behavior, self-supervised objectives may learn convenient regularities rather than robust interaction priors. Curating data for density balance involves identifying underrepresented agent configurations, interaction intensities, and environmental conditions, then ensuring that pretraining and evaluation reflect those distributions. This is difficult because density itself is multidimensional. It includes agent count, spatial proximity, speed variance, maneuver diversity, and temporal clustering. A system that balances only one dimension may create new blind spots in another.

Evaluation must therefore move beyond a single leaderboard score. Self-supervised representations should be assessed for transfer to downstream tasks such as trajectory prediction, occupancy forecasting, risk scoring, and planning. They should be tested under distribution shift across cities, sensor suites, and traffic regimes. They should be probed for whether they encode socially relevant variables such as right-of-way, yielding behavior, and pedestrian hesitation. They should also be audited for leakage, because self-supervised pretraining on data that overlaps with evaluation scenarios can inflate apparent generalization. Data infrastructure thus becomes an epistemic infrastructure: it determines what the model can know, what the benchmark can reveal, and what the institution can claim with confidence.

5. Training and Deployment Trade-Offs

Training self-supervised models for vehicle motion prediction at scale involves trade-offs among compute, data freshness, representation generality, and operational latency. Large pretraining runs can learn broad priors, but they are expensive and slow to update. Continual pretraining on fleet data can adapt to changing conditions, but it risks catastrophic forgetting, feedback loops, and uncontrolled drift. Fine-tuning on task-specific labels can improve precision, but it may narrow the representation and reduce robustness to rare events. The system must decide which components are centralized, which are distributed, and which are updated only after extensive validation.

Deployment adds further constraints. Inference may occur on-vehicle, at the edge, or in a data center. On-vehicle inference prioritizes latency, power, and safety isolation. Edge inference can aggregate local traffic context but introduces communication dependencies. Data-center inference can use larger models but may be unacceptable for real-time control due to network

variability. A practical architecture may use a pretrained self-supervised encoder to generate compact embeddings that are then consumed by lightweight predictors, planners, or monitors. This separation can improve modularity, but it also creates interface risks: if the embedding omits a safety-relevant distinction, downstream modules may not recover it.

Sustainability concerns intersect with these choices. Training large models consumes significant energy and hardware, while storing multi-sensor data for pretraining creates long-term storage and governance burdens. A system that continuously retrains on all available data may be technically impressive but institutionally unsustainable. More sustainable designs may use selective sampling, data distillation, model pruning, and scheduled retraining. They may also separate representation learning from operational adaptation, so that only high-value updates propagate to vehicles. The key insight is that deployment trade-offs are not merely engineering optimizations. They determine who can afford to build, validate, and maintain large-scale motion prediction systems, and they shape the concentration of capability in a small number of well-resourced organizations.

6. Robustness, Safety, and Uncertainty

Motion prediction is safety-critical because errors can propagate into planning and control. Self-supervised representations can improve robustness by exposing models to diverse unlabeled scenarios, but they can also create hidden failure modes. A representation may be invariant to an augmentation that is actually safety-relevant, such as temporary occlusion of a pedestrian. It may learn shortcuts from map geometry that do not generalize when construction zones alter lane topology. It may produce confident embeddings for out-of-distribution scenes because self-supervised objectives rarely enforce calibrated uncertainty.

Robustness therefore requires both representation design and system-level safeguards. Ensembles, uncertainty heads, conformal prediction, and runtime monitors can help detect when a prediction is unreliable. However, these safeguards depend on representations that preserve information about ambiguity. If the encoder collapses multimodal futures into a single latent mode, downstream uncertainty estimation becomes difficult. Self-supervised objectives should encourage multiple plausible representations without rewarding arbitrary diversity. They should also be evaluated under sensor degradation, missing tracks, adversarial perturbations, and rare interaction patterns. A system that performs well on clean benchmark data but fails under rain, occlusion, or aggressive merging is not ready for large-scale deployment.

Safety also requires operational transparency. When a vehicle motion prediction system informs braking, lane selection, or route planning, engineers and regulators need to understand why a representation produced a particular forecast. Full interpretability may be unrealistic for large neural models, but meaningful traceability is possible through data provenance, scenario retrieval, uncertainty decomposition, and counterfactual evaluation. The goal is not to reduce safety to explanation. It is to build institutions that can investigate failures, attribute causes, and update representations without waiting for repeated incidents. Self-supervised learning can support this process if it is designed with metadata, versioning, and auditability from the outset.

7. Fairness, Governance, and Policy

Fairness in vehicle motion prediction is often framed as a question of model bias, but it is better understood as a socio-technical allocation problem. Predictive systems influence how vehicles distribute risk among pedestrians, cyclists, passengers, and other drivers. If self-

supervised representations are learned predominantly from regions with particular road designs, enforcement patterns, or pedestrian behaviors, they may mispredict in communities with different mobility practices. If evaluation prioritizes vehicle flow over pedestrian safety, the system may optimize for smooth traffic rather than equitable protection. These are governance questions as much as technical ones.

Policy frameworks for autonomous and assisted driving increasingly require risk assessment, data governance, and accountability. Self-supervised learning complicates these frameworks because pretraining data may be large, heterogeneous, and only weakly documented. Regulators may need mechanisms for dataset provenance, representation auditing, and post-deployment monitoring. Municipalities may need channels to report problematic interactions. Operators may need to demonstrate that updates do not silently degrade performance for vulnerable road users. The danger of large-scale pretraining is that it can obscure responsibility behind diffuse data and opaque representations. Governance must therefore make the chain of custody visible, from sensor collection to curation, pretraining, fine-tuning, deployment, and incident review.

Language-model critiques have warned that scale alone does not guarantee fairness, accountability, or social benefit [20]. Similar caution applies to motion prediction. A model trained on massive driving data may reproduce historical biases in policing, road investment, or mobility access. It may perform well for the average driver while failing for children, older adults, or disabled pedestrians whose behavior is underrepresented. Fairness requires more than disaggregated metrics, although such metrics are necessary. It requires participatory design, domain expertise from transportation and urban planning, and policy mechanisms that can constrain deployment when evidence is insufficient. Self-supervised representation learning should be treated as part of public infrastructure when it shapes mobility safety, not merely as a private model artifact.

8. Sustainability and Lifecycle Considerations

The sustainability of large-scale motion prediction systems includes environmental, economic, and institutional dimensions. Environmental sustainability concerns compute energy, data storage, sensor hardware, and fleet power consumption. Economic sustainability concerns the cost of maintaining datasets, retraining models, validating updates, and supporting regulatory compliance. Institutional sustainability concerns whether the organizations responsible for these systems can retain expertise, document decisions, and manage turnover. A self-supervised model that cannot be maintained over years is not truly large-scale infrastructure, regardless of its benchmark performance.

Lifecycle design should therefore treat representations as versioned assets. Each pretrained encoder should have documented data sources, preprocessing steps, augmentation policies, evaluation results, and known limitations. Downstream fine-tuning should record which representation version was used and why. Deployment should include rollback plans, drift detection, and scheduled audits. Data retention should balance the value of long-term pretraining against privacy, storage, and environmental costs. Selective forgetting and data minimization may be technically difficult but institutionally necessary. The most sustainable system is not necessarily the one with the largest model or dataset; it is the one that can be governed, updated, and retired responsibly.

9. Case Illustrations and Cross-Domain Comparisons

Cross-domain comparison with language and vision pretraining clarifies both the potential and the limits of self-supervision for vehicle motion. In language, self-supervised pretraining succeeds because text offers enormous scale, relatively stable symbolic structure, and tasks that can be evaluated through generation and understanding [2]. In vision, contrastive and bootstrapped methods succeed because augmentation can approximate invariances without destroying semantic identity [3], [4], [5]. In motion prediction, augmentation is more dangerous because small changes can alter safety meaning. Shifting a pedestrian slightly may be irrelevant for one prediction horizon but critical for another. Masking a vehicle may simulate occlusion, but it may also remove the very agent whose intention matters most. The pretext task must therefore be grounded in physical and social plausibility, not merely statistical convenience.

Case illustrations from autonomous driving datasets show that scale alone does not resolve these tensions. Multimodal datasets provide rich sensor and map context [17], forecasting benchmarks enable comparative evaluation [18], and interactive motion datasets support multi-agent reasoning [19]. Yet each dataset encodes particular cities, sensor configurations, and annotation priorities. A representation pretrained on one corpus may transfer unevenly to another. This is not a reason to abandon self-supervision. It is a reason to treat transfer, calibration, and curation as central research problems. The most promising systems may combine self-supervised pretraining with targeted human annotation, simulation, and adversarial scenario generation. In that hybrid model, self-supervision provides breadth, while curation and governance provide the boundaries that safety requires.

10. Future Directions

Future research should develop self-supervised objectives that explicitly model interaction, uncertainty, and social norms. Masked trajectory modeling can be extended to masked interaction graphs, where the model must infer missing relational context rather than only missing positions. Contrastive learning can be adapted to compare behaviorally equivalent scenarios across different map representations. Redundancy-reduction methods can be used to preserve multimodal latent structure while preventing collapse. Evaluation should include transfer to planning, risk assessment, and human-robot interaction, not only trajectory error. Benchmarks should report tail performance, calibration, and subgroup analysis across vulnerable road users and diverse environments.

System research should also address federated and privacy-preserving pretraining. Fleet data are distributed across vehicles, regions, and operators. Centralizing all data may be impractical or unacceptable. Federated representation learning could allow models to benefit from distributed experience while limiting raw data movement, but it introduces challenges in heterogeneity, communication cost, verification, and poisoning resistance. Governance research should develop audit mechanisms for self-supervised encoders, including data provenance, representation probing, and scenario-based certification. Policy research should clarify liability when pretrained representations contribute to downstream failures, especially when multiple organizations own different parts of the pipeline. Finally, sustainability research should measure the full lifecycle cost of large-scale motion prediction and identify when smaller, curated, or domain-specific models are preferable to ever-larger pretraining runs.

11. Conclusion

Self-supervised representation learning offers a powerful approach to large-scale vehicle motion prediction because it can exploit abundant unlabeled trajectory and sensor data to learn transferable structure. Its technical promise lies in capturing temporal, spatial, and social regularities before task-specific supervision is applied. However, this paper has argued that the decisive challenges are system-level. Architecture must balance interaction richness with deployability. Data infrastructure must support curation, evaluation, provenance, and distributional balance. Deployment must manage latency, update frequency, and sustainability. Safety requires uncertainty-aware representations and operational safeguards. Fairness and governance require explicit attention to how predictive systems allocate risk and how institutions remain accountable for their effects. Cross-domain comparisons with language and vision clarify that self-supervision is not a universal solution but a design orientation that must be adapted to embodied, multi-agent, and safety-critical domains. The research agenda ahead is therefore interdisciplinary: it must advance representation learning while building the data pipelines, evaluation regimes, governance structures, and lifecycle practices that make large-scale vehicle motion prediction dependable. If these dimensions are treated as integral rather than secondary, self-supervised learning can contribute not only to more accurate prediction but to more responsible mobility infrastructure.

References

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (pp. 5998-6008).
2. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT* (pp. 4171-4186).
3. He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9729-9738).
4. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning* (pp. 1597-1607).
5. Grill, J.-B., Strub, F., Althé, F., Tallec, C., Richemond, P. H., Buchatskaya, E., ... Valko, M. (2020). Bootstrap your own latent: A new approach to self-supervised learning. In *Advances in Neural Information Processing Systems* (pp. 21271-21284).
6. Bardes, A., Ponce, J., & LeCun, Y. (2022). VICReg: Variance-invariance-covariance regularization for self-supervised learning. In *International Conference on Learning Representations*.
7. Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (2016). Social LSTM: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 961-971).
8. Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S., & Alahi, A. (2018). Social GAN: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2255-2264).

9. Salzmann, T., Ivanovic, B., Chakravarty, P., & Pavone, M. (2020). Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. In *Computer Vision – ECCV 2020* (pp. 683-700). Springer.
10. Liang, M., Yang, B., Hu, R., Chen, Y., Liao, R., Feng, S., & Urtasun, R. (2020). Learning lane graph representations for motion forecasting. In *Computer Vision – ECCV 2020* (pp. 541-556). Springer.
11. Zhao, H., Gao, J., Lan, T., Sun, C., Sapp, B., Varadarajan, B., Shen, Y., Shen, Y., Chai, Y., Schmid, C., Li, C., & Anguelov, D. (2021). TNT: Target-driven trajectory prediction. In *Proceedings of the 5th Conference on Robot Learning* (pp. 895-904). PMLR.
12. Ngiam, J., Caine, B., Vasudevan, V., Zhang, Z., Chiang, H.-T. L., Ling, J., ... Anguelov, D. (2022). Scene transformer: A unified architecture for predicting future trajectories of multiple agents. In *International Conference on Learning Representations*.
13. Shi, S., Wang, Z., & Li, H. (2021). AgentFormer: Agent-aware transformers for socio-temporal multi-agent forecasting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9813-9823).
14. Zhou, Z., Ye, L., Wang, J., Wu, K., & Lu, K. (2022). HiVT: Hierarchical vector transformer for multi-agent motion prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8823-8833).
15. Nayakanti, N., Al-Rfou, R., Zhou, A., Goel, K., Refaat, K. S., & Sapp, B. (2023). Wayformer: Motion forecasting via simple and efficient attention networks. In *2023 IEEE International Conference on Robotics and Automation* (pp. 2980-2987). IEEE.
16. Yang, R., Xu, Y., Fu, Y., & Su, L. (2026). Den-tp: A density-balanced data curation and evaluation framework for trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10632-10641).
17. Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., ... Beijbom, O. (2020). nuScenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11621-11631).
18. Chang, M.-F., Lambert, J., Sangkloy, P., Singh, J., Bak, S., Hartnett, A., ... Hays, J. (2019). Argoverse: 3D tracking and forecasting with rich maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8748-8757).
19. Ettinger, S., Cheng, S., Caine, B., Liu, C., Zhao, H., Pradhan, S., ... Anguelov, D. (2021). Large scale interactive motion forecasting for autonomous driving: The Waymo open motion dataset. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9710-9719).
20. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623).