# Introduction

The rapid adoption of electric vehicles (EVs) is reshaping transport and energy systems, driven by ambitious decarbonisation targets and urban air quality policies [8]. In London, this transition has been supported by the expansion of charging infrastructure [10], yet challenges persist. The charging landscape remains fragmented across multiple operators, with uneven station distribution [9] and tariffs that combine per-kWh rates, session fees, idle penalties, and membership discounts [12, 13]. For drivers, the key concerns are minimising costs, reducing delays, and ensuring trip feasibility, while operators and regulators focus on infrastructure adequacy and grid stability. Existing optimisation efforts have often prioritised the grid perspective [1–3], but have struggled to capture the complex, user-centred realities of urban charging.

Several gaps in the literature constrain the applicability of current approaches to dense metropolitan contexts such as Inner London. Many studies assume uniform or simplified tariffs, overlooking operator heterogeneity, idle fees, and subscription models [32–35]. Charging dynamics are frequently simplified, with constant or generic rates replacing vehicle-specific nonlinear charging curves [36–38]. Hybrid objectives balancing cost and time remain underexplored: weighted sums and Pareto methods are sometimes applied, but calibration is often ad hoc [39–41]. Reinforcement learning (RL) has shown promise [5–7, 21, 22], yet many studies rely on oversimplified environments that neglect congestion, station cycling, or realistic constraints [16, 25, 45]. Reward design also poses difficulties, with prior work often fixing priorities or underusing shaping and penalties [42–44].

This thesis addresses these gaps by applying Proximal Policy Optimisation (PPO) to model EV charging decisions in Inner London. Using structured datasets, the study integrates operator-specific tariffs, station metadata, and vehicle charging curves to simulate realistic costs and charging times. A custom RL environment was built using SUMO for traffic modelling and OpenAI Gym for agent training, with constraints to prevent unrealistic behaviours such as repeated station cycling. Reward functions were designed to capture cost, time, and hybrid objectives, incorporating shaping and feasibility penalties. The methodology emphasises reproducibility through structured datasets, fixed seeds, and consistent evaluation.

The research is guided by four questions:

1. How can reinforcement learning minimise total trip cost while ensuring state of charge (SoC) remains above a predefined reserve threshold?
2. How can reinforcement learning minimise total trip duration while maintaining sufficient SoC for journey completion?
3. How can a hybrid reinforcement learning framework balance cost and time objectives under varying operational constraints?
4. How do cost-based, time-based, and hybrid optimisation objectives differ in shaping charging behaviour and route selection, and what trade-offs emerge between them?

By addressing these questions, the study contributes a reinforcement learning framework that captures the technical and behavioural complexities of EV charging in Inner London.

# Literature Review

Before reinforcement learning gained prominence, EV charging optimisation was addressed through deterministic and heuristic approaches. Mixed-integer linear programming (MILP) and linear programming (LP) have been widely applied to siting, scheduling, and grid integration, valued for transparency and their ability to guarantee optimal solutions under fixed assumptions [19]. However, their computational burden scales poorly with the number of vehicles and stations, making them impractical for real-time urban applications. To improve scalability, metaheuristic techniques such as genetic algorithms (GA) and particle swarm optimisation (PSO) have been employed [20]. A related strand of research focuses on the electric vehicle routing problem (EVRP), which extends traditional vehicle routing by incorporating charging constraints. Early EVRP formulations treated stations as deterministic service nodes, overlooking congestion, waiting times, and nonlinear charging curves [16]. These limitations have motivated the move towards adaptive, data-driven approaches, with reinforcement learning offering a compelling alternative.

Technical constraints from vehicle charging curves further complicate optimisation. Battery charging is nonlinear: power is high at lower states of charge (SoC) but tapers significantly as SoC approaches full capacity, meaning marginal charging times increase disproportionately. Some studies have incorporated this tapering into optimisation models [17, 18], yet many still assume constant charging rates, leading to underestimation of charging duration. Few studies account for the vehicle-specific nature of EV charging curves. Most models use a generic nonlinear charging profile or even assume constant charging power, failing to capture how different battery chemistries and vehicle BMS algorithms yield distinct tapering behaviours [36]. In reality, charging performance can vary widely across EV models [38], but a lack of high-quality public data has constrained realism. Until recently, empirical charging datasets were small and limited in scope, impeding fine-grained modelling [37]. This data scarcity has forced researchers to rely on simplified or synthetic curves, and to treat all vehicles uniformly [36]. New large-scale datasets now highlight the heterogeneity in charging behaviour across vehicle types [38], underscoring that generic assumptions may misestimate charging duration and energy delivered. Improving data availability and incorporating vehicle-specific charging characteristics remain critical to enhance model accuracy [37, 38]. Combined with tariff heterogeneity, this creates cost–time trade-offs: users may minimise cost by choosing slower, cheaper stations at the expense of longer travel or waiting times, while those prioritising time gravitate towards ultra-rapid but more expensive chargers. These challenges illustrate why optimisation cannot focus solely on grid stability; any practical framework must reflect user behaviour, tariff fragmentation, and charging dynamics in dense urban settings such as London.

Despite the complexity of real-world tariff structures, many prior EV routing and charging optimisation studies have assumed uniform or simplified pricing, overlooking operator-specific tariffs and membership discounts [32, 33]. Only recently have researchers begun to address this heterogeneity: for example, game-theoretic models consider competition between charging providers with different pricing schemes [34], and user preference studies show strong interest in interoperability across networks [32]. Ignoring features like membership plans or idle (overstay) fees can lead to suboptimal or unrealistic recommendations [35]. By contrast, incorporating these tariff nuances (e.g., discounted member rates vs. higher guest prices, per-minute idle penalties) can significantly alter optimal routing decisions [35, 34], as drivers may trade off cost savings from memberships against travel time or detour length [32].

RL offers significant advantages by enabling agents to learn adaptive policies through interaction with dynamic environments, something deterministic and heuristic methods cannot achieve. Applications in energy systems demonstrate RL’s ability to reduce costs, flatten grid loads, and improve user satisfaction [21, 22]. Early EV charging studies used DQN, which showed success in cost reduction and scheduling efficiency [5, 6], though discrete action spaces limited their suitability for continuous charging problems. Policy gradient methods, such as PPO, address this limitation by directly optimising policies for continuous action spaces [7]. PPO has gained traction for its stability and suitability for multi-objective problems, and has outperformed value-based methods in scalability and convergence [23]. Extensions such as DDPG and SAC have also been explored for EV charging [31]. Despite these advances, few studies apply PPO at the driver level in dense cities, where tariffs, nonlinear charging curves, and congestion interact.

The effectiveness of RL also depends on training frameworks and simulation tools. Value-based methods such as DQN have been widely used for scheduling [5, 6], but their discrete action space makes them unsuitable for continuous decision problems. For this project, PPO is implemented in OpenAI Gym, which provides a modular and reproducible structure for defining states, actions, and rewards [24]. To capture mobility constraints, the Simulation of Urban Mobility (SUMO) is incorporated, providing spatial and temporal modelling of traffic and congestion [25]. Together, Gym and SUMO provide a realistic framework for evaluating PPO in dense urban EV charging scenarios. Many earlier EV routing studies made simplifying assumptions that omit key real-world dynamics. For instance, deterministic travel times and unconstrained charger access were often assumed – early EV routing formulations ignored road congestion and charging station queues entirely [16]. In contrast, recent works stress simulation realism by integrating traffic models and operational constraints. The use of SUMO coupled with OpenAI Gym interfaces allows researchers to expose learning agents to time-varying traffic congestion, realistic travel delays, and competition for charging stalls [45]. Multi-agent simulations of electric taxi fleets, for example, now incorporate SUMO to prevent implausible manoeuvres and station overuse [45]. While adding such realism increases computational complexity, it avoids the optimistic bias of overly simplified environments. Studies deploying these high-fidelity simulators report more credible and policy-relevant results, as agents learn strategies (e.g., congestion avoidance, smart station selection) that would actually work under real urban conditions [45].

Reward function design is central to RL’s success. Studies typically target cost minimisation, time reduction, or hybrid combinations, but most fix reward priorities instead of dynamically balancing them [21, 22]. Reward function formulation in EV charging RL has evolved to facilitate learning of complex behaviours. Potential-based reward shaping is one proven technique: by adding an extra reward signal derived from a heuristic “potential” (such as remaining distance or battery level), one can guide the agent toward good decisions without changing the optimal policy [42]. Researchers also integrate penalty terms for undesirable or infeasible actions. A common practice is assigning a large negative reward if the vehicle’s battery is depleted en route [43], which penalises infeasible paths and trains the agent to avoid running out of charge. Likewise, penalties can discourage behaviours that, while technically feasible in simulation, are impractical – such as excessive detours or unnecessary charging stops. Charging costs themselves are typically represented as negative rewards in the RL framework [43]. By defining the reward as the negative of monetary cost, plus penalties for delay or battery wear, prior studies ensure the RL agent treats charging duration and expense as something to minimise [43, 44]. Overall, careful reward shaping – combining informative incentives with strong penalties – is crucial for efficient and realistic learning in EV charging scenarios [42, 43].

Finally, EV charging inevitably involves trade-offs between cost and time. Cost-oriented approaches minimise user expenditure by leveraging dynamic pricing or demand response, often reducing bills while supporting grid stability [26, 27]. However, these generally overlook time constraints, which are critical in urban mobility. Time-focused approaches instead reduce charging duration, waiting, or travel delays, often directing users to ultra-rapid but more expensive stations [28, 16]. Hybrid frameworks explicitly address this challenge, employing weighted sums or Pareto methods to capture trade-offs between cost and time [29]. Prior work on multi-objective EV charging optimisation has explored various scaling techniques to balance cost and time. A common approach is to convert time into an equivalent monetary cost via a value-of-time parameter, enabling a single-objective formulation [39]. Other studies employ weighted sums with tuned coefficients to reflect user preferences or policy priorities [40]. Alternatively, some researchers use Pareto-front or evolutionary reinforcement learning techniques to avoid preset weights [40]. These methods generate a set of efficient solutions, capturing the inherent trade-off so that stakeholders can choose based on an implied value-of-time. Overall, incorporating an explicit cost–time trade-off (either through value-of-time scaling [41] or multi-objective frameworks [40]) is essential to produce realistic routing policies that neither minimise cost at unacceptable delay nor save time at unreasonable expense. RL, particularly PPO, is well suited to this setting, adapting policies to continuous spaces and balancing competing rewards [7]. Hybrid RL models have shown promise in jointly minimising cost, time, and even secondary objectives such as battery degradation [30]. Yet hybrid optimisation remains underexplored in real-world urban contexts where tariffs are operator-specific, charging curves nonlinear, and congestion highly variable. The present study therefore applies PPO to design cost-based, time-based, and hybrid charging strategies in Inner London, explicitly integrating real-world tariffs, nonlinear charging curves, and traffic constraints into a unified optimisation framework.

Methodology

**Data Preparation**

The methodology relied on four structured data components: tariffs, charging stations, vehicle specifications with charging curves, and simulated user demands.

**Tariff data** were collected directly from operator websites (SureCharge, Believ etc.). The raw data exhibited heterogeneous formats, including per-kWh energy rates, session fees, idle charges, and membership discounts etc.

A screenshot of a computer

AI-generated content may be incorrect.A close-up of a message

AI-generated content may be incorrect.

Figure 1 Raw Tariff Information Examples

To enable consistent cost computation, we converted all rates to £/kWh and decomposed the tariffs into three structured datasets. This ensured realistic and provider-specific cost estimation in the simulation environment.

A diagram of a manual selection

AI-generated content may be incorrect.

Figure 2 Conversion Process

**Charging station data** were obtained from the UK National ChargePoint Registry (NCR) [1]. After removing duplicates and fixing coordinate errors, we split the dataset into two layers: station-level metadata (for spatial queries) and connector-level features (for action feasibility in the RL environment).

A diagram of a diagram

AI-generated content may be incorrect.

Figure 3 Splitting the Charging Data

**Vehicle specifications** were sourced from Open EV Data v2 [2]. Charging curves were interpolated over 0–100% state of charge (SoC), expressed in kWh/km, and capped by station power limits to model charging times and SoC changes.

A diagram of an electrical system

AI-generated content may be incorrect.

Figure 4 Splitting the EV Data

**Simulated users** were generated as structured trip instances combining user profiles, EV states, trip definitions, and behavioural constraints (see Table 1).

|  |  |
| --- | --- |
| **Feature Category** | **Examples** |
| User Profile | Membership type (Member/Payg), Subscription status, Sessions per month |
| Trip Definition | Origin-Destination coordinates, Departure time, Trip distance |
| EV State | Assigned EV model, Start SoC, Reserve SoC, Available energy |
| Preferences & Constraints | Optimization objective, max detour km |
| Energy Efficiency | Consumption in kWh/km, estimated required kWh per trip |

Table 1 Feature Categories & Examples

Together, these four components provided a unified and reproducible dataset for training and evaluating the RL agents. By integrating these, the environment captures both system-level constraints and realistic user behaviour.

**Environment Design**

The charging decision-making task was modelled as a Markov Decision Process (MDP),

where **S** is the state space, **A** the action space, **P** the transition dynamics, **R** the reward function, and the discount factor.

**State representation.** Each state contained the vehicle’s SoC, position, remaining distance, and features of up to k candidate stations (connector type, power, tariff, detour). Telemetry such as visited stations and last charging time was also tracked for analysis.

**Action space.** The agent could select one of the candidate stations or continue driving. Invalid options (repeat visits, cooldown violations, exceeding charge limits) were masked before routing, with a low-SoC override ensuring stations were always available in emergencies

**Transitions.** Driving reduced SoC in proportion to distance:

Where is energy consumption andbattery capacity. Travel time was obtained from SUMO’s microscopic simulation:

**.**

Charging followed nonlinear battery curves capped by connector power, with efficiency and a fixed 3-minute overhead:

**Termination.** Episodes ended when the destination was reached or if SoC fell below a reserve threshold

**Constraints and reproducibility.** Rules limited charges per trip, enforced minimum gaps between sessions, and prohibited repeat stations. Fixed seeds across environment and training components ensured reproducibility of results.

**Design trade-offs.** Simplifications were introduced to ensure meaningful charging decisions. Table 2 summarises the key adjustments and their effects.

|  |  |  |  |
| --- | --- | --- | --- |
| **Aspect** | **Real-world baseline** | **Simulation choice** | **Rationale** |
| Initial SoC | Often >50% | Sampled 10–30% | Ensures charging is frequently required, avoids trivial trips |
| Trip lengths | Many <10 km | Calibrated 12–25 km | 30–60% of trips require ≥1 charge, providing a learning signal |
| Charging | Overheads vary | Fixed 3-min per session | Penalises “nibbling” charges and station hopping |
| Traffic | Complex congestion patterns | SUMO microsim (vs. constant speed or multipliers) | Provides realistic congestion while keeping simulation deterministic |
| Station use | Drivers may revisit stations | No repeats, cooldowns, max charges | Prevents unrealistic cycling behaviour |
| Variability | High randomness in trips | Fixed seeds (environment + training) | Enables reproducibility and controlled comparisons |

Table 2 Environment Design Trade-offs

Together, these design elements produced an RL environment that was both computationally feasible and sufficiently realistic to evaluate charging strategies under urban driving conditions. While simplified in some respects, the use of SUMO-backed traffic and structured user demands ensured that the environment captured the essential challenges of EV route planning in Inner London.

**Reward Design**

The environment supported three optimization objectives. Terminal bonuses and penalties enforced feasibility, while charging was strictly net-negative. Potential-based shaping rewarded driving progress, and small anti-dither penalties reduced oscillatory behaviours. Table 3 summarises the key design choices, their implementation, and the rationale for each.

**1. Cost minimisation**

where ​ is the charging session cost in GBP.

**2. Time minimisation**

where is detour time and is charging duration.

**3. Hybrid objective**

Where is total journey time and cost in GBP, with denominators normalising scales to prevent dominance.

**4. Shaping.** To provide denser feedback, potential-based shaping rewarded reductions in estimated time-to-arrive (ETA):

Where  **. 60**. Shaping was neutral during charging, so only forward progress yielded positive signals.

**Summary of design choices.**

|  |  |  |
| --- | --- | --- |
| Design choice | Implementation | Rationale |
| Success / failure signals | +50 on trip completion, –200 on depletion | Rewards feasibility and strongly penalises stranding |
| Infeasible actions | –2 penalty per invalid station choice | Discourages wasted steps while keeping exploration possible |
| Charging cost | All charge steps net-negative (time + energy + overhead) | Ensures charging is necessary but never intrinsically rewarding |
| Potential shaping | ETA-based potential function | Provides denser feedback without biasing against cost minimisation |
| Anti-dither penalties | Small penalties for idle drive and micro-charges | Prevents oscillatory behaviour and “nibbling” charges |
| Charging overhead | Fixed 3 min per session | Reflects real-world setup delays and discourages frequent short sessions |
| Efficiency & limits | η=0.92\eta=0.92, connector caps | Models technical constraints, making some cheap stations unattractive |
| Hybrid scaling | Value of time £0.05/min | Keeps cost and time comparable, avoiding dominance of one metric |

Table 3 Summary of Design Choices

**Proximal Policy Optimization**

Policy learning used **Proximal Policy Optimisation (PPO)** from Stable-Baselines3, chosen for its stability and efficiency. The algorithm was left unmodified; the study’s contribution lay in environment, reward, and evaluation design.

The clipped surrogate objective is defined as:

Where is the probability ratio and is the advantage estimate.

Advantages were computed using Generalised Advantage Estimation (GAE) {CITATION NEEDED}:

Where is the temporal-difference error.

**Training setup.** Fixed seeds were applied across PPO, numpy, torch, and the environment to ensure reproducibility. Short smoke runs confirmed that KPI trajectories matched baseline expectations, providing a stable reference point for later experiments.

**Design choice rationale.**

* **PPO** was chosen for robustness and wide adoption in RL research.
* **No algorithm modifications** kept the focus on environment and reward design rather than algorithmic novelty.
* **GAE** reduced variance in advantage estimates, accelerating learning.
* **Reproducibility controls** ensured fair comparisons across reward definitions and environment settings.

**Evaluation overview.** All trained policies were assessed on fixed evaluation subsets under consistent seeds, tracking key performance indicators such as success rate, journey time, and charging cost. Full evaluation design and results are presented in Section 4.

Overall, the methodology combined structured datasets, a custom RL environment, carefully designed rewards, and reproducible PPO training, enabling the evaluation of EV charging strategies under realistic Inner London traffic conditions.

# References

[1] Mullan, M., Harries, D., Bräunl, T., & Whitely, S. (2018). The technical, economic and commercial viability of the electric vehicle to grid concept. Energy Policy, 109, 403–417.

[2] Kiani, H., Hesami, K., Azarhooshang, A., Pirouzi, S., & Safaee, S. (2021). Adaptive robust operation of the active distribution network including renewable and flexible sources. Sustainable Energy, Grids and Networks, 26, 100476.

[3] Weckx, S., & Driesen, J. (2015). Load balancing with EV chargers and PV inverters in unbalanced distribution grids. IEEE Transactions on Sustainable Energy.

[4] Xydas, E., Marmaras, C., & Cipcigan, L. (2016). A data-driven approach for characterising the charging demand of electric vehicles: A UK case study. Applied Energy, 162, 763–771.

[5] Qian, T., Shao, C., Wang, X., & Shahidehpour, M. (2019). Deep reinforcement learning for EV charging navigation by coordinating smart grid and intelligent transportation system. IEEE Transactions on Smart Grid, 11(2), 1714–1723.

[6] Han, Y., Li, T., & Wang, Q. (2024). A DQN-based approach for large-scale EVs charging scheduling. Complex & Intelligent Systems, 10, 8319–8339.

[7] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimisation algorithms. arXiv preprint arXiv:1707.06347.

[8] Sovacool, B. K., Noel, L., Axsen, J., & Kempton, W. (2018). The neglected social dimensions to a vehicle-to-grid (V2G) transition: a critical and systematic review. Environmental Research Letters, 13(1), 013001.

[9] Nicholas, M., & Lutsey, N. (2020). Quantifying the EV charging infrastructure gap in the UK. ICCT Report.

[10] Transport for London. (2021). London’s 2030 Electric Vehicle Infrastructure Strategy. TfL.

[11] Robinson, C., Blythe, P., Bell, M., Hübner, Y., & Hill, G. (2013). Analysis of electric vehicle driver recharging demand profiles and subsequent impacts on the electricity distribution network. Energy Policy, 61, 337–348.

[12] Figenbaum, E., & Kolbenstvedt, M. (2016). Learning from Norwegian battery electric and plug-in hybrid vehicle users: Results from a survey of vehicle owners. Institute of Transport Economics Report.

[13] Hardman, S., Jenn, A., Tal, G., Axsen, J., Beard, G., Daina, N., ... & Turrentine, T. (2018). A review of consumer preferences of and interactions with electric vehicle charging infrastructure. Transportation Research Part D, 62, 508–523.

[14] Zavvos, E., Gerding, E. H., & Brede, M. (2021). A comprehensive game-theoretic model for EV charging station competition. IEEE Transactions on Intelligent Transportation Systems.

[15] Lin, R., Chu, H., Gao, J., & Chen, H. (2024). Charging management and pricing strategy of electric vehicle charging station based on mean field game theory. Asian Journal of Control, 26(2), 803-813.

[16] Montoya, A., Guéret, C., Mendoza, J. E., & Villegas, J. G. (2017). The electric vehicle routing problem with nonlinear charging function. Transportation Research Part B, 103, 87–110.

[17] Froger, A., Mendoza, J. E., Jabali, O., & Laporte, G. (2019). Improved formulations and algorithmic components for the EV routing problem with nonlinear charging functions. Computers & Operations Research, 104, 256–294.

[18] Luo, C., Huang, Y. F., & Gupta, V. (2019). Stochastic dynamic programming for EV charging station placement in urban areas. IEEE Transactions on Smart Grid, 10(2), 2272–2282.

[19] He, F., Wu, D., Yin, Y., & Guan, Y. (2013). Optimal deployment of public charging stations for plug-in hybrid electric vehicles. Transportation Research Part B: Methodological, 47, 87–101.

[20] García-Álvarez, J., González, M. A., & Vela, C. R. (2018). Metaheuristics for solving a real-world electric vehicle charging scheduling problem. Applied Soft Computing, 65, 292–306.

[21] Perera, A. T. D., & Kamalaruban, P. (2021). Applications of reinforcement learning in energy systems. Renewable and Sustainable Energy Reviews, 137, 110618.

[22] Xu, B., Rubenis, A., & Long, C. (2024). Reinforcement learning based smart charging for electric vehicle fleet. In International Symposium on Intelligent Technology for Future Transportation (pp. 375–383).

[23] Yu, C., Velu, A., Vinitsky, E., Gao, J., Wang, Y., Bayen, A., & Wu, Y. (2022). The Surprising Effectiveness of PPO in Cooperative Multi-Agent Games. NeurIPS 2022.

[24] Brockman, G., et al. (2016). OpenAI Gym. arXiv preprint arXiv:1606.01540.

[25] Behrisch, M., Bieker, L., Erdmann, J., & Krajzewicz, D. (2011). SUMO – Simulation of Urban Mobility: An Overview. Proceedings of SIMUL 2011, 55–60.

[26] Gunasekaran, R., Mohanraj, M.R., Senthilkumar, R., & Kamalakannan, R.S. (2025). Enhancing electric vehicle charging station utilization by reducing users waiting times through WOA-GLNN approach. Electrical Engineering.

[27] Zhang, L., Ji, Y., Li, X., Huang, Z., Cui, D., Chen, H., Gong, J., Breer, F., Junker, M., & Sauer, D.U. (2025). Multi-objective charging scheduling for electric vehicles at charging stations with renewable energy generation. Green Energy and Intelligent Transportation, 4(4), 100283.

[28] Cui, F., Lin, X., Zhang, R., & Yang, Q. (2022). Multi-objective optimal scheduling of charging stations based on deep reinforcement learning. Frontiers in Energy Research, 10, 1042882.

[29] Lin, R., Chu, H., Gao, J., & Chen, H. (2025). A Multi‑Objective Evolutionary Reinforcement Learning framework for optimizing EV charging strategies. Applied Energy, in press.

[30] Jia, R., Gao, K., Cui, S., et al. (2025). A multi‑objective RL‑based velocity optimization for electric trucks, considering battery degradation mitigation. Transportation Research Part E, in press.

[31] Yan, Z., Liu, Q., Wang, J., & Li, B. (2025). Deep reinforcement learning-based plug-in EV charging scheduling under uncertainty: a constrained SAC vs DDPG comparison. Energy.

[32] Visaria, A.A., Jensen, A.F., Thorhauge, M., Mabit, S.L. (2022). User preferences for EV charging, pricing schemes, and charging infrastructure. *Transportation Research Part A*, 165, 120–143.

[33] Hardman, S., Jenn, A., Tal, G., Axsen, J., Beard, G., Daina, N., et al. (2018). A review of consumer preferences and interactions with electric vehicle charging infrastructure. *Transportation Research Part D*, 62, 508–523.

[34] Zavvos, E., Gerding, E.H., Brede, M. (2021). A comprehensive game-theoretic model for EV charging station competition. *IEEE Transactions on Intelligent Transportation Systems*, 22(11), 7135–7149.

[35] Lin, R., Chu, H., Gao, J., Chen, H. (2024). Charging management and pricing strategy of electric vehicle charging stations based on mean field game theory. *Asian Journal of Control*, 26(2), 803–813.

[36] Froger, A., Mendoza, J.E., Jabali, O., Laporte, G. (2019). Improved formulations and algorithmic components for the EV routing problem with nonlinear charging functions. *Computers & Operations Research*, 104, 256–294.

[37] Baek, K., Lee, E., Kim, J. (2024). A dataset for multi-faceted analysis of electric vehicle charging transactions. *Scientific Data*, 11, 262.

[38] Zhan, W., Liao, Y., Deng, J., Wang, Z., Yeh, S. (2025). Large-scale empirical study of electric vehicle usage patterns and charging infrastructure needs. *npj Sustainable Mobility and Transport*, 2, 9.

[39] Alam, M.R., Guo, Z. (2023). Charging infrastructure planning for ride-sourcing electric vehicles considering drivers’ value of time. *Transportation Letters*, 15(6), 573–583.

[40] Kemper, N., Heider, M., Pietruschka, D., Hähner, J. (2025). Multi-objective and neuroevolutionary reinforcement learning for electric vehicle charging and load management. *Applied Energy*, 391, 125890.

[41] Dorsey, E., et al. (2025). Optimizing EV charging infrastructure deployment considering value-of-time. *Working Paper*.

[42] Lv, K., Pei, X., Chen, C., Xu, J. (2022). A safe and efficient lane change decision-making strategy of autonomous driving based on deep reinforcement learning. *Mathematics*, 10(9), 1551.

[43] Dorokhova, M., Ballif, C., Wyrsch, N. (2021). Routing of electric vehicles with intermediary charging stations: a reinforcement learning approach. *Frontiers in Big Data*, 4, 586481.

[44] Hu, Z., Wan, K., Gao, X., Zhai, Y. (2019). A dynamic adjusting reward function method for deep reinforcement learning with adjustable parameters. *Mathematical Problems in Engineering*, 2019, 7619483.

[45] Liu, S., Wang, Y., Chen, X., Fu, Y., Di, X. (2022). SMART-eFlo: An integrated SUMO-Gym framework for multi-agent reinforcement learning in electric fleet management. In *Proc. 25th IEEE Int. Conf. Intelligent Transportation Systems (ITSC)*, 2884–2890. IEEE.