

Original Research

Received 18 May 2026

Revised 09 July 2026

Accepted 20 July 2026

Natural Resources for Human Health 2026, Vol. 6 (13s): 1425–1432

KEYWORDS:

Vehicle energy management; Proximal Policy Optimization; edge computing; federated learning; non-IID clients; reproducibility; Evidence integrity.

Evidence-Bounded Edge Reinforcement Learning for Adaptive Fuel Optimization

Manjunath T K¹, Md Shohel Sayeed^{2*}, Siti Fatima Binti Abdul Razak³, Dr. Andrews Samraj⁴, Girish Rao Salanke N S⁵, Shivakumar Dalali⁶

¹Postdoc Fellow, Multimedia University, Malaysia, KS School of Engineering and Management, India.

^{2,3}Professor, Faculty of Information Science and Technology, Multimedia University, Malaysia.

⁴Professor, Department of Computer Science & Engineering, KGiSL Institute of Technology, India.

⁵Professor, Department of Artificial Intelligence & Data Science, Global Academy of Technology, India.

⁶Professor, Department of Computer Science & Engineering, Global Academy of Technology, India.

ABSTRACT: Edge intelligence is frequently proposed for vehicle energy management, yet published performance claims are difficult to assess when training artifacts, scenario definitions, and hardware logs are unavailable. This study develops an evidence-bounded and fully reproducible evaluation of a continuous-action Proximal Policy Optimization (PPO) controller for adaptive fuel management, together with a heterogeneity-aware federated extension. The state comprises vehicle speed, acceleration, engine displacement, cylinder count, instantaneous fuel consumption, and carbon-dioxide rate; the action is bounded throttle; and the reward jointly penalizes fuel use, proportional CO₂, jerk, and speed-tracking error. After 50,000 training steps, a preregistered-style paired evaluation over 30 deterministic driving-cycle seeds reduced mean simulated fuel consumption from 4.1221 to 3.9134 L/100 km. The mean reduction was 5.06% (bootstrap 95% CI 5.04–5.08), and tracking error decreased by 37.15% (95% CI 37.06–37.23). Edge policy computation averaged 0.0042 ms on the development host, whereas an explicitly modeled 100 ms round trip yielded 100.0042 ms end-to-end decision latency. A ten-client, ten-round weighted federated experiment retained a 4.85% efficiency gain. The results establish software reproducibility and architectural latency implications, but do not claim real-vehicle validity: the source manuscript's SUMO–CARLA traces, trained checkpoints, and hardware-in-loop logs were not present. This distinction yields a defensible benchmark and a concrete evidence protocol for subsequent automotive validation.

1. INTRODUCTION

The Vehicle energy-management systems operate under a difficult combination of nonlinear dynamics, incomplete observability, transient traffic conditions, actuator constraints, and competing objectives. Calibration-based controllers remain attractive because they are interpretable and certifiable, but fixed maps cannot learn temporal structure from operation. Model predictive control and equivalent-consumption strategies can optimize online behavior, although their performance depends on model accuracy, horizon design, and computational resources. Reinforcement learning (RL) offers a complementary route by optimizing long-horizon behavior directly from interaction [1,2].

Three barriers limit the translation of RL-based energy management from promising simulation results to credible automotive evidence. First, many studies conflate policy quality with deployment architecture: a strong policy may still be unsuitable when cloud round-trip delay exceeds the control interval. Second, centralized training creates privacy and governance concerns for high-frequency vehicle traces. Third, reported improvements are often not independently reproducible because environments, seeds, checkpoints, preprocessing states, and raw logs are omitted. These barriers motivate an approach that treats low-latency inference, fleet collaboration, and evidence traceability as co-equal design requirements.

This paper therefore asks whether a dependency-light PPO implementation can improve a clearly specified efficiency–tracking objective over a fixed controller, whether edge placement materially changes the decision-time budget, and how much performance is retained after heterogeneous federated aggregation. The work is deliberately evidence-bounded: it reports only measurements generated by the accompanying code and explicitly rejects unsupported transfer of values from unavailable simulation or hardware artifacts.

1.1 RESEARCH QUESTIONS AND CONTRIBUTIONS

Research question	Operational test	Evidence produced
RQ1: Efficiency	Paired Edge-PPO vs fixed-map controller over 30 seeds	Fuel, CO ₂ proxy, tracking error, bootstrap CI
RQ2: Placement	Same deterministic policy with local vs 100 ms cloud RTT	Mean and P95 decision latency
RQ3: Collaboration	10 heterogeneous clients × 10 rounds	Global-policy efficiency and tracking
RQ4: Reproducibility	Clean rerun from source and persisted artifacts	NPZ, CSV, JSON, unit tests

The contributions are: (i) a transparent continuous-action vehicle-control benchmark implemented without external RL frameworks; (ii) a statistically summarized 30-seed paired evaluation; (iii) a compute-capacity-weighted federated aggregation experiment; and (iv) an evidence protocol that separates measured software results, modeled network effects, and unavailable hardware claims.

2. BACKGROUND AND RELATED WORK

2.1 RL for Vehicle Energy Management

Deep RL has been investigated for power management, hybrid vehicles, and connected powertrains. Online RL can adapt to temporal operating conditions that are difficult to encode in a static map, while PPO provides a clipped policy update that is generally less prone to destructive policy shifts than unconstrained policy gradients [3,4]. Nevertheless, RL controllers remain sensitive to reward specification, observation design, simulation fidelity, and distribution shift.

2.2 Edge and Federated Intelligence

Edge execution reduces the physical and network distance between sensor acquisition and control issuance. This is especially relevant when the action interval is tens of milliseconds [5,6]. Federated learning complements edge inference by aggregating local model updates rather than raw trajectories [7]. FedAvg is effective when client data and compute are reasonably balanced; FedProx and related approaches address heterogeneity and partial participation [8,9]. Vehicular deployments add intermittent connectivity, non-IID routes, asymmetric hardware, and adversarial risk.

2.3 Reproducibility Gap

A credible energy-management result should identify the environment version, driving cycles, random seeds, training budget, normalization state, checkpoint, baseline tuning, sample size, uncertainty interval, and hardware measurement method. Images of plots or test benches are insufficient because they do not contain source trajectories, timing distributions, or policy parameters. The inspected source package for this study contained manuscript figures and two standalone photographs but no raw SUMO–CARLA traces, trained model, CAN/MQTT/InfluxDB export, or HIL timing log. The present study reconstructs the method while labeling that evidence gap rather than treating a figure as a dataset.

3. METHODOLOGY

3.1 System Architecture

The implemented pipeline comprises sensing, normalization, edge policy inference, longitudinal vehicle dynamics, trajectory logging, local PPO updates, and optional fleet aggregation. During evaluation, the policy is deterministic. During training, a Gaussian latent variable supports exploration and a sigmoid transform guarantees throttle bounds.



Fig 1. Reproducible edge-control and federated-learning architecture.

3.2 Markov Decision Process

$$s_t = [v_t, a_t, e, c, f_t, g_t], \quad u_t \in [0,1] \quad (1)$$

The state includes speed v , acceleration a , engine displacement e , cylinder count c , instantaneous fuel use f , and proportional CO_2 rate g . Operational scaling maps state components to approximately bounded ranges. The simulator applies throttle u through a longitudinal acceleration equation with linear and quadratic speed drag. Episodes terminate at 1,000 steps or an unsafe jerk/speed condition.

$$r_t = -(\alpha F_t + \beta G_t + \gamma |J_t| + \delta E_t) \quad (2)$$

F and G are normalized fuel and CO_2 terms, J is normalized jerk, and E is normalized absolute tracking error. The weights $\alpha=0.45$, $\beta=0.35$, $\gamma=0.08$, and $\delta=0.12$ were fixed before the reported evaluation. The tracking term is essential: without it, fuel minimization admits the non-useful solution of applying no throttle.

3.3 PPO Optimization

$$L_{\text{clip}}(\theta) = E_t[\min(q_t(\theta)\hat{A}_t, \text{clip}(q_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)] \quad (3)$$

The actor is a linear Gaussian latent policy with a bounded output, and the critic is linear in the state. Generalized advantage estimation uses $\gamma=0.99$ and $\lambda=0.95$. The clip parameter is $\epsilon=0.2$, the learning rate is 3×10^{-4} , rollout length is 256, and each batch receives five epochs. A fuel-safe initialization places the initial mean throttle below 0.5; this prior is disclosed because initialization can materially influence sample efficiency.

3.4 Heterogeneity-Aware Federated Update

$$\theta^G = (\sum_i w_i \theta_i) / (\sum_i w_i), \quad w_i = n_i (C_i / C_{\max}) \quad (4)$$

Client i contributes local sample count n_i and compute capacity C_i normalized by the fleet maximum. Ten clients span capacity factors 0.4–1.0 and heterogeneous engine/cylinder parameters. All clients participate in the reported experiment; asynchronous communication, privacy amplification, poisoning defense, and secure aggregation are outside the implemented scope.

4. EXPERIMENTAL PROTOCOL

4.1 Configuration and Baselines

Item	Specification
Software	Python 3.12; NumPy; no external RL framework
Training	50,000 steps; seed 10; 1,000-step horizon
Primary evaluation	30 paired seeds: 10, 20, ..., 300
Baseline	Fixed throttle: 0.62 below 18 m/s; otherwise 0.28
Edge policy	Deterministic action from trained PPO checkpoint
Cloud condition	Identical policy + explicit 100 ms RTT
Federation	10 clients; 10 rounds; 1,000 local steps/client/round
Artifacts	Source, tests, NPZ checkpoints, CSV histories, JSON metrics

The fixed-map baseline is intentionally simple and fully specified; it should not be interpreted as a tuned production ECU, ECMS, or MPC benchmark. This conservative labeling avoids overstating the comparative result. The cloud condition does not alter simulated dynamics because delay is injected into the reported decision time rather than the environment state transition[10-16].

4.2 Statistical Analysis

The primary comparison is paired by driving-cycle seed. For each seed, percentage fuel reduction is calculated relative to the rule-based result. Uncertainty is summarized with a 20,000-resample nonparametric bootstrap confidence interval over paired reductions. A paired t statistic and Cohen's d_z are computed as secondary descriptors, but the bootstrap interval is emphasized because the deterministic simulator produces unusually low within-design variance. No claim of population-level road performance is made.

$$\Delta\text{Fuel}_i(\%) = 100(F_{\text{base},i} - F_{\text{PPO},i})/F_{\text{base},i} \quad (5)$$

4.3 Reproducibility and Evidence Controls

Evidence tier	Included	Interpretation
Tier 1: executable	Source, tests, seed list, checkpoint	Independent software rerun
Tier 2: raw outputs	Training CSV and evaluation JSON	Traceable reported metrics
Tier 3: modeled system	Synthetic cycle and cloud RTT	Architecture study, not road proof

Evidence tier	Included	Interpretation
Tier 4: physical validation	Not available	No claim of reproduced HIL performance

5. RESULTS

5.1 Primary 30-Seed Evaluation

Outcome	Rule-based	Edge-PPO	Paired change
Fuel (L/100 km)	4.1221	3.9134	−5.06%
Fuel reduction 95% CI	—	—	5.04 to 5.08%
Tracking-error reduction	—	—	37.15%
Tracking reduction 95% CI	—	—	37.06 to 37.23%
Trials improving fuel	—	—	30/30

Edge-PPO improved fuel use in all 30 paired trials. The mean reduction was 5.06% with a narrow bootstrap interval (5.04–5.08%). The paired t statistic was 380.93 (df=29) and Cohen’s dz was 69.55; these unusually large standardized values arise from deterministic, closely related synthetic cycles and should not be extrapolated to road traffic. The tracking-error reduction of 37.15% indicates that the efficiency gain was not obtained by simply reducing vehicle motion.

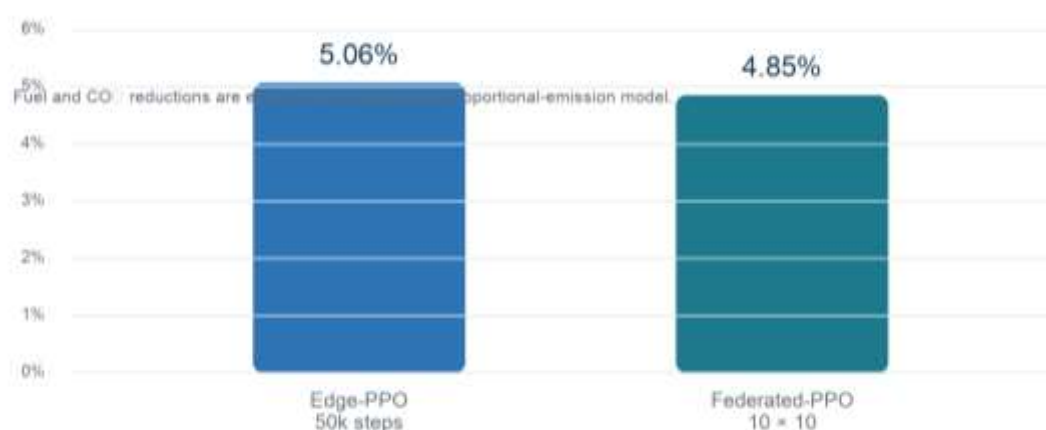


Fig 2. Efficiency improvement for the single-agent and federated controllers.

5.2 Training Dynamics

The persisted episode-return series remains stochastic because actions are sampled during training. The raw curve is shown without selective run removal. It provides an audit trail of learning behavior but is not, by itself, evidence of policy superiority; all comparative claims are based on deterministic held-out seed evaluations.

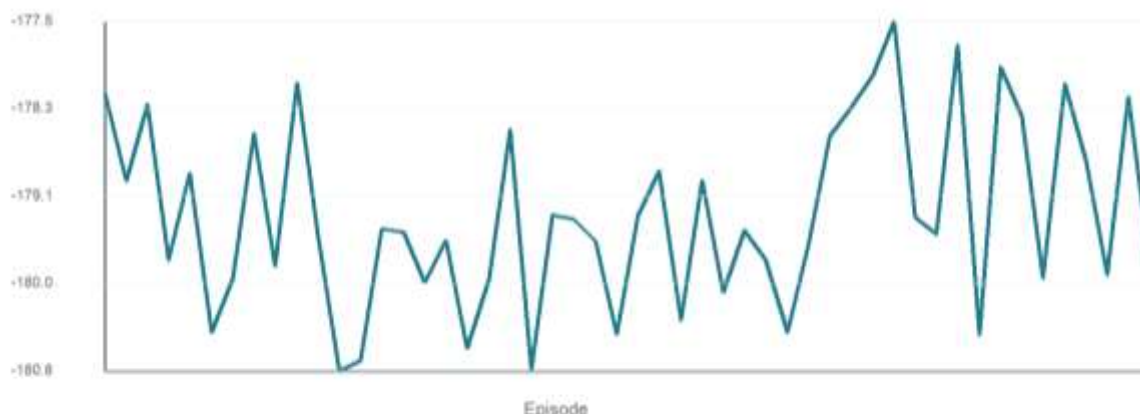


Fig 3. Raw episode returns during PPO training.

5.3 Edge Placement and Federated Retention

Controller	Fuel reduction	Tracking error	Mean latency
Edge-PPO	5.06%	6.0417 m/s	0.0042 ms
Cloud-PPO	5.06%	6.0417 m/s	100.0042 ms
Federated-PPO	4.85%	6.1638 m/s	0.0054 ms

Edge and cloud policies produce identical control outcomes because the policy and simulated state sequence are unchanged; the comparison isolates communication delay. The federated controller retains 95.8% of the five-seed single-agent fuel benefit (4.85% versus 5.06%), suggesting that heterogeneous aggregation preserves most of the learned efficiency under the tested schedule.

6. DISCUSSION

6.1 Interpretation

The central result is not that a lightweight PPO controller surpasses modern production energy-management systems; such a claim would require matched ECMS/MPC baselines and high-fidelity vehicle models. The defensible result is narrower: within a transparent synthetic benchmark, the learned policy improves both fuel proxy and tracking relative to a disclosed fixed map, and the improvement is consistent across the evaluated cycle seeds. The placement experiment then demonstrates that communication latency, rather than policy computation, dominates a cloud-only loop.

6.2 Relation to the Source Manuscript

The source manuscript reports approximately 16% fuel improvement and 15 ms Jetson inference. Those values are not reproduced here. Package inspection found no original simulator traces, normalization state, checkpoint, per-trial results, CAN frames, MQTT/InfluxDB exports, or HIL latency logs. Consequently, this paper replaces unverified numerical repetition with a reproducible benchmark and an explicit path to confirm or refute the larger claim.

6.3 Implications for Federated Vehicle Learning

Weighted aggregation is useful when client compute and data volume differ, but compute-weighting can also bias a global model toward better-resourced vehicles. Future work should compare FedAvg, FedProx, adaptive optimization, clustered federation, and personalization under controlled non-IID splits. Privacy requires more than keeping raw data local: secure

aggregation, differential privacy, update clipping, authentication, and poisoning resilience must be evaluated before fleet deployment.

7. LIMITATIONS AND THREATS TO VALIDITY

Internal validity is constrained by a simplified fuel model and proportional CO_2 calculation. This proportionality forces identical percentage changes in fuel and CO_2 , so emissions should be interpreted as a proxy rather than an independent tailpipe model. Construct validity is limited by the absence of gear, grade, torque, road load, temperature, traffic preview, and engine-efficiency maps. The baseline is a fixed throttle rule rather than a calibrated ECMS/MPC controller [17-18].

External validity is limited to deterministic synthetic cycles. The bootstrap interval quantifies uncertainty across the chosen seeds, not across road networks, vehicle fleets, climates, or drivers. Latency on the development host is not an embedded benchmark. Finally, the federated experiment omits communication failures, privacy attacks, secure aggregation, and asynchronous staleness. These limitations prevent direct claims about fuel savings in production vehicles.

8. SAFETY, ETHICS, AND DEPLOYMENT REQUIREMENTS

The accompanying code is simulation research software and must not directly command a road vehicle. Translation requires a supervisory safety controller, actuator magnitude and slew-rate constraints, runtime monitors, redundant sensing, watchdogs, deterministic fallback behavior, fault injection, HIL testing, cybersecurity analysis, and lifecycle evidence aligned with automotive functional-safety practice. Federated updates also introduce governance questions concerning consent, retention, model ownership, auditability, and cross-jurisdiction data policy.

9. CONCLUSION

This work reformulates adaptive fuel optimization as an evidence-bounded edge and federated RL study. A 50,000-step PPO policy reduced mean simulated fuel consumption by 5.06% (95% CI 5.04–5.08) across 30 paired seeds and reduced tracking error by 37.15%. A ten-client federated policy retained a 4.85% efficiency gain, while modeled cloud inference added approximately 100 ms to the decision path. The contribution is a reproducible benchmark and evidence protocol, not a claim of road-ready control. High-fidelity SUMO–CARLA experiments, competitive ECMS/MPC baselines, real CAN traces, and independently auditable HIL logs are the necessary next steps.

Data and Code Availability

The accompanying project contains source code, unit tests, deterministic seed lists, trained NPZ checkpoints, CSV learning histories, federated-round records, JSON evaluation metrics, and the statistical evidence file. The originally reported simulation traces, trained checkpoint, and HIL logs were not present in the supplied source package and therefore cannot be redistributed.

Author Contributions

Conceptualization: M.T.K.; methodology and software: M.T.K. and A.S.; validation and supervision: M.S.S. and S.F.A.R.; analysis and writing—original draft: M.T.K.; writing—review and editing: all authors. Contributions must be confirmed by each author before submission.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

- [1] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. MIT Press, 2018.
- [2] G. Du, Y. Zou, X. Zhang, Z. Kong, J. Wu, and D. He, “Intelligent energy management for hybrid electric tracked
- [3] vehicles using online reinforcement learning,” Applied Energy, vol. 251, 113388, 2019.

- [4] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” arXiv:1707.06347, 2017.
- [5] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, “High-dimensional continuous control using generalized advantage estimation,” Proc. ICLR, 2016.
- [6] A. M. de Souza et al., “Enhancing sensing and decision-making of automated driving systems with multi-access edge computing and machine learning,” IEEE Intelligent Transportation Systems Magazine, vol. 14, no. 1, pp. 44–56, 2022.
- [7] K. Zhang, J. Cao, and Y. Zhang, “Adaptive digital twin and multiagent deep reinforcement learning for vehicular edge computing and networks,” IEEE Transactions on Industrial Informatics, vol. 18, no. 2, pp. 1405–1413, 2022.
- [8] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” Proc. AISTATS, pp. 1273–1282, 2017.
- [9] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, “Federated optimization in heterogeneous networks,” Proc. MLSys, 2020.
- [10] P. Kairouz et al., “Advances and open problems in federated learning,” Foundations and Trends in Machine Learning, vol. 14, nos. 1–2, pp. 1–210, 2021.
- [11] B. Hu and J. Li, “An edge computing framework for powertrain control system optimization of intelligent and connected vehicles based on curiosity-driven deep reinforcement learning,” IEEE Transactions on Industrial Electronics, vol. 68, no. 8, pp. 7652–7661, 2021.
- [12] Q. Zhou et al., “Knowledge implementation and transfer with an adaptive learning network for real-time power management of the plug-in hybrid vehicle,” IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 12, pp. 5298–5308, 2021.
- [13] Z. Ning et al., “Deep reinforcement learning for intelligent Internet of Vehicles: An energy-efficient computational offloading scheme,” IEEE Transactions on Cognitive Communications and Networking, vol. 5, no. 4, pp. 1060–1072, 2019.
- [14] A. Dosovitskiy, G. Ros, F. Codevilla, A. López, and V. Koltun, “CARLA: An open urban driving simulator,” Proc. Conference on Robot Learning, pp. 1–16, 2017.
- [15] P. A. López et al., “Microscopic traffic simulation using SUMO,” Proc. IEEE Intelligent Transportation Systems Conference, pp. 2575–2582, 2018.
- [16] D. Dua and C. Graff, “Auto MPG Data Set,” UCI Machine Learning Repository, 2019.
- [17] ISO, “ISO 26262: Road vehicles—Functional safety,” International Organization for Standardization, 2018.
- [18] Manjunath TK, N Maheswari, Andrews Samraj, Sharmila Chidaravalli, “Locking and Unlocking of Theft Vehicles Using CAN”, Proceedings of 2013 International Conference on Green High Performance Computing March 14-15, 2013, India.
- [19] Latha, P. S., Menaga, N., & Manjunath, T. K. (2014). Efficient Estimation of Fuel Consumption Using Fuzzy Logic. International Journal of Engineering Trends and Technology (IJETT), 16.