

Self-Supervised Post-Training for Robust End-to-End Vehicle Control

Wiktor Baldys

Department of Computer Science, University of New Hampshire, Durham, NH, USA.
wiktor1970@unh.edu

Abstract

End-to-end vehicle control has emerged as a compelling alternative to modular autonomous driving stacks because it can learn perception, prediction, planning, and actuation within a unified differentiable architecture. Yet the same integration that enables graceful behavior in nominal conditions also creates fragility under distribution shift, sensor degradation, rare edge cases, and evolving road environments. This paper examines self-supervised post-training as a system-level strategy for improving the robustness of end-to-end vehicle control after initial supervised or reinforcement-learning-based training. Rather than treating self-supervision as a mere pretext for representation learning, the discussion positions post-training as a socio-technical process that couples model adaptation, fleet data infrastructure, verification, governance, and lifecycle sustainability. The analysis considers architectural trade-offs between monolithic policies and modular safety envelopes, the role of large-scale unlabeled driving data, and the operational challenges of continuous over-the-air updates. It also explores fairness, privacy, and policy implications when self-supervised adaptation depends on fleet telemetry and simulation pipelines. The central argument is that robust end-to-end control cannot be achieved through algorithmic innovation alone; it requires disciplined post-training regimes, auditable data provenance, safety-case integration, and deployment governance that treats model updates as safety-critical changes. By connecting representation learning with infrastructure and regulation, the paper outlines an interdisciplinary agenda for self-supervised post-training that is technically effective, operationally accountable, and socially sustainable.

Keywords

end-to-end vehicle control, self-supervised learning, post-training, autonomous driving, robustness, safety assurance, deployment governance, sustainability.

1. Introduction

End-to-end vehicle control represents a significant shift in the design of autonomous driving systems. Instead of separating perception, localization, prediction, planning, and control into independently engineered modules, end-to-end approaches map raw sensor inputs directly to driving actions or trajectories. The intellectual foundations of this shift lie in deep representation learning, which demonstrated that hierarchical neural networks can discover useful features from high-dimensional data without handcrafted pipelines [1], [2]. More recently, transformer architectures have provided scalable mechanisms for attending to long-range dependencies across temporal and multimodal streams, making them attractive for driving policies that must integrate camera, lidar, radar, and vehicle state information [3]. At the same time, the promise of end-to-end control is tempered by a central systems problem: a policy trained on a finite dataset may fail unpredictably when the operational environment diverges from its training distribution.

Self-supervised learning has become a major response to this problem. Methods such as masked autoencoding and contrastive representation learning allow models to learn from unlabeled data by solving tasks that require understanding of structure, invariance, and temporal consistency [4], [5], [6]. In the driving domain, unlabeled data is abundant because fleet vehicles continuously collect video, lidar, radar, and kinematic signals. The challenge is not merely to pretrain a backbone, but to post-train an entire control policy in a way that improves robustness without undermining safety, interpretability, or regulatory compliance. Recent work on post-training in end-to-end autonomous driving [7] signals a growing recognition that the adaptation phase after initial training is where many practical robustness gains and risks are concentrated.

This paper argues that self-supervised post-training for end-to-end vehicle control should be understood as a socio-technical infrastructure problem rather than a narrow machine learning technique. The discussion is organized around several interconnected themes. First, it clarifies the conceptual foundations of self-supervised post-training and distinguishes it from pretraining, fine-tuning, and conventional data augmentation. Second, it examines how post-training interfaces interact with end-to-end architectures, including the tension between monolithic policies and modular safety monitors. Third, it analyzes robustness under distribution shift and safety-critical edge cases. Fourth, it considers data, compute, deployment, and update infrastructure. Fifth, it explores governance, fairness, and policy implications. Sixth, it discusses sustainability and lifecycle management. Finally, it outlines evaluation and assurance practices needed to make self-supervised post-training credible in real-world vehicle fleets.

2. Conceptual Foundations of Self-Supervised Post-Training

Self-supervised learning refers to training regimes in which supervisory signals are derived from the structure of the data itself rather than from human annotations. In vision, masked modeling reconstructs hidden portions of an input, encouraging the model to learn semantic and spatial regularities [4]. Contrastive learning pulls augmented views of the same instance together while pushing apart representations of different instances, thereby shaping invariances that can transfer to downstream tasks [5]. Predictive coding extends this logic to sequential data by learning representations that anticipate future latent states [6]. These approaches are attractive for driving because they can exploit enormous quantities of unlabeled sensor data, which would be prohibitively expensive to annotate at the scale required for every possible edge case.

Post-training, however, is not identical to pretraining. Pretraining typically produces a general-purpose representation that is later adapted to a task. Post-training occurs after an initial policy has been trained and is concerned with refining behavior, calibration, and robustness under operational constraints. In end-to-end vehicle control, post-training may involve adapting a perception backbone, adjusting a trajectory head, recalibrating uncertainty estimates, or aligning a policy with safety envelopes. Self-supervision can enter this phase through auxiliary objectives such as temporal consistency, future-state prediction, cross-modal alignment, or novelty detection. The goal is not simply to improve benchmark accuracy but to make the policy more reliable when the vehicle encounters weather changes, construction zones, unusual traffic participants, or sensor faults.

A system-level view reveals several structural trade-offs. Self-supervised post-training can improve data efficiency by reducing reliance on labeled incidents, but it often increases compute requirements because large models must be updated repeatedly across fleet data. It

can produce representations that generalize across geographies, but only if the unlabeled data is representative; otherwise, it may amplify biases toward well-instrumented regions or common driving cultures. It can support continuous improvement through over-the-air updates, but each update introduces new verification burdens. These trade-offs are not incidental. They shape the architecture of data pipelines, the economics of fleet operations, and the legal exposure of manufacturers.

The conceptual foundation also has governance implications. Self-supervised objectives are often opaque: a model may learn invariances that are useful for prediction but undesirable for safety, such as ignoring small objects because they rarely affect reconstruction loss. Data provenance becomes critical because unlabeled fleet data may contain personally identifiable information, proprietary location traces, or sensitive behavioral patterns [16]. Ethical frameworks for algorithmic systems emphasize transparency, accountability, and contestability [17], all of which are complicated when post-training is continuous and distributed. Thus, self-supervised post-training must be designed with documentation, auditability, and human oversight from the outset, not as an afterthought.

3. End-to-End Vehicle Control Architectures and Post-Training Interfaces

End-to-end driving architectures have evolved from relatively simple convolutional networks that map camera images to steering commands [9] to conditional imitation learning systems that incorporate navigational commands and route intentions [8]. Some systems learn directly in simulation and transfer to real vehicles [10], while others rely on large-scale multimodal datasets and simulation environments to support training and evaluation [11], [12], [13]. The architectural choice determines where self-supervised post-training can be inserted. A monolithic policy may allow end-to-end gradient flow and holistic adaptation, but it offers few natural intervention points for safety monitoring. A modular or hybrid architecture may sacrifice some end-to-end elegance but provides clearer interfaces for verification and fallback.

Post-training interfaces can be understood as the boundaries through which adaptation occurs. If the perception backbone is frozen and only a control head is updated, post-training is cheaper and easier to validate, but the system may lack the capacity to correct deep representational failures. If the entire policy is updated, the system can adapt more powerfully, but catastrophic forgetting and unpredictable side effects become serious risks [14]. Adapter layers, low-rank updates, and parameter-efficient fine-tuning offer intermediate options. They allow targeted adaptation while preserving much of the original policy behavior, which is attractive for regulated deployment where change management must be explicit.

The tension between end-to-end learning and safety engineering is particularly acute. Formal models of safe self-driving cars emphasize verifiable rules, responsibility-sensitive safety, and conservative braking or avoidance behavior [18]. End-to-end policies are difficult to verify exhaustively because their behavior emerges from high-dimensional learned representations. Self-supervised post-training can help by improving novelty detection and uncertainty estimation, but it cannot by itself provide formal guarantees. A practical architecture therefore often combines a learned policy with a runtime safety monitor that can override unsafe actions or request minimal-risk maneuvers [19]. Post-training must respect this boundary; otherwise, improvements in nominal performance may degrade the effectiveness of the safety envelope.

Deployment considerations further shape architecture. Vehicles have limited compute, power, and thermal budgets, so post-training must eventually produce models that run efficiently at the edge. Continuous updates over the air require versioning, rollback capability, and staged deployment. A self-supervised post-training pipeline that operates on fleet data may discover new patterns, but if those patterns are integrated without rigorous validation, the fleet becomes a distributed experiment. The governance challenge is to balance learning speed with change control, ensuring that every adapted policy has a documented rationale, evaluation record, and fallback configuration.

4. Robustness Under Distribution Shift and Safety-Critical Edge Cases

Robustness is the central motivation for self-supervised post-training. Driving environments are non-stationary: weather, lighting, road maintenance, traffic laws, vehicle mixes, and pedestrian behavior vary across time and place. A policy trained on data from one city may encounter unfamiliar intersections, unusual vehicles, or temporary signage in another. Self-supervised objectives can improve robustness by encouraging representations that capture stable structure across views and time. Contrastive learning, for example, can teach a model that the same scene under different illumination should have similar latent features [5]. Masked modeling can force the model to infer occluded or missing information, which may support better handling of partial sensor failure [4]. Predictive objectives can help the model anticipate future motion and detect anomalies [6].

Yet robustness gains are not automatic. Self-supervised post-training may learn shortcuts that correlate with the pretext task but do not support safe control. A model trained to predict future frames may focus on dominant background motion rather than vulnerable road users. A contrastive objective may treat two different pedestrians as negative examples and learn overly specific features that do not generalize. Therefore, post-training must be coupled with targeted evaluation on safety-critical scenarios, including pedestrians, cyclists, emergency vehicles, construction workers, and degraded sensor conditions. Simulation platforms and multimodal datasets provide controlled environments for such testing [11], [12], [13], but simulation-to-reality gaps remain a fundamental limitation.

Safety-critical edge cases also raise questions of fairness and distributional justice. If self-supervised post-training relies on fleet data, regions with fewer vehicles or less digital infrastructure may be underrepresented. The resulting policy may perform worse in rural areas, low-income neighborhoods, or countries with different road norms. Ethical analysis of algorithms has long warned that seemingly neutral optimization can reproduce structural inequities [17]. In autonomous driving, fairness concerns include equitable access to mobility, consistent protection for vulnerable road users, and transparent accountability when harm occurs. Post-training data governance must therefore include representativeness auditing and impact assessment, not only aggregate performance metrics.

The operational design domain concept is useful here. A vehicle control system is only safe within specified conditions, and post-training may expand or contract that domain [20]. Self-supervised adaptation could allow a fleet to operate in new weather conditions or cities, but it could also create hidden dependencies that are not captured in the documented domain. Regulators and operators need mechanisms to trace how post-training changed the policy, which data contributed to the change, and whether the safety case remains valid. Without such traceability, robustness claims become difficult to distinguish from overfitting to recent fleet experience.

5. Infrastructure and Deployment Considerations

Self-supervised post-training depends on infrastructure that spans data collection, storage, compute, simulation, validation, and deployment. Fleet vehicles generate continuous streams of video, lidar, radar, kinematics, and system logs. This data must be ingested, filtered, and stored with attention to privacy and security. Contextual integrity suggests that information flows should be evaluated against the norms of the context in which data was collected [16]. Location traces and cabin recordings are especially sensitive. A post-training pipeline that indiscriminately uploads raw data may create legal and ethical risks that outweigh its technical benefits.

Compute infrastructure is equally important. Self-supervised learning often benefits from large batch sizes, long training schedules, and repeated adaptation cycles. The energy and carbon footprint of such training can be substantial [23], [24]. If post-training is performed centrally, data movement and storage dominate costs. If it is performed at the edge, model updates must be constrained by vehicle hardware and thermal limits. Hybrid approaches may use centralized representation learning and edge-side personalization, but they introduce synchronization and versioning complexity. The system must decide which data remains local, which is aggregated, and which is used for safety-critical updates.

Deployment infrastructure includes over-the-air update mechanisms, shadow mode evaluation, canary fleets, and rollback procedures. A self-supervised post-training update should not be released directly to all vehicles. It should first be evaluated in simulation, then in closed courses, then in limited geographic areas, and finally in broader deployment. Each stage requires logging, monitoring, and incident reporting. Regulatory frameworks increasingly expect such post-market surveillance for automated driving systems [21]. The infrastructure for post-training is therefore not merely an MLOps concern; it is part of the safety case and the legal record.

Cross-domain comparisons are instructive. Aviation relies on rigorous change control, redundant systems, and incident reporting. Cloud computing relies on continuous deployment but with canary releases and observability. Medical devices rely on premarket approval and post-market surveillance. Autonomous driving must combine elements of all three: rapid learning from fleet data, strict safety assurance, and public accountability. Self-supervised post-training intensifies this combination because it enables continuous adaptation from unlabeled data, which is both a resource and a governance challenge.

6. Governance, Fairness, and Policy Implications

Governance for self-supervised post-training must address accountability, transparency, and fairness. When a vehicle control policy is updated using unlabeled fleet data, who is responsible for the resulting behavior? The manufacturer, the data provider, the model developer, or the operator? Existing liability frameworks may struggle to assign responsibility when harm emerges from an emergent property of continuous adaptation. Policy proposals such as the European Union's Artificial Intelligence Act attempt to classify high-risk systems and impose requirements for risk management, data governance, technical documentation, and human oversight [22]. Autonomous driving clearly falls within high-risk categories, and post-training processes must be included in compliance documentation.

Transparency is difficult because end-to-end models are often opaque, and self-supervised objectives add another layer of indirection. Regulators may not need full access to model weights, but they need evidence that post-training did not degrade safety, introduce

discriminatory behavior, or violate operational constraints. This suggests a role for third-party auditing, standardized evaluation suites, and disclosure of data sources and update histories. Fairness requires attention to who benefits and who bears risk. If self-supervised post-training improves performance in data-rich regions while leaving data-poor regions behind, it may exacerbate mobility inequities. Policy should therefore encourage inclusive data collection and ensure that safety performance is evaluated across diverse populations and geographies.

Privacy and security are also central. Fleet data can reveal home locations, routines, and passenger conversations. Self-supervised learning may not require explicit labels, but it still consumes raw data. Differential privacy, federated learning, and on-device processing are potential mitigations, but each introduces trade-offs in model quality, compute cost, and system complexity. Security is equally important: an adversary who can poison unlabeled data or manipulate the post-training objective could degrade fleet behavior in subtle ways. Governance must therefore include adversarial robustness, anomaly detection, and secure update pipelines.

The policy landscape is still evolving. Standards such as SAE J3016 provide common terminology for driving automation [20], while incident reporting requirements create feedback loops between operators and regulators [21]. Self-supervised post-training adds a new dimension because it changes the temporal dynamics of certification. A vehicle that was approved at deployment may behave differently after several post-training cycles. Regulatory frameworks must therefore shift from one-time approval toward continuous assurance, with clear triggers for re-evaluation, audit, and recall.

7. Sustainability and Lifecycle Management

Sustainability is often overlooked in discussions of autonomous driving, but it is a first-order concern for self-supervised post-training. Training large models consumes significant energy, and repeated post-training cycles multiply that footprint [23], [24]. Inference on fleet vehicles also consumes energy, though the incremental cost per vehicle may be small compared with propulsion. The lifecycle of data collection, storage, training, deployment, and decommissioning must be assessed holistically. Self-supervision can reduce the need for human labeling, which saves labor and some energy, but it may increase compute intensity. The net effect depends on model size, update frequency, data filtering, and hardware efficiency.

Sustainability also intersects with system design. A post-training pipeline that uploads all raw data to centralized data centers may be less efficient than one that filters, compresses, or processes data at the edge. Federated learning and distillation can reduce data movement, but they add coordination overhead. Fleet-level optimization can reduce energy use through smoother driving, platooning, and route selection, but these benefits must be weighed against the energy cost of the autonomy stack itself. Lifecycle governance should require carbon accounting for training and deployment, not just vehicle emissions.

Finally, sustainability includes social and institutional durability. A self-supervised post-training system that depends on continuous data donation may erode public trust if users feel surveilled or if benefits are not shared. Long-term sustainability requires transparent value exchange, robust privacy protections, and inclusive access. The governance of foundation models has raised similar concerns about concentration of power, environmental impact, and equitable access [25]. Autonomous driving should learn from these debates rather than treating sustainability as a public relations add-on.

8. Evaluation, Verification, and Operational Assurance

Evaluation of self-supervised post-training must go beyond standard driving benchmarks. Metrics should cover safety, comfort, progress, rule compliance, uncertainty calibration, fairness, and robustness to distribution shift. A model that improves average displacement error may still perform worse in rare but safety-critical scenarios. Evaluation should therefore include scenario-based testing, adversarial simulation, and field monitoring. Simulation platforms provide repeatability and coverage [11], while real-world datasets provide ecological validity [12], [13]. Neither alone is sufficient.

Verification of end-to-end policies remains difficult. Formal methods can verify safety monitors and rule-based envelopes, but not the full learned policy [18], [19]. Self-supervised post-training can support verification by producing uncertainty estimates or novelty scores that trigger conservative behavior. However, these signals must themselves be validated. A model may be confidently wrong, especially under distribution shift. Operational assurance therefore requires a combination of runtime monitoring, fallback controllers, and incident analysis. Every post-training update should be accompanied by a safety case that documents what changed, what was tested, and what residual risks remain.

Continuous assurance is a cultural and institutional challenge as much as a technical one. Organizations must maintain independence between model development and safety validation, invest in red teaming, and treat fleet data as evidence rather than marketing material. Regulators need access to standardized records. The public needs understandable explanations of how vehicles learn and what protections exist. Self-supervised post-training, if governed well, can make autonomous driving more adaptable and resilient. If governed poorly, it can create opaque, rapidly changing systems whose failures are difficult to diagnose or prevent.

Conclusion

Self-supervised post-training offers a powerful pathway for improving the robustness of end-to-end vehicle control by exploiting vast quantities of unlabeled fleet data. It can help policies adapt to new environments, detect novelty, and maintain temporal consistency beyond what supervised datasets alone can provide. However, the technical promise is inseparable from structural trade-offs in architecture, infrastructure, governance, fairness, and sustainability. A monolithic policy may adapt holistically but resist verification; a modular architecture may be safer but less flexible. Centralized post-training may be easier to manage but raises privacy and energy concerns; edge adaptation may protect data but complicate version control. These trade-offs cannot be resolved by algorithms alone.

The central argument of this paper is that self-supervised post-training should be treated as a safety-critical socio-technical process. It requires auditable data provenance, representative and privacy-preserving data pipelines, staged deployment, continuous evaluation, and regulatory frameworks capable of overseeing models that change over time. Robustness is not a static property of a trained network; it is an emergent property of the entire system, including the people, institutions, and infrastructures that collect data, validate updates, and respond to failures. By integrating representation learning with governance and lifecycle management, the field can move toward end-to-end vehicle control that is not only more capable but also more accountable, equitable, and sustainable.

References

1. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
<https://doi.org/10.1038/nature14539>
2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998-6008.
4. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked autoencoders are scalable vision learners. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16000-16009.
<https://doi.org/10.1109/CVPR52688.2022.01553>
5. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. *Proceedings of the 37th International Conference on Machine Learning*, 1597-1607.
6. Oord, A. v. d., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.
7. 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*.
8. Codevilla, F., Müller, M., López, A., Koltun, V., & Dosovitskiy, A. (2018). End-to-end driving via conditional imitation learning. *2018 IEEE International Conference on Robotics and Automation*, 4693-4700. <https://doi.org/10.1109/ICRA.2018.8460487>
9. Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., ... & Zieba, K. (2016). End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*.
10. Kendall, A., Hawke, J., Janz, D., Mazur, P., Reda, D., Allen, J. M., ... & Shah, A. (2019). Learning to drive in a day. *2019 International Conference on Robotics and Automation*, 8248-8254. <https://doi.org/10.1109/ICRA.2019.8793742>
11. Dosovitskiy, A., Ros, G., Codevilla, F., López, A., & Koltun, V. (2017). CARLA: An open urban driving simulator. *Proceedings of the 1st Annual Conference on Robot Learning*, 1-16.
12. Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., ... & Beijbom, O. (2020). nuScenes: A multimodal dataset for autonomous driving. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11621-11631.
<https://doi.org/10.1109/CVPR42600.2020.01164>
13. Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., ... & Angelov, D. (2020). Scalability in perception for autonomous driving: Waymo open dataset. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2446-2454. <https://doi.org/10.1109/CVPR42600.2020.00252>
14. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*.
15. Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.

16. Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, 79(1), 119-157.
17. 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.
<https://doi.org/10.1177/2053951716679679>
18. Shalev-Shwartz, S., Shammah, S., & Shashua, A. (2017). On a formal model of safe and scalable self-driving cars. *arXiv preprint arXiv:1708.06374*.
19. Koopman, P., & Wagner, M. (2017). Autonomous vehicle safety: An interdisciplinary challenge. *IEEE Intelligent Transportation Systems Magazine*, 9(1), 90-96.
<https://doi.org/10.1109/MITS.2016.2583491>
20. SAE International. (2021). Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles (J3016_202104). SAE International.
https://doi.org/10.4271/J3016_202104
21. National Highway Traffic Safety Administration. (2021). Standing general order on crash reporting for incidents involving ADS and Level 2 ADAS. NHTSA.
22. European Commission. (2021). Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). European Commission.
23. Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54-63. <https://doi.org/10.1145/3381831>
24. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3645-3650. <https://doi.org/10.18653/v1/P19-1355>
25. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.