

Data-Centric Optimization of Autonomous Driving Trajectory Prediction via Intelligent Dataset Generation

Iisakki Gronroos

Department of Computer Science, Binghamton University, Binghamton, NY, USA.
iisakkimail@binghamton.edu

Gijs Rijcken

Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL,
USA.
gijs523@uab.edu

Rigo Koster

Department of Computer Science, University of North Texas, Denton, TX, USA.
rigokoster201@unt.edu

Annette Salazar

School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.
annette2005@uc.edu

Abstract

Autonomous driving trajectory prediction has largely advanced through model-centric innovations, yet its real-world reliability remains constrained by data scarcity, imbalance, distribution shift, and weak representativeness. This paper examines data-centric optimization of trajectory prediction through intelligent dataset generation as a system-level strategy rather than a narrow preprocessing technique. It argues that dataset generation must be treated as an infrastructural capability that couples perception, simulation, curation, validation, governance, and deployment feedback. By synthesizing literature on driving datasets, motion forecasting, simulation, data-centric artificial intelligence, and safety governance, the paper identifies structural trade-offs among realism, coverage, cost, fairness, and maintainability. It further analyzes architectural implications for autonomous driving stacks, including how generated data can support rare-event augmentation, interactive scenario stress testing, and continuous lifecycle improvement. Special attention is given to robustness, fairness, sustainability, and policy because generated datasets can either amplify or mitigate long-standing asymmetries in road-user representation. The paper also considers evaluation regimes that move beyond aggregate error metrics toward density balance, scenario validity, and socio-technical accountability. The central claim is that intelligent dataset generation should be governed as a public-interest infrastructure, with transparent provenance, versioning, auditing, and cross-institutional coordination. Without such governance, data-centric gains may remain brittle and difficult to certify. With it, trajectory prediction can become more adaptable, equitable, and deployable across diverse urban and regulatory contexts. The conclusion outlines a research agenda spanning synthetic-real hybrid pipelines, simulation fidelity, evaluation pluralism, and lifecycle sustainability.

Keywords

autonomous driving, trajectory prediction, data-centric artificial intelligence, dataset generation, simulation, robustness, fairness, governance, infrastructure, safety.

1. Introduction

Autonomous driving systems depend on the ability to anticipate the future motion of vehicles, pedestrians, cyclists, and other road users. Trajectory prediction is therefore not a peripheral perception task but a core component of planning, risk assessment, and vehicle-to-infrastructure coordination. Early progress in this field was closely tied to the release of large-scale driving datasets such as KITTI, nuScenes, Argoverse, the Interaction dataset, and the Waymo Open Motion Dataset [1][2][3][4][5]. These resources enabled reproducible benchmarking and accelerated the development of deep forecasting architectures. However, the field has increasingly recognized that model architecture alone does not determine deployment performance. The distribution, balance, annotation quality, geographic coverage, and temporal validity of data often shape generalization more decisively than incremental architectural refinement.

This recognition has given rise to data-centric artificial intelligence, an orientation that treats data quality, curation, and lifecycle management as first-class engineering concerns [11]. In autonomous driving, data-centricity is especially consequential because the operational environment is open-ended, safety-critical, and socially contested. A trajectory prediction model trained on dense urban traffic may fail in suburban intersections, low-light conditions, construction zones, or mixed-traffic environments. A model optimized on average displacement error may appear strong while systematically underestimating rare but dangerous interactions involving children, wheelchair users, or emergency vehicles. These failures are not merely statistical; they become questions of infrastructure design, regulatory compliance, and public trust.

Intelligent dataset generation offers one response to these challenges. It encompasses synthetic scene generation, scenario mining, rare-event augmentation, density balancing, simulation-based rendering, and automated curation. Rather than simply adding more data, intelligent generation seeks to produce data that is diagnostically useful, representative of under-served conditions, and aligned with the operational design domain of a deployed system. This paper develops a system-level account of that proposition. It asks how intelligent dataset generation can optimize trajectory prediction while remaining robust, fair, sustainable, and governable. The discussion is organized around architecture, infrastructure, evaluation, policy, and lifecycle concerns. The central argument is that data-centric optimization should not be reduced to a pipeline component. It must be institutionalized as a socio-technical capability that connects simulation, field operations, auditing, and public accountability.

2. From Model-Centric to Data-Centric Autonomy

Model-centric trajectory prediction has produced impressive benchmarks, yet its limitations become visible when systems move from curated test sets to open roads. Planning and decision-making for autonomous vehicles require predictions that are calibrated, interpretable, and responsive to interaction dynamics [6]. Surveys of motion prediction and risk assessment have long noted that uncertainty, multi-modality, and context dependence make trajectory forecasting fundamentally different from static object detection [7]. A vehicle does not merely need to know where a pedestrian is; it must reason about intention, social norms, and the possible futures that emerge from mutual adaptation. These qualities are difficult to learn from datasets that underrepresent negotiation, hesitation, and rare behavioral patterns.

Data-centric autonomy responds by shifting attention to the provenance and composition of training and evaluation data. Vulnerable road user perspectives, for example, reveal that autonomous vehicles must be trusted by those outside the vehicle, not only by passengers or engineers [8]. Research on pedestrian interaction shows that communication, gaze, gesture, and cultural context shape crossing behavior in ways that generic trajectory datasets may not capture [9]. End-to-end autonomous driving research similarly highlights the difficulty of aligning perception, prediction, and planning when training data lacks causal structure or counterfactual coverage [10]. The data-centric turn therefore asks not only whether a model can fit existing data but whether the data itself encodes the operational realities the model will face.

The shift from model-centric to data-centric optimization also changes organizational responsibilities. In a model-centric paradigm, data collection is often a precursor to training, while engineering prestige concentrates on architecture, loss functions, and benchmark rankings. In a data-centric paradigm, dataset design becomes an ongoing function with its own metrics, versioning, audits, and feedback loops. Data engineers, simulation specialists, safety analysts, legal experts, and domain ethnographers must collaborate. This expansion of stakeholders can slow development, but it also creates the possibility of more durable certification. The challenge is to build governance structures that make data-centric work rigorous without making it paralyzed by procedural overhead.

A further tension concerns scale. Large datasets have been essential for progress, but scale alone does not guarantee coverage of rare or consequential events. Waymo, nuScenes, and Argoverse provide millions of frames and rich annotations, yet they remain samples of particular cities, sensor configurations, and traffic cultures [2][3][5]. Intelligent dataset generation is valuable precisely because it can target gaps in the long tail. It can synthesize occluded pedestrians, adversarial lane changes, unusual weather, and conflict-prone intersections. However, synthetic data introduces its own distributional biases. If simulation assets, traffic rules, or agent behaviors reflect narrow assumptions, generated data may reproduce the same blind spots it was meant to correct. Data-centric optimization therefore requires continuous comparison between synthetic and real distributions, not blind augmentation.

3. The Role of Synthetic and Curated Datasets in Trajectory Prediction

Intelligent dataset generation encompasses a spectrum from fully synthetic simulation to lightly curated real-world mining. At the synthetic end, driving simulators such as CARLA and AirSim provide controllable environments for scenario generation, sensor simulation, and safety-critical testing [14][15]. These platforms allow researchers to vary weather, lighting, road geometry, and agent behavior in ways that would be expensive or unethical to reproduce on public roads. At the curated end, real-world logs can be mined for edge cases, rebalanced for density, and annotated with additional contextual labels. A density-balanced data curation and evaluation framework illustrates how trajectory prediction can benefit from explicit attention to the distributional shape of scenarios rather than raw volume alone [12]. The term density balance here refers to the need for training and evaluation sets that do not overrepresent common cruising behavior while underrepresenting interactions that are rare but safety-relevant.

Synthetic and curated datasets are not mutually exclusive. The most promising pipelines combine simulation-generated scenes with real-world validation and human review. Synthetic data can expand coverage of rare events, while real data anchors the model in sensor noise,

behavioral realism, and legal relevance. This hybrid approach raises architectural questions about provenance. Each generated sample should be traceable to its source assets, simulation parameters, sampling policy, and validation status. Without provenance, a model may be trained on synthetic artifacts that are impossible to audit after deployment. Provenance also supports fairness analysis. If a generated dataset overrepresents certain pedestrian behaviors or vehicle types, auditors need to know whether the imbalance comes from simulation design, scenario mining, or annotation practice.

Waymo Open Dataset and similar large-scale resources demonstrate the value of standardized formats and benchmark protocols [13]. They also reveal the limits of static releases. Traffic patterns, vehicle technologies, and urban policies change over time, so a dataset that was representative at release may become stale. Intelligent dataset generation can support continuous refresh, but continuous refresh creates versioning challenges. A model trained on one dataset version may behave differently on another, and regulators may need to certify not only the model but the data lifecycle that produced it. This is especially important for safety cases, where evidence must be linked to specific operational conditions. A data-centric approach therefore treats datasets as living infrastructure with maintenance schedules, deprecation policies, and compatibility guarantees.

The interaction between synthetic generation and fairness deserves particular attention. If simulation engines derive pedestrian behavior from simplified rules, they may exclude cultural, age-related, or disability-related variation. A wheelchair user navigating a curb cut, a child crossing between parked cars, or a street vendor moving through traffic may not appear in standard scenario libraries. Generating such scenarios requires not only technical capability but also social consultation. Data-centric optimization thus becomes a design activity with ethical content. The question is not merely how to generate more trajectories, but whose trajectories are considered relevant, who validates their realism, and how conflicts between realism and safety conservatism are resolved. These questions cannot be answered by engineering metrics alone.

4. Architectural and Infrastructure Implications of Intelligent Dataset Generation

From an architectural perspective, intelligent dataset generation should be integrated into the autonomous driving stack rather than treated as an offline preprocessing step. A typical stack includes perception, prediction, planning, control, and fleet management. Dataset generation can inform each layer. Perception benefits from synthetic sensor variation; prediction benefits from trajectory diversity and interaction coverage; planning benefits from stress-test scenarios; fleet management benefits from operational feedback that identifies where models underperform. This integration requires data contracts between subsystems. Prediction teams need to specify what scenario attributes matter, such as intersection type, right-of-way rules, and pedestrian density. Simulation teams need to expose controllable parameters and validity checks. Safety teams need to define acceptance criteria for generated scenarios. Without these contracts, data generation risks producing volume without relevance.

Infrastructure is equally important. Generating, storing, annotating, and serving large-scale synthetic and curated datasets requires substantial compute, network, and human labor. Cloud-based pipelines can scale, but they also raise concerns about cost concentration, energy consumption, and data sovereignty. An organization that relies on a small number of cloud providers may gain flexibility but lose control over where data resides and how it is processed. Regulations such as data localization requirements, privacy rules, and sector-specific safety standards may constrain cross-border dataset generation. A sustainable infrastructure strategy

therefore balances centralized training efficiency with distributed validation and regional adaptation. Edge cases may need to be collected and validated locally, then aggregated through privacy-preserving mechanisms. This is not only a technical architecture problem but also a governance architecture problem.

The deployment of generated datasets also affects model maintenance. In traditional machine learning, a model is trained, validated, and released. In autonomous driving, the model operates in a changing environment, so maintenance is continuous. Intelligent dataset generation can support this by detecting distribution shifts, generating targeted scenarios, and retraining models. However, continuous retraining introduces risks of feedback loops. If the fleet collects data based on routes it already prefers, the dataset may become more biased toward those routes. If simulation generates scenarios based on current model failures, it may overfit to known weaknesses while missing unknown ones. A robust architecture therefore includes exploratory data collection, randomized scenario generation, and independent auditing. The system should not merely optimize for the failures it can already see.

Cross-domain comparisons offer useful lessons. In recommender systems, data-centric efforts focus on feedback loops, exposure bias, and fairness. In healthcare, data-centric AI emphasizes provenance, consent, and distribution shift across patient populations. In industrial control, it emphasizes simulation-to-reality transfer and safety envelopes. Autonomous driving trajectory prediction combines all these challenges because it is safety-critical, socially embedded, and physically grounded. The infrastructure must therefore support reproducibility, auditability, and rapid iteration. That combination is difficult to achieve with ad hoc scripts and undocumented datasets. It requires platform investment, standards, and organizational commitment. The cost of such investment is significant, but the cost of brittle data infrastructure may be higher when failures occur in public space.

5. Robustness, Fairness, and Governance in Data-Centric Trajectory Prediction

Robustness in trajectory prediction is often framed as performance under noise, occlusion, or adverse weather. A data-centric perspective broadens robustness to include performance under distributional novelty, value conflicts, and adversarial manipulation. Generated datasets can improve robustness by exposing models to rare combinations of agents and environments. They can also degrade robustness if synthetic artifacts create shortcuts. For example, a model may learn to rely on unrealistic rendering cues rather than genuine social interaction. Robustness testing must therefore include synthetic-to-real transfer, adversarial scenario generation, and stress tests that vary non-causal features. The goal is not to make the model invariant to everything, since some features are causal, but to ensure that predictions remain safe when irrelevant features change.

Fairness in trajectory prediction concerns both outcomes and representation. Outcome fairness asks whether the system performs equitably across locations, road-user types, and demographic contexts. Representation fairness asks whether the dataset includes the people and scenarios that the system will encounter. These concerns intersect with historical inequities in transportation. Low-income neighborhoods may have different street designs, policing patterns, and pedestrian behaviors than wealthier areas. If datasets are dominated by wealthier, well-mapped regions, deployed systems may underperform in marginalized communities. Intelligent dataset generation can address this by deliberately sampling underrepresented areas and validating scenarios with community stakeholders. However, such efforts require resources and political will. Without governance, fairness may remain an optional add-on rather than a design constraint.

Governance frameworks for data-centric autonomous driving must address provenance, consent, transparency, and accountability. The ethics of algorithms literature provides a foundation for mapping these concerns, including issues of opacity, bias, and responsibility gaps [19]. Safety models for self-driving cars emphasize the need for formal, verifiable rules and clear operational boundaries [20]. Data generation sits between these two domains. It is not purely a technical artifact, nor is it purely a policy instrument. It is a mechanism by which certain futures are made visible to the system and others are rendered invisible. Governance should therefore require documentation of scenario selection, synthetic generation methods, validation procedures, and known limitations. It should also establish channels for external audit and public contestation. Without such channels, data-centric optimization may become a technocratic exercise that excludes those most affected by autonomous vehicles.

Incident reports and safety investigations illustrate why governance matters. Analyses of autonomous vehicle accidents in California and the 2018 Uber fatal crash show that failures often involve complex interactions among perception, prediction, planning, and human factors [16][17]. A generated dataset cannot solve all these failures, but it can help if it includes the conditions under which they occur. Conversely, a dataset that omits vulnerable road users or mixed-traffic ambiguity may create a false sense of safety. Governance should therefore require post-deployment monitoring that feeds back into dataset generation. This feedback loop should be auditable and subject to independent review. The goal is not to eliminate all risk, which is impossible, but to ensure that risk is distributed and addressed transparently.

6. Sustainability, Deployment, and Lifecycle Considerations

Sustainability in autonomous driving data-centric optimization has at least three dimensions: environmental, economic, and institutional. Environmental sustainability concerns the energy and carbon footprint of generating, storing, and training on large datasets. Synthetic data can reduce the need for physical road testing, which may lower emissions from test fleets, but simulation and training themselves consume substantial compute. The net environmental effect depends on how efficiently synthetic data is used. Generating millions of low-value frames is wasteful. A more sustainable approach uses targeted generation, active learning, and data pruning. The density-balanced curation framework mentioned earlier points in this direction by emphasizing evaluation quality over raw quantity [12]. Economic sustainability concerns the cost of maintaining data pipelines, simulation assets, and annotation teams. Many organizations can afford initial dataset creation but struggle with long-term maintenance. This creates a risk of stale datasets and undocumented drift.

Institutional sustainability concerns the organizational capacity to govern data over time. Autonomous driving programs often span multiple teams, vendors, and jurisdictions. If data governance is fragmented, no single entity may be accountable for dataset quality. A sustainable model requires clear ownership, version control, and lifecycle policies. Datasets should have release notes, deprecation schedules, and migration paths. Models should be linked to the datasets that trained them, and deployments should be traceable to specific data versions. This may seem bureaucratic, but it is comparable to safety engineering in aviation or medical devices. In those domains, traceability is not optional because failures have severe consequences. Autonomous driving should adopt similar discipline, especially as systems move from pilots to public operations.

Deployment considerations also vary by region. A dataset generated for dense European cities may not transfer well to sprawling North American suburbs or to mixed-traffic environments

in Asia, Africa, and Latin America. Intelligent dataset generation can support localization, but localization requires local knowledge. Simulation assets must reflect local road geometry, signage, vehicle types, and behavioral norms. Validation must involve local experts and communities. This suggests a federated approach in which central platforms provide tools and standards while regional partners generate and validate context-specific data. Such an approach can improve robustness and fairness while respecting data sovereignty. It also creates challenges for benchmarking, because models may need to be evaluated across multiple regional test sets rather than a single global leaderboard.

Lifecycle sustainability further requires attention to model retirement. If a model is replaced, its associated datasets may still be valuable for auditing or for training future models. However, retaining all data indefinitely raises privacy and security risks. A lifecycle policy should specify retention periods, anonymization requirements, and secure deletion procedures. It should also consider the rights of individuals whose trajectories appear in the data. Even if data is collected in public spaces, aggregation and inference can reveal sensitive information. Data-centric optimization must therefore be paired with privacy engineering. Techniques such as differential privacy, federated learning, and synthetic data generation can help, but they introduce trade-offs in fidelity and utility. These trade-offs should be evaluated explicitly rather than assumed away.

7. Evaluation, Policy, and Cross-Domain Perspectives

Evaluation of data-centric trajectory prediction must move beyond aggregate displacement error. A model may have low average error while failing systematically in rare scenarios or for specific road users. Evaluation should therefore include density-balanced metrics, worst-case analysis, and subgroup reporting. It should distinguish between in-distribution and out-of-distribution performance. It should test for calibration, because planners need to know when predictions are uncertain. It should also assess scenario validity, especially for synthetic data. A generated scene may be technically plausible but socially impossible, such as a pedestrian crossing against a red light in a jurisdiction where this is extremely rare. Evaluation should include human review and, where possible, field validation. These practices are more expensive than benchmark scoring, but they are necessary for safety cases.

Policy implications follow from these evaluation requirements. Regulators may need to require documentation of dataset provenance, coverage, and known biases. They may need to establish standards for simulation fidelity and synthetic data validation. They may also need to support public datasets that are independent of any single company. Public infrastructure for autonomous driving data could reduce concentration of power and improve transparency. However, public datasets raise questions about funding, maintenance, and privacy. A hybrid model may be appropriate, with government funding for foundational datasets and industry contributions for operational data. The key is to ensure that data governance serves the public interest, not only commercial competitiveness.

Cross-domain perspectives can enrich this policy discussion. In climate science, data curation and uncertainty quantification are central to policy credibility. In public health, data equity and community consent shape research legitimacy. In cybersecurity, threat intelligence sharing requires trust and standardization. Autonomous driving can learn from these domains by treating data as a shared resource with rules of access, quality, and accountability. The alternative is a fragmented landscape in which each company builds proprietary datasets that are neither comparable nor auditable. Such fragmentation may slow safety improvements and

erode public trust. A more collaborative approach could accelerate progress while distributing the costs of data generation and validation.

Finally, the research agenda for intelligent dataset generation should include several priorities. First, methods for measuring and improving synthetic-to-real transfer without requiring exhaustive real-world testing. Second, frameworks for density balancing that are sensitive to safety relevance rather than only frequency. Third, tools for participatory scenario design that include marginalized road users. Fourth, standards for provenance and versioning across organizations. Fifth, evaluation protocols that combine quantitative metrics with qualitative and contextual analysis. Sixth, governance mechanisms that link dataset generation to post-deployment monitoring. These priorities are not purely technical. They require interdisciplinary collaboration among computer scientists, transportation engineers, social scientists, legal scholars, and community representatives. Data-centric optimization is therefore as much a social and institutional project as it is an algorithmic one.

8. Conclusion

This paper has argued that data-centric optimization of autonomous driving trajectory prediction via intelligent dataset generation should be understood as a system-level capability. The field has already benefited from large-scale datasets and simulation platforms, but the next stage of maturity requires deliberate attention to coverage, balance, provenance, fairness, and lifecycle governance. Intelligent dataset generation can target rare events, support continuous improvement, and adapt models to diverse operational contexts. Yet it can also reproduce biases, create synthetic artifacts, and concentrate power if left ungoverned. The difference depends on architecture, infrastructure, and policy. Organizations must integrate dataset generation into the driving stack, document data provenance, validate synthetic scenarios, and establish feedback loops with deployment. Regulators and public institutions must ensure that data practices are transparent, accountable, and oriented toward public safety rather than narrow performance metrics. Researchers must develop evaluation methods that capture density balance, subgroup performance, and socio-technical validity. Ultimately, trajectory prediction will be only as robust and fair as the data that shapes it. Intelligent dataset generation is therefore not a shortcut around hard engineering and governance problems. It is a way of confronting them more systematically. If pursued with appropriate safeguards, it can help autonomous vehicles anticipate human behavior more reliably across the varied and contested spaces of everyday mobility.

References

1. Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? The KITTI vision benchmark suite. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3354-3361). IEEE.
2. Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., & 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). IEEE.
3. Chang, M.-F., Lambert, J., Sangkloy, P., Singh, J., Bak, S., Hartnett, A., Wang, D., Carr, P., Lucey, S., Ramanan, D., & 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). IEEE.

4. Zhan, W., Sun, L., Wang, D., Shi, H., Clausse, A., Naumann, M., Kummerle, J., Konigshof, H., Stiller, C., de La Fortelle, A., & Tomizuka, M. (2019). Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps. arXiv preprint arXiv:1910.03088.
5. Ettinger, S., Cheng, S., Caine, B., Liu, C., Zhao, H., Pradhan, S., Chai, Y., Sapp, B., Qi, C. R., Zhou, Y., Yang, Z., Chouard, A., Sun, P., Ngiam, J., Vasudevan, V., McCauley, A., Shlens, J., & 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). IEEE.
6. Schwarting, W., Alonso-Mora, J., & Rus, D. (2018). Planning and decision-making for autonomous vehicles. Annual Review of Control, Robotics, and Autonomous Systems, 1, 187-210.
7. Lefevre, S., Vasquez, D., & Laugier, C. (2014). A survey on motion prediction and risk assessment for intelligent vehicles. ROBOMECH Journal, 1, 1.
8. Saleh, K., Hossny, M., & Nahavandi, S. (2018). Towards trusted autonomous vehicles from vulnerable road users perspective. In Proceedings of the IEEE Systems Conference (pp. 1-7). IEEE.
9. Rasouli, A., & Tsotsos, J. K. (2020). Autonomous vehicles that interact with pedestrians: A survey of theory and practice. IEEE Transactions on Intelligent Transportation Systems, 21(3), 900-918.
10. Chen, L., Wu, P., Chitta, K., Jaeger, B., Geiger, A., & Li, H. (2024). End-to-end autonomous driving: Challenges and frontiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12), 9866-9883.
11. Zha, D., Bhat, Z. P., Lai, K.-H., Yang, F., Jiang, Z., Zhong, S., & Hu, X. (2023). Data-centric artificial intelligence: A survey. arXiv preprint arXiv:2303.10158.
12. 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).
13. Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., Vasudevan, V., Han, W., Ngiam, J., Zhao, H., Timofeev, A., Ettinger, S., Krivokon, M., Gao, A., Joshi, A., Zhang, Y., Shlens, J., Chen, M., & Anguelov, D. (2020). Scalability in perception for autonomous driving: Waymo open dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2446-2454). IEEE.
14. Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017). CARLA: An open urban driving simulator. In Proceedings of the 1st Annual Conference on Robot Learning (pp. 1-16). PMLR.
15. Shah, S., Dey, D., Lovett, C., & Kapoor, A. (2018). AirSim: High-fidelity visual and physical simulation for autonomous vehicles. In Field and Service Robotics (pp. 621-635). Springer.
16. Favarò, F. M., Nader, N., Eurich, S. O., Tripp, M., & Varadarajan, N. (2017). Examining accident reports involving autonomous vehicles in California. PLOS ONE, 12(9), e0184952.

17. Kohli, P., & Chadha, A. (2020). Enabling pedestrian safety using computer vision techniques: A case study of the 2018 Uber fatal crash. In Proceedings of the IEEE International Conference on Big Data (pp. 5246-5252). IEEE.
18. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. arXiv preprint arXiv:1606.06565.
19. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1-21.
20. Shalev-Shwartz, S., Shammah, S., & Shashua, A. (2017). On a formal model of safe and scalable self-driving cars. arXiv preprint arXiv:1708.06374.