

Multimodal Sensor Fusion Optimization through Post-Training for Autonomous Vehicles

Tijn van Arnsberg

Department of Computer Science, University of Central Florida, Orlando, FL, USA.
vanarnsberg129@ucf.edu

Corrie C. Sammon

Department of Computer Science, University of Houston, Houston, TX, USA.
ccsammon@uh.edu

Elladio Galuppi

Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
elladio.galuppi@ku.edu

Stef Boer

Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
stefboer86@uab.edu

Abstract

Multimodal sensor fusion is central to autonomous vehicle perception because no single sensing modality can provide reliable geometry, semantics, and resilience across weather, lighting, and traffic conditions. Although architectural advances have improved fusion, deployed systems remain constrained by distribution shift, calibration drift, sensor degradation, and long-tail events. This paper examines post-training as a system-level optimization layer for multimodal sensor fusion. Post-training includes supervised fine-tuning, parameter-efficient adaptation, uncertainty calibration, knowledge distillation, and preference or reward-based alignment. The argument is that these techniques should not be treated as model-centric refinements but as infrastructure-aware mechanisms that shape the trade-offs among latency, redundancy, robustness, energy, and regulatory accountability. The paper analyzes fusion architectures, describes how post-training alters representation alignment and failure behavior, and considers deployment constraints in edge and vehicle-cloud environments. It further discusses governance implications, including validation, auditing, fairness, and sustainable mobility. The conclusion is that post-training can improve robustness and adaptation, but its benefits depend on data governance, sensor health monitoring, uncertainty communication, and institutional capacity to certify continuously updated systems.

Keywords

multimodal sensor fusion; autonomous vehicles; post-training; robustness; calibration; edge computing; governance; sustainable mobility.

1. Introduction

Autonomous driving is often framed as a perception problem, but it is more accurately a socio-technical infrastructure problem in which sensing, computation, communication,

validation, and governance are tightly coupled. Vehicle safety models assume bounded perception error, predictable behavior, and timely fallback, yet real roads produce ambiguous scenes, occlusions, adverse weather, and adversarial conditions that challenge every sensor individually [1]. The number of miles required to demonstrate reliability statistically is enormous, which means that post-deployment learning and targeted adaptation become unavoidable if systems are to improve after initial certification [2]. Early benchmarks such as KITTI and later multimodal datasets such as nuScenes established the empirical foundation for comparing fusion methods, but they also revealed that benchmark performance does not directly translate into operational safety [3], [4].

Multimodal fusion is attractive because cameras provide dense semantics and texture, lidar provides precise depth and geometry, radar offers velocity and weather robustness, and positioning or inertial sensors contribute ego-motion context. Fusion, however, is not merely concatenation or late averaging. It requires alignment across coordinate frames, time synchronization, calibration, modality reliability estimation, and representation learning that preserves complementary information while suppressing contradictory signals [5]. Classic fusion architectures range from camera-lidar projection and region proposal fusion to point-based augmentation and multi-task learning [6], [7], [8], [9]. Reviews of deep sensor fusion emphasize that the field has matured technically but remains fragmented across datasets, hardware configurations, and safety cases [10].

Post-training has emerged as a practical way to adapt large perception and driving models after pretraining. Attention-based architectures and large-scale representation learning provide the backbone for many modern systems, but they also create models that are expensive to retrain and difficult to certify [11]. Post-training can refine fusion behavior without full retraining, adjust uncertainty estimates, distill large models into edge-deployable variants, and align outputs with human preferences or safety constraints [12], [13], [14], [15]. Recent work on post-training in end-to-end autonomous driving [16] highlights that adaptation after pretraining is not a minor engineering step but a distinct optimization phase with its own data requirements, evaluation risks, and deployment consequences.

This paper argues that post-training for multimodal sensor fusion should be understood as a systems optimization layer. Its purpose is not simply to improve average accuracy. It must manage redundancy, latency, calibration, energy, and accountability under conditions of partial observability and institutional oversight. The discussion proceeds through architecture, post-training mechanisms, robustness and safety, infrastructure and sustainability, governance and fairness, and comparative case illustrations. The central claim is that post-training can improve autonomous vehicle perception, but only when it is coupled with sensor health monitoring, uncertainty-aware validation, and governance mechanisms that treat model updates as changes to a safety-critical infrastructure.

2. Multimodal Fusion Architectures and System Constraints

Fusion architectures can be broadly divided into early, intermediate, and late fusion, although real systems often combine them. Early fusion merges raw or lightly processed signals, which can preserve fine-grained correlations but demands strict synchronization and calibration. Intermediate fusion learns shared representations from modality-specific encoders, offering flexibility but introducing opacity in how conflicts are resolved. Late fusion combines modality-specific decisions, which simplifies modular verification but may discard cross-modal cues that would have prevented a failure. Camera-lidar fusion methods such as multi-view detection and point painting illustrate the trade-offs between geometric precision and

semantic richness [6], [9]. Radar-camera fusion adds robustness in rain, fog, and darkness, but radar returns are sparse and noisy, requiring careful association and false-positive suppression [10].

The system constraints are as important as the algorithmic choices. Autonomous vehicles operate under strict latency and power budgets. A fusion model that improves accuracy by increasing computation may reduce the time available for planning and fallback, thereby worsening overall safety. Redundancy is often proposed as a solution, but redundant sensors increase cost, weight, calibration complexity, and maintenance burden. A camera may fail from glare, a lidar from reflective surfaces, and radar from interference. The fusion system must therefore estimate not only what is present but also which sensors are trustworthy at a given moment. This is a reliability engineering problem, not only a representation learning problem [1], [5].

Dataset and benchmark design shape what fusion systems learn. KITTI and nuScenes provide standardized evaluation, but they cannot represent every operational design domain, weather pattern, or cultural driving norm [3], [4]. A model trained predominantly on daytime urban scenes may perform poorly in rural night conditions, construction zones, or regions with different lane-marking conventions. Post-training can address these gaps through targeted adaptation, but it can also overfit to small fine-tuning sets and erode previously learned competencies. The risk is especially high when post-training data is collected from a limited fleet or from simulation environments that do not capture sensor degradation and human behavior faithfully [10], [16].

Architectural choice also affects governance. Modular pipelines with explicit interfaces between perception, prediction, and planning are easier to audit component by component, but they may propagate errors and lack end-to-end optimization. End-to-end models can learn richer policies, but their internal representations are difficult to inspect and attribute [16]. Fusion models that rely on attention across modalities can dynamically weight sensors, which is powerful but makes failure explanation harder. For regulators and safety assessors, the key question is not whether a model is end-to-end or modular, but whether the system can produce evidence that its behavior remains within acceptable bounds after updates. That evidence depends on architecture, logging, calibration, and post-training procedures [12], [21].

3. Post-Training as a System-Level Optimization Layer

Post-training refers to any adaptation performed after an initial pretraining phase. In multimodal fusion, it may include supervised fine-tuning on domain-specific data, parameter-efficient adaptation such as low-rank updates, quantized fine-tuning for embedded hardware, uncertainty calibration, knowledge distillation, and alignment through human feedback or reward models [14], [15], [17], [18]. These methods are often discussed as model compression or transfer learning techniques, but in autonomous vehicles they function as infrastructure operations. They determine how the vehicle responds to new sensor configurations, seasonal changes, regulatory updates, and newly discovered edge cases. Post-training is therefore a continuous operational capability, not a one-time model release [16].

Supervised fine-tuning can improve fusion on rare scenarios, such as unprotected left turns, emergency vehicle encounters, or nighttime pedestrian detection. However, fine-tuning on rare events can destabilize common-case behavior if the data distribution is skewed. Parameter-efficient methods mitigate some cost by updating a small subset of weights, which makes fleet-wide deployment more feasible [15]. Quantized fine-tuning further reduces

memory and energy demands, enabling updates on vehicle-grade hardware [17]. Yet these efficiencies introduce new risks: low-rank or quantized updates may alter calibration, compress away rare but safety-critical features, or interact unpredictably with sensor preprocessing. Validation must therefore cover not only accuracy but also uncertainty, latency, and failure modes [13].

Knowledge distillation is particularly relevant for multimodal fusion because large teacher models can integrate many sensors and modalities, while smaller student models must run in real time. Distillation can transfer cross-modal associations and soften hard labels, improving robustness when one modality is degraded [14]. The danger is that a student may imitate a teacher's overconfidence without inheriting its redundancy. If the teacher learned to rely on lidar in clear conditions, the student may fail when lidar is occluded or dirty. Post-training should therefore include modality dropout, sensor fault injection, and adversarial calibration tests. The goal is not merely to compress a model but to preserve a safety case under component degradation [1], [12].

Preference-based and reward-based post-training, developed largely in language and decision-making systems, offers a framework for aligning driving behavior with human values and policy constraints [18]. In autonomous driving, feedback may come from safety drivers, simulation evaluators, or regulatory criteria. The challenge is that preferences are context-dependent and culturally variable. A maneuver considered polite in one region may be unsafe in another. Reward models can also be gamed, leading to aggressive or overly timid behavior. Post-training must therefore be constrained by rule-based safety envelopes, formal monitors, and interpretable fallback logic [19]. The most credible systems will combine learned adaptation with verifiable runtime checks that can override unsafe outputs.

4. Robustness, Calibration, and Safety Assurance

Robustness in multimodal fusion is not the absence of errors but the ability to degrade gracefully. A robust system recognizes when a modality is unreliable, shifts weight to remaining sensors, and communicates uncertainty to planning and human-machine interfaces. Calibration is central because a model that is accurate on average but overconfident in rare cases can cause dangerous planning decisions [13]. Post-training can recalibrate outputs using held-out data, temperature scaling, or uncertainty-aware losses, but calibration must be maintained across firmware updates, sensor replacements, and environmental changes. A calibrated model in the laboratory may become miscalibrated after a minor hardware change [12], [16].

Safety assurance traditionally relies on requirements, hazard analysis, traceability, and independent validation. These practices assume a relatively stable system boundary. Post-training challenges that assumption because the model may change frequently while the physical vehicle remains in service. A safety case must therefore include change management: what data triggered the update, what tests were performed, what performance regressions were observed, and what rollback procedures exist. Fleet-level learning can improve safety if updates are validated across diverse conditions, but it can also propagate a localized failure to thousands of vehicles if the update is not representative [2], [16]. Governance must treat model updates as engineering changes with documented risk assessments.

Uncertainty quantification is one bridge between post-training and safety assurance. Bayesian approximations, ensembles, and evidential methods can estimate epistemic and aleatoric uncertainty, helping the system distinguish unfamiliar scenes from sensor noise [12].

However, uncertainty estimates are only useful if downstream planning and human oversight can act on them. A vehicle that reports high uncertainty but continues at speed provides little safety benefit. Post-training should therefore optimize the entire perception-to-action chain, not only perception metrics. It should reward conservative behavior under uncertainty and ensure that fallback maneuvers are available. This systems view is consistent with formal safety models that define safe fallback states and responsibility-sensitive planning [1].

Fairness and robustness are also linked. A fusion model may perform well for typical vehicles and pedestrians but fail for wheelchair users, children, cyclists, or individuals wearing clothing that reduces visual contrast. These are not merely fairness concerns; they are safety concerns. Post-training data must include diverse body types, mobility aids, skin tones, clothing, and lighting conditions. Otherwise, the model's failure modes will be systematically distributed across social groups [20], [21]. Because post-training often uses smaller, targeted datasets, there is a heightened risk of representational bias. Validation should therefore report disaggregated performance across relevant conditions and populations, not only aggregate detection scores.

5. Deployment Infrastructure, Energy, and Sustainability

Deployment infrastructure shapes what post-training can achieve. Autonomous vehicles are edge devices with limited power, thermal, and memory budgets, but they are also connected nodes in a fleet-level data pipeline. Post-training can occur on the vehicle, in a depot, or in the cloud. On-vehicle adaptation enables immediate personalization but risks uncontrolled divergence and difficult certification. Depot-based updates allow more computation and testing but require downtime. Cloud-based training can leverage large datasets but introduces communication costs, privacy risks, and latency. A hybrid architecture is likely, with cloud training, depot validation, and on-vehicle inference plus limited adaptation [10], [17].

Energy and sustainability are often overlooked in fusion research. More sensors, higher-resolution data, and larger models increase energy consumption both during training and inference. Post-training can reduce inference energy through distillation and quantization, but frequent cloud retraining can offset those savings. Lifecycle emissions include sensor manufacturing, vehicle integration, data transmission, and end-of-life disposal. If autonomous vehicles encourage more driving or larger vehicles, the sustainability benefits of efficient perception may be overwhelmed by induced demand [22], [23]. Post-training should therefore be evaluated not only in accuracy per watt but in system-level mobility outcomes, including vehicle occupancy, mode shift, and land use.

Infrastructure also includes communication networks, maps, and roadside units. Cooperative perception can extend sensing beyond line of sight, but it depends on reliable low-latency connectivity and trust in external data. Post-training for cooperative fusion must address message loss, malicious inputs, and heterogeneous sensor configurations. A vehicle that learns to rely on infrastructure may fail when connectivity drops. Conversely, a vehicle that ignores infrastructure forfeits potential safety gains. The design must balance autonomy and cooperation, with graceful degradation when external inputs are unavailable [5], [10]. This is a structural trade-off between capability and resilience.

Maintenance and calibration are ongoing infrastructure costs. Sensors degrade, lenses become dirty, and mounting positions shift after minor collisions. Post-training can compensate for some drift, but it cannot replace physical calibration and sensor health monitoring. A sustainable deployment model must include remote diagnostics, predictive maintenance, and

transparent reporting of sensor status. If post-training is used to mask degraded hardware, it may create a false sense of safety. The better approach is to treat sensor health as a first-class input to fusion, allowing the model to down-weight faulty modalities and request service when needed [12], [16].

6. Governance, Fairness, and Policy Implications

Governance for post-trained autonomous vehicle fusion must address accountability across a distributed supply chain. Original equipment manufacturers, sensor vendors, software suppliers, fleet operators, and cloud providers all contribute to the final behavior. When an update changes perception, responsibility is difficult to assign. Regulatory frameworks that focus only on pre-deployment approval cannot keep pace with continuous learning. A more adaptive approach includes registration of model versions, audit trails for training data, independent testing of updates, and mandatory incident reporting [20], [21]. Transparency should be proportionate to risk: the public may not need source code, but regulators need evidence of calibration, failure modes, and change control.

Privacy and data protection are central because post-training often relies on fleet-collected sensor data. Cameras capture faces, license plates, and private property. Lidar can reveal detailed spatial information. Data minimization, on-vehicle processing, anonymization, and retention limits are necessary, but they can reduce the diversity of post-training data. There is a tension between privacy and robustness. Synthetic data and simulation can help, but simulated sensor artifacts may not match reality. Federated learning offers a partial solution by training across vehicles without centralizing raw data, yet it introduces challenges in secure aggregation, model poisoning, and auditability [21], [24]. Policy must enable safety improvement without creating unchecked surveillance infrastructure.

Fairness requires attention to both allocation and outcomes. If post-training improves performance only in regions where data is abundant, autonomous vehicles may be safest in wealthy, well-mapped areas and less safe in marginalized communities. This can exacerbate transport inequities. Regulators and operators should require geographic and demographic performance reporting, inclusive data collection, and community engagement. Fairness is not only about equal detection accuracy; it is about equitable access to mobility benefits and equitable distribution of risk [20], [22]. Post-training can support fairness if it is designed with explicit equity objectives, but it can also entrench bias if optimization targets only aggregate safety or efficiency.

Policy implications extend to liability and insurance. If a post-trained update causes a crash, the liability may fall on the manufacturer, the fleet operator, or the software provider depending on contracts and jurisdiction. Insurance markets need actuarial data on model versions and update histories. Governments may need to define minimum standards for post-training documentation, independent validation, and recall procedures. International harmonization would reduce fragmentation, but it must accommodate local road rules and cultural expectations. The governance challenge is to support innovation while ensuring that continuous adaptation does not erode public accountability [24], [2].

7. Case Illustrations and Cross-Domain Comparisons

A useful illustration is winter driving. Cameras may be obstructed by snow, lidar by spray, and radar by clutter. A post-trained fusion model can learn to trust radar and thermal sensors more in these conditions, but only if it has been exposed to representative winter data and can detect sensor occlusion. If post-training is performed in simulation, the model may learn

unrealistic spray patterns. A robust deployment would combine post-training with heater diagnostics, lens cleaning, and uncertainty-triggered speed reduction. The case shows that post-training is necessary but insufficient without hardware and infrastructure support [10], [12].

Another illustration is urban construction zones. Lane markings disappear, workers and cones appear unexpectedly, and maps become stale. A fusion model must integrate camera semantics, lidar geometry, and possibly vehicle-to-infrastructure messages. Post-training on construction data can improve detection, but the operational design domain may change daily. A governance system might require geofenced speed limits and remote assistance when confidence is low. This mirrors aviation, where continuous airworthiness and crew procedures manage uncertainty rather than assuming perfect automation [19], [24]. Cross-domain comparison suggests that autonomous vehicles need a safety management system, not just a better model.

A third illustration is fleet-level update propagation. Suppose a post-training update improves night-time pedestrian detection by ten percent but slightly increases false positives for roadside signs. In a small test, the trade-off may seem acceptable. Across a fleet, false positives could cause unnecessary braking and rear-end collisions. The update should be rolled out gradually, with shadow mode evaluation, geographic canaries, and rollback triggers. This is analogous to software canarying in cloud infrastructure, but the safety consequences are physical. Post-training optimization must therefore include release engineering and operational monitoring as core components [16], [2].

Cross-domain comparisons with language models are instructive but limited. Preference-based post-training has improved helpfulness and safety in language systems, yet driving has hard physical constraints and irreversible consequences [18]. A language model can refuse a request; a vehicle cannot refuse to brake. Knowledge distillation and low-rank adaptation are transferable techniques, but their evaluation criteria differ. In language, perplexity and human preference may suffice. In driving, calibration, latency, failure severity, and regulatory compliance dominate. The transfer should be methodological, not metaphorical [14], [15], [17].

8. Conclusion

Post-training for multimodal sensor fusion in autonomous vehicles is best understood as a system-level optimization layer that connects model adaptation to infrastructure, safety, sustainability, and governance. It can improve robustness to weather, sensor degradation, and long-tail events, and it can make large fusion models deployable on constrained hardware. However, its benefits are conditional. Post-training can overfit, miscalibrate, propagate failures across fleets, and obscure accountability if it is treated as a purely technical update. The most credible systems will combine parameter-efficient adaptation, uncertainty calibration, distillation, and constrained alignment with runtime safety monitors and transparent change management.

The paper has argued that architectural choices and post-training methods must be evaluated together. Early, intermediate, and late fusion create different failure modes and governance requirements. Calibration and uncertainty quantification are essential for planning and human oversight. Deployment infrastructure determines whether updates improve safety or merely shift risk. Energy, maintenance, privacy, and equity are not secondary concerns; they are structural constraints on sustainable adoption. Policy frameworks should require audit trails,

disaggregated performance reporting, independent validation, and rollback procedures for post-trained models.

Future research should develop evaluation suites that measure robustness, calibration, fairness, latency, and energy under realistic sensor degradation. It should also investigate federated and privacy-preserving post-training, simulation-to-reality transfer, and methods for explaining attention across modalities. Most importantly, post-training should be integrated into a safety management system that treats autonomous vehicles as continuously evolving socio-technical infrastructures. The goal is not to eliminate uncertainty but to manage it responsibly through technical rigor, institutional accountability, and public legitimacy.

References

1. Shalev-Shwartz, S., Shammah, S., & Shashua, A. (2017). On a formal model of safe and scalable self-driving cars. arXiv preprint arXiv:1708.06374.
2. Kalra, N., & Paddock, S. M. (2016). Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transportation Research Part A: Policy and Practice*, 94, 182-193. <https://doi.org/10.1016/j.tra.2016.09.010>
3. Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? The KITTI vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3354-3361). IEEE. <https://doi.org/10.1109/CVPR.2012.6248074>
4. 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 *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11621-11631). IEEE. <https://doi.org/10.1109/CVPR42600.2020.01164>
5. Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423-443. <https://doi.org/10.1109/TPAMI.2018.2798607>
6. Chen, X., Ma, H., Wan, J., Li, B., & Xia, T. (2017). Multi-view 3D object detection network for autonomous driving. In *2017 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1907-1915). IEEE. <https://doi.org/10.1109/CVPR.2017.652>
7. Ku, J., Mozifian, M., Lee, J., Harakeh, A., & Waslander, S. L. (2018). Joint 3D proposal generation and object detection from view aggregation. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 1-8). IEEE. <https://doi.org/10.1109/IROS.2018.8594049>
8. Liang, M., Yang, B., Wang, S., & Urtasun, R. (2019). Multi-task multi-sensor fusion for 3D object detection. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7345-7353). IEEE. <https://doi.org/10.1109/CVPR.2019.00752>
9. Vora, S., Lang, A. H., Helou, B., & Beijbom, O. (2020). PointPainting: Sequential fusion for 3D object detection. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4604-4612). IEEE. <https://doi.org/10.1109/CVPR42600.2020.00466>
10. Fayyad, J., Jaradat, M. A., Gruyer, D., & Najjaran, H. (2020). Deep learning sensor fusion for autonomous vehicle perception and localization: A review. *Sensors*, 20(15), 4220. <https://doi.org/10.3390/s20154220>

11. 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* 30 (pp. 5998-6008). Curran Associates.
12. Kendall, A., Gal, Y., & Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7482-7491). IEEE.
<https://doi.org/10.1109/CVPR.2018.00781>
13. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning* (Vol. 70, pp. 1321-1330). PMLR.
14. Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.
15. Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2022). LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.
16. Yang, R., Wang, M., Chen, Y., Guo, T., Xu, Y., Cui, C., ... & Su, L. (2026). Post-Training in End-to-End Autonomous Driving. *arXiv preprint arXiv:2607.08072*.
17. Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023). QLoRA: Efficient finetuning of quantized LLMs. In *Advances in Neural Information Processing Systems* 36.
18. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems* 35 (pp. 27730-27744).
19. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*.
20. Bryson, J. J., Diamantis, M. E., & Grant, T. D. (2017). Of, for, and by the people: The legal lacuna of synthetic persons. *Artificial Intelligence and Law*, 25(3), 273-291.
<https://doi.org/10.1007/s10506-017-9214-9>
21. Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law*, 7(2), 76-99. <https://doi.org/10.1093/idpl/ix005>
22. Banister, D. (2008). The sustainable mobility paradigm. *Transport Policy*, 15(2), 73-80.
<https://doi.org/10.1016/j.tranpol.2007.10.005>
23. Litman, T. (2020). Autonomous vehicle implementation predictions: Implications for transport planning. Victoria Transport Policy Institute.
24. Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A*, 376(2133), 20180080. <https://doi.org/10.1098/rsta.2018.0080>