Pre-Draft Notes:

**🎯 Summary of Target**

* **Intro** → 200–250 words
* **Infrastructure + Curves/Tariffs/User** → 450–550 words combined (currently ~1450, so major trimming)
* **Traditional Approaches** → 200–250 words
* **Reinforcement Learning** → 350–400 words (heaviest section)
* **Frameworks (PPO/Gym/SUMO)** → 200–250 words
* **Multi-Objective Optimization** → 200–250 words

This gets us to **~1500 words** total, keeps balance, and gives emphasis to **RL + PPO + hybrid optimization**, which is your core novelty.

Literature Review Draft 2

The rapid adoption of electric vehicles (EVs) is transforming the transport and energy sectors, driven by ambitious decarbonisation targets and air quality policies. Cities such as London are expanding charging networks to meet rising demand, yet urban charging remains characterised by fragmented operators, uneven charger distribution, and complex tariffs that combine per-kWh fees, session costs, idle penalties, and membership discounts {need a paper referencing London operator tariffs}. For users, the immediate concerns are minimising charging costs, reducing delays, and ensuring route feasibility, while operators and regulators focus on infrastructure adequacy and grid stability. Traditional optimisation approaches have addressed charging primarily from a grid perspective, using techniques such as adaptive robust optimisation [3] and reactive power management [4] to stabilise distribution networks. Although effective for system reliability, these studies do not reflect user-side constraints, including travel times, heterogeneous tariffs, and individual charging preferences [1, 2].

This gap has motivated the exploration of methods that jointly consider mobility and cost in dynamic settings. Reinforcement learning (RL) has shown promise in such contexts, offering adaptability to uncertainty and the ability to learn policies through interaction with complex environments [6, 7]. Deep Q-Networks (DQN) have been applied to charging scheduling [6], but their discrete action space limits applicability. Proximal Policy Optimisation (PPO), by contrast, is well-suited to continuous decisions and hybrid objectives [8]. This study builds on that potential, applying PPO to optimise cost-based, time-based, and hybrid charging strategies in Inner London, explicitly incorporating real-world tariffs, nonlinear charging curves, and operator heterogeneity.

The effectiveness of EV adoption depends on the accessibility, affordability, and spatial distribution of charging infrastructure. Studies frame the EV charging problem across several dimensions: siting and sizing of chargers, day-to-day operational management, user behaviour, and market or tariff structures [10]. In the UK, infrastructure deployment has been uneven, with studies showing significant “infrastructure gaps” between adoption projections and available charging capacity [11]. These gaps are pronounced in metropolitan areas such as London, where high demand clustering and limited space for station expansion exacerbate congestion {need a paper specifically addressing Inner London infrastructure constraints}.

User behaviour introduces further uncertainty. Empirical analyses of UK charging sessions reveal clustering in evening hours, reflecting commuting patterns and household dependence [12]. Household charging is influenced by routine daily schedules and time-of-use tariffs, while public charging behaviour reflects range anxiety and limited access to off-street charging [13]. These behavioural variations are crucial for optimisation models, as neglecting them risks underestimating congestion and overestimating user flexibility. Full data-chain approaches that integrate transport and grid datasets highlight the importance of modelling heterogeneity in charging demand {need a paper explicitly combining mobility and charging demand in UK}.

Tariff structures add additional complexity. Unlike simplified models assuming uniform per-kWh pricing, real operators impose layered cost schemes including session fees, idle penalties, and membership discounts. Some operators also introduce time-of-day variability, which creates incentives to shift charging to cheaper periods {need a London tariff case study}. Game-theoretic models of station competition show how dynamic tariffs affect station choice and congestion [14], but most are stylised market models rather than fine-grained simulations of urban public charging. Mean-field approaches similarly model strategic pricing under grid constraints, but they do not capture the diversity of tariffs that urban drivers face in practice [3].

Technical constraints from vehicle charging curves further complicate optimisation. Battery charging is nonlinear: power is high at lower states of charge but tapers significantly as SoC approaches full capacity, meaning marginal charging times increase disproportionately [1]. Some studies have attempted to incorporate this tapering into scheduling [2], yet many optimisation models still assume constant charging rates, leading to underestimation of charging duration. When combined with tariff heterogeneity, this creates strong cost-time trade-offs: users may save money by charging at slower, cheaper stations but incur longer delays, whereas prioritising time pushes them towards ultra-rapid stations with higher costs.

Taken together, these challenges illustrate why optimisation cannot focus solely on grid stability. Any practical framework must reflect the realities of heterogeneous user behaviour, tariff fragmentation, and nonlinear charging curves. These dynamics are especially acute in London, where multiple private operators compete across dense urban geography, producing highly variable charging experiences {need London-specific study on operator fragmentation}. Reinforcement learning offers a natural way forward, as it can learn adaptive policies in response to these interacting uncertainties.

Before reinforcement learning gained prominence, EV charging optimisation was primarily addressed through deterministic and heuristic approaches. Classical formulations such as mixed-integer linear programming (MILP) and linear programming (LP) have been widely applied to problems of charging station siting, scheduling, and grid integration. These methods are valued for their transparency and ability to guarantee optimal solutions under fixed assumptions [1, 2]. However, their computational burden scales poorly with the number of vehicles and stations, making them impractical for real-time urban applications.

To address scalability, researchers have employed meta-heuristic techniques such as genetic algorithms (GA), particle swarm optimisation (PSO), and ant colony optimisation (ACO), as well as multi-criteria decision-making (MCDM) frameworks that balance grid stability, accessibility, and cost [3]. While these approaches are more flexible, they lack adaptability to dynamic environments where tariffs, demand, and mobility patterns fluctuate rapidly.

A related strand of research focuses on the electric vehicle routing problem (EVRP), which extends traditional vehicle routing by incorporating battery capacity and charging constraints [4]. Early formulations treated stations as deterministic service nodes, overlooking congestion, waiting times, and nonlinear charging curves. As a result, both deterministic and heuristic methods provide valuable baselines but struggle to capture the stochastic and user-centric nature of urban EV charging. This limitation has motivated the exploration of adaptive, data-driven approaches capable of responding to uncertainty, with reinforcement learning offering a compelling alternative.

EV charging optimisation must account for the combined influence of vehicle dynamics, tariff structures, and user behaviour. A critical technical factor is the nonlinearity of charging curves: batteries accept high power at lower states of charge (SoC) but taper significantly as they approach full capacity. This tapering means marginal charging times increase disproportionately, complicating route and cost forecasting [1]. While some studies incorporate these dynamics [2], many optimisation models still assume constant charging rates, leading to underestimation of session duration.

Tariff heterogeneity adds further complexity. Operators implement diverse pricing schemes including per-kWh rates, session fees, idle penalties, and membership discounts. Some also adopt time-of-day pricing, incentivising off-peak charging {need London tariff study}. Game-theoretic frameworks have shown how strategic pricing affects congestion and station selection [3, 4], yet most treat tariffs at an abstract market level rather than modelling the fine-grained operator diversity characteristic of London.

User behaviour introduces an equally important dimension. UK demand studies show charging sessions cluster in evening hours, reflecting commuting and household routines [5], while household charging behaviour is shaped by range anxiety and tariff sensitivity [6]. Ignoring behavioural heterogeneity risks overestimating network flexibility and underestimating congestion in dense urban environments.

Together, nonlinear charging curves, tariff fragmentation, and behavioural variability make it difficult to balance cost and time objectives. Users may minimise cost by choosing slower, cheaper stations, but this often increases waiting or travel time. Conversely, minimising time pushes them towards ultra-rapid stations with higher fees. Few existing optimisation frameworks explicitly model this trade-off in real-world multi-operator settings {need paper addressing hybrid trade-offs}.

Reinforcement learning (RL) has emerged as a promising alternative to traditional optimisation methods for EV charging. Unlike deterministic or heuristic approaches, RL allows agents to learn adaptive policies by interacting with dynamic environments, making it particularly suitable for problems shaped by uncertainty in tariffs, traffic, and charging demand. Applications in energy systems demonstrate RL’s capacity to reduce costs, flatten grid loads, and improve user satisfaction compared to classical methods [1, 2].

Early EV charging studies applied value-based methods such as Deep Q-Networks (DQN), which approximate action values in high-dimensional spaces. DQN and its extensions (Double and Dueling DQN) have shown success in cost reduction and scheduling efficiency [3, 4]. However, DQN’s reliance on discrete action spaces limits its applicability to charging route optimisation, where decisions often involve continuous variables such as charging duration and delivered energy.

Policy gradient methods address these limitations by optimising policies directly for continuous action spaces. Among them, Proximal Policy Optimisation (PPO) has gained traction for its stability, sample efficiency, and suitability for multi-objective problems [5]. PPO has been applied in energy management and EV fleet charging, where it outperforms value-based methods in scalability and convergence [6]. Actor–critic extensions such as Deep Deterministic Policy Gradient (DDPG) and Soft Actor–Critic (SAC) have also been studied, with SAC offering insights into integrating objectives such as battery degradation [7]. Despite these advances, relatively few works apply PPO to urban charging at the driver level, where decisions must simultaneously account for tariffs, nonlinear charging curves, and congestion.

Reward function design is central to RL’s success in EV optimisation. Existing studies have targeted cost minimisation through per-kWh and session charges [8], time minimisation through charging duration and waiting penalties [9], or hybrid approaches combining multiple objectives [10]. Yet most frameworks fix reward priorities rather than allowing dynamic balancing, limiting their flexibility in user-driven contexts.

Several gaps remain: current models often focus on grid coordination rather than end-user mobility, real-world operator tariffs are rarely integrated, nonlinear charging curves are often ignored, and applications to dense, multi-operator urban settings are scarce. This motivates the use of PPO in this project to optimise cost, time, and hybrid objectives in London, explicitly incorporating tariffs, charging curves, and user mobility constraints into the RL environment.

The effectiveness of reinforcement learning for EV charging depends not only on algorithm choice but also on the frameworks used for training and simulation. Value-based methods such as Deep Q-Networks (DQN) have been widely applied in charging optimisation, reducing costs and improving scheduling efficiency [1, 2]. However, DQN is constrained by its discrete action space, making it unsuitable for problems that require continuous decisions, such as optimising charging duration or energy delivered**(note: compare with more than just DQN**).

This project instead employs Proximal Policy Optimisation (PPO) as the primary algorithm. PPO stabilises policy gradient training by constraining updates, making it both robust and sample-efficient [3]. Its ability to operate in continuous and hybrid action spaces makes it well suited to EV charging, where agents must balance not only *where* to charge but also *how long* to charge. PPO’s suitability for multi-objective optimisation further supports its application in this study, which targets cost, time, and hybrid objectives in Inner London.

The simulation environment is implemented using OpenAI Gym, which provides a modular and reproducible structure for defining states, actions, and rewards [4]. Gym’s flexibility allows the seamless integration of tariffs, charging curves, and SoC dynamics into the RL pipeline. To capture mobility constraints, the Simulation of Urban Mobility (SUMO) is incorporated, providing spatial and temporal modelling of vehicle movements and congestion [5]. Together, Gym and SUMO create a realistic framework for evaluating PPO’s performance in dense urban charging scenarios.

EV charging optimisation inevitably involves trade-offs between cost and time. Cost-oriented approaches typically minimise user expenditure by exploiting dynamic pricing or demand response programmes, often achieving bill reductions while supporting grid stability [1, 2]. However, such models generally overlook time constraints, which are crucial in urban mobility. Time-focused approaches instead prioritise reducing charging duration, waiting, or travel delays, frequently directing users towards ultra-rapid stations despite higher tariffs [3, 4]. Both strands illustrate the difficulty of optimising a single objective in isolation.

Hybrid frameworks explicitly address this challenge by combining cost and time objectives. Some employ weighted sums of both metrics, while others adopt Pareto-optimal methods to produce trade-off solutions [5]. Reinforcement learning is particularly well suited to this setting, as policy-gradient algorithms such as PPO can adapt policies in continuous action spaces and balance competing rewards [6]. Hybrid RL models have been shown to minimise not only charging cost and time but also secondary objectives such as battery degradation [7].

Despite these advances, hybrid optimisation remains underexplored in real-world urban contexts where tariffs are operator-specific, charging curves nonlinear, and congestion highly variable. This gap is especially pronounced in London, where multiple operators impose diverse cost structures and user behaviour creates clustered demand {need a paper explicitly addressing hybrid optimisation in urban EV networks}. Addressing this gap, the present study applies PPO to design cost-based, time-based, and hybrid charging strategies in Inner London, integrating tariffs, nonlinear charging curves, and traffic constraints into a unified optimisation framework.

📌 References for first block:  
[1] Bian et al. (2019). Optimising EV charging: a customer’s perspective. *IEEE Trans. Smart Grid*.  
[2] Xydas, E. (2016). A data-driven approach for characterising EV charging demand: a UK case study. *Appl. Energy*.  
[3] Kiani, H. et al. (2021). Adaptive robust operation of active distribution networks with flexible sources. *Sustainable Energy, Grids and Networks*.  
[4] Weckx, S. et al. (2014). Reactive power support in distribution grids with distributed generation. *IEEE Trans. Sustainable Energy*.  
[6] Liu, L. et al. (2024). RL-based optimisation for EV charging scheduling. *POWERCON*.  
[7] Bertolini, A. (2022). Power output optimisation of EV smart charging hubs using DRL. *Expert Syst. Appl.*  
[8] Schulman, J. et al. (2017). Proximal policy optimisation algorithms. *arXiv preprint arXiv:1707.06347*.

📌 References for second block:  
[1] Yalçın, S. & Herdem, M. S. (2024). Optimising EV battery management with hybrid RL models. *Energies*.  
[2] Bertolini, A. (2022). Power output optimisation of EV smart charging hubs using DRL. *Expert Syst. Appl.*  
[3] Lin, R. et al. (2024). Charging management and pricing strategy of EV stations using mean-field game theory. *Asian J. Control*.  
[10] EV Literature Review doc – Section 1 synthesis.  
[11] Nicholas, M. & Lutsey, N. (2020). Quantifying the EV charging infrastructure gap in the UK. *ICCT Report*.  
[12] Xydas, E. (2016). A data-driven approach for characterising EV charging demand: a UK case study. *Appl. Energy*.  
[13] Almaghrebi, A. et al. (2024). Insights into household EV charging behaviour. *Energies*.  
[14] Gerding, E. H. (2021). A comprehensive game-theoretic model for EV charging station competition. *IEEE Trans. Intell. Transp. Syst.*

📌 References for block three:  
[1] Schmidt, M. et al. (2025). Multiple-criteria-based EV charging infrastructure design. *Energies*.  
[2] (Author truncated) (2024). Optimal siting and sizing of EV charging stations using stochastic power flow analysis. *IEEE Trans. Transp. Electrification*.  
[3] Nicholas, M. (2012). Estimating EV charging infrastructure costs across US metropolitan areas. *Transp. Res. Rec.*  
[4] Anagnostopoulou, A. (2014). EV routing problem with industry constraints. *Transp. Res. Procedia*.

📌 References for block four:  
[1] Yalçın, S. & Herdem, M. S. (2024). Optimising EV battery management with hybrid RL models. *Energies*.  
[2] Bertolini, A. (2022). Power output optimisation of EV smart charging hubs using DRL. *Expert Syst. Appl.*  
[3] Lin, R. et al. (2024). Charging management and pricing strategy using mean-field game theory. *Asian J. Control*.  
[4] Gerding, E. H. (2021). A game-theoretic model of EV charging station competition. *IEEE Trans. Intell. Transp. Syst.*  
[5] Xydas, E. (2016). Characterising EV charging demand: a UK case study. *Appl. Energy*.  
[6] Almaghrebi, A. et al. (2024). Household EV charging behaviour: analysis and predictive modelling. *Energies*.

📌 References for block 5:  
[1] Gautam, M. (2023). Deep reinforcement learning for resilient power and energy systems. *Electricity*.  
[2] Bian, J. et al. (2019). Optimising EV charging: a customer’s perspective. *IEEE Trans. Smart Grid*.  
[3] Liu, L. et al. (2024). RL-based optimisation for EV charging scheduling. *POWERCON*.  
[4] Li, S. et al. (2022). EV charging management using DRL. *Sensors*.  
[5] Schulman, J. et al. (2017). Proximal policy optimisation algorithms. *arXiv preprint arXiv:1707.06347*.  
[6] {need a paper applying PPO in EV fleet charging or energy scheduling}.  
[7] Yalçın, S. & Herdem, M. S. (2024). Hybrid RL models for EV battery management. *Energies*.  
[8] Lin, R. et al. (2024). Charging management and pricing with mean-field game theory. *Asian J. Control*.  
[9] Gerding, E. H. (2021). EV charging station competition: a game-theoretic model. *IEEE Trans. Intell. Transp. Syst.*  
[10] Kiani, H. et al. (2021). Adaptive robust operation of active distribution networks. *Sustainable Energy, Grids and Networks*.

📌 References for block 6:  
[1] Li, S. et al. (2022). EV charging management based on DRL. *Sensors*.  
[2] Liu, L. et al. (2024). RL-based optimisation for EV charging scheduling. *POWERCON*.  
[3] Schulman, J. et al. (2017). Proximal policy optimisation algorithms. *arXiv preprint arXiv:1707.06347*.  
[4] Brockman, G. et al. (2016). OpenAI Gym. *arXiv preprint arXiv:1606.01540*.  
[5] Behrisch, M. et al. (2011). SUMO – simulation of urban mobility: an overview. *SIMUL 2011*.

📌 References for final block:  
[1] Kiani, H. et al. (2021). Adaptive robust operation of active distribution networks. *Sustainable Energy, Grids and Networks*.  
[2] Bian, J. et al. (2019). Optimising EV charging: a customer’s perspective. *IEEE Trans. Smart Grid*.  
[3] Bertolini, A. (2022). Power output optimisation of EV smart charging hubs using DRL. *Expert Syst. Appl.*  
[4] Anagnostopoulou, A. (2014). EV routing problem with industry constraints. *Transp. Res. Procedia*.  
[5] Nicholas, M. & Lutsey, N. (2020). Quantifying the EV charging infrastructure gap in the UK. *ICCT Report*.  
[6] Schulman, J. et al. (2017). Proximal policy optimisation algorithms. *arXiv preprint arXiv:1707.06347*.  
[7] Yalçın, S. & Herdem, M. S. (2024). Hybrid RL models for EV battery management. *Energies*.

NEXT STEPS:

1 from first to last references correct.

2. add references where {needed}.

3. make your own references table.

4. when all 3 above are done, put it in chatgpt ,claude, gemini and ask it to fact check it.

5. revise if needed.

6. send it to elisa.

7. make new stuff if you want.

It feels like we are trying to compare dqn and ppo in lit review.

We have to make it sound like we prefer ppo over dqn, blah, blah , blah.

EXACTLY. We have to show why ppo works in our case better by listing dqn, blah etc ke where it works best, and ppo kyu yaha better karti hai.