الجمهورية الجزائرية الديمقراطية الشعبية

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

وزارة التعليم العالي و البحث العلمي

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

جامعة عمار ثليجي بالأغواط

UNIVERSITY OF AMMAR TELIDJI LAGHOUAT

كلية العلوم

FACULTY OF SCIENCES

قسم الإعلام الآلي

DEPARTMENT OF COMPUTER SCIENCE

MASTER'S THESIS

Field : Mathematics and Computer Science

Option : Computer Science

Specialization : Networks, Distributed Systems and Applications

Submitted by : Kouidri Imane

Theme

Deployment and Optimization of Electric Taxi Vehicles in Urban Areas

Publicly defended on 07/07/2025, before the jury composed of:

Mr.	Tahar Allaoui	MC(B)	(University of Laghouat, Algeria)	President
Mr.	Abdelmajid Ben Arffa	MC(B)	(University of Laghouat, Algeria)	Examiner
Mr.	Lakhdar Oulad Djedid	MC(A)	(University of Laghouat, Algeria)	Examiner
Mr.	Tahar Bendouma	MC(A)	(University of Laghouat, Algeria)	Supervisor
Mr.	Omar Sami Oubbati	MC(A)	(University of Gustave Eiffel, France.)	Co-Supervisor

Academic Year 2025/2026

الجمهورية الجزائرية الديمقراطية الشعبية
RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
وزارة التعليم العالي و البحث العلمي
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE
جامعة عمار ثليجي بالأغواط
UNIVERSITÉ AMMAR TELIDJI LAGHOUAT

كلية العلوم
FACULTÉ DES SCIENCES
قسم الإعلام الآلي
DÉPARTEMENT D'INFORMATIQUE

THÈSE DE MASTER

Domaine : Mathématique et Informatique

Filière : Informatique

Option : Réseaux, Systèmes et Applications Répartis

Présenté par : Kouidri Imane

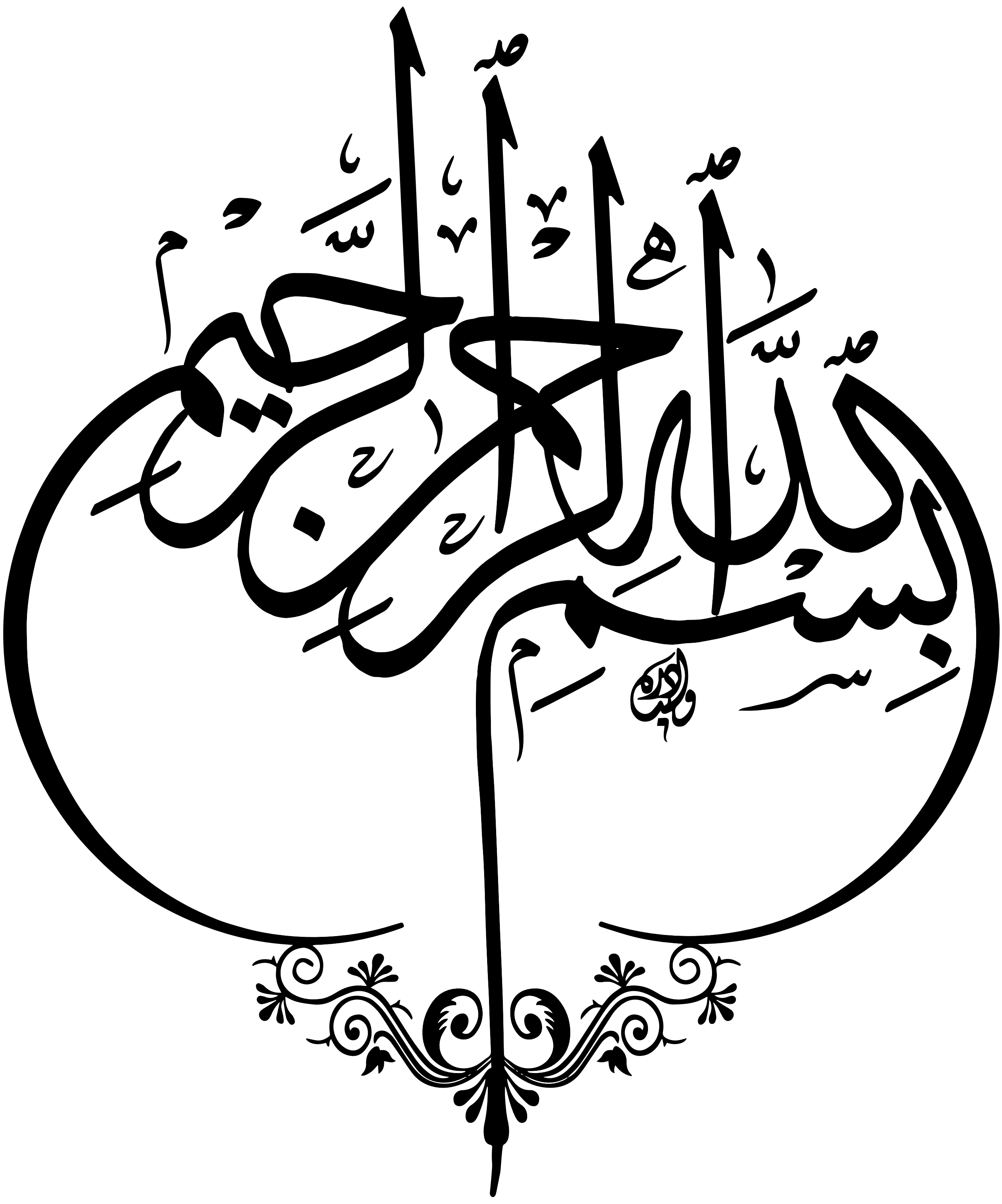
Thème

Déploiement et optimisation des taxis électriques dans les zones urbaines

Soutenue publiquement, le 07/07/2025, devant le jury composé de :

Mr.	Tahar Allaoui	MC(B)	(Université de Laghouat, Algérie.)	President
Mr.	Abdelmajid Ben Arffa	MC(B)	(Université de Laghouat, Algérie.)	Examinateur
Mr.	Lakhdar Oulad Djedid	MC(A)	(Université de Laghouat, Algérie.)	Examinateur
Mr.	Tahar Bendouma	MC(A)	(Université de Laghouat, Algérie.)	Encadreur
Mr.	Omar Sami Oubbati	MC(A)	(Université de Gustave Eiffel, France.)	Co-Encadreur

Année Universitaire 2025/2026


بسم الله الرحمن الرحيم

Dedications

In the name of Allah, the Most Gracious, the Most Merciful.

I dedicate this work, first and foremost, to Allah, whose infinite mercy, strength, and guidance granted me the will, perseverance, and health to complete this journey.

To my beloved parents, whose unconditional love, prayers, and sacrifices are the true foundation of my achievements.

To my dear sisters and brothers, for their constant encouragement and unwavering support.

To my precious nephews, whose joy and presence inspired me and filled my heart with love and motivation.

To my cousin Karim, whose thoughtful encouragement and shared passion for this field offered me both motivation and perspective whenever our paths crossed.

And finally, to myself for the resilience, dedication, and countless hours of effort that turned this goal into reality.

Acknowledgments

I would like to express my deepest gratitude to my supervisors Dr. Tahar Bendouma and Dr. Omar Sami Oubbati, whose exceptional guidance, patience, and support were instrumental throughout this journey. Their belief in my potential, professionalism, and rich knowledge not only shaped this work but also helped me grow as a learner and researcher. I am truly thankful for all I have learned under their mentorship.

I extend my sincere appreciation to my university and to all the professors and doctors within my department. Their dedication to teaching, and the knowledge and values they imparted, have greatly enriched my academic path and inspired me over the years.

I am also deeply thankful to my family—my parents, siblings, nephews, and cousin—for their love, encouragement, and support, which have been my strength during this journey.

A special mention goes to my former high school French teacher, whose warmth and kind words of pride remain a source of inspiration to me.

To everyone who offered guidance, support, or kindness—no matter how small—thank you from the bottom of my heart.

Abstract

Electric vehicles (EVs) are at the forefront of the transition toward sustainable urban transportation. Among them, electric taxis have emerged as an eco-friendly alternative to traditional internal combustion engine vehicles, offering reduced emissions and operational noise. However, integrating electric taxis into urban mobility systems introduces new challenges—ranging from limited battery capacities to inadequate charging infrastructure and complex service optimization needs. This thesis addresses these challenges through the development of an intelligent decision-making framework based on Deep Q-Network (DQN) reinforcement learning. We investigate methods to optimize taxi dispatching, customer service, and charging behavior while minimizing downtime and maximizing system efficiency. The proposed system architecture integrates a realistic simulation environment with performance evaluation metrics tailored to electric taxi services. A comparative analysis with existing optimization methods highlights the strengths and limitations of various approaches. Experimental results demonstrate the efficacy of our proposed solution in improving energy efficiency, customer service rate, and operational sustainability. This work offers practical insights for deploying intelligent EV taxi systems and contributes to the broader goal of sustainable smart city development.

Keywords: Electric Vehicles (EVs), Electric Taxis, Reinforcement Learning, Deep Q-Networks (DQN), Charging Optimization, Urban Mobility.

ملخص

تُعد المركبات الكهربائية، وخاصة سيارات الأجرة الكهربائية، جزءًا أساسيًا من مستقبل التنقل المستدام في المدن. يتناول هذا العمل التحديات المرتبطة بإدماج سيارات الأجرة الكهربائية في البيئة الحضرية، مثل محدودية سعة البطارية، وجدولة الشحن المثلى، والحاجة إلى خدمة أكبر عدد ممكن من الركاب بكفاءة. نقترح نهجًا يعتمد على التعلم المعزز باستخدام شبكة Q العميقة (DQN) لتحسين تشغيل سيارات الأجرة الكهربائية. يتعلم النموذج المقترح متى يجب تقديم الخدمة أو التوجه للشحن، بهدف تقليل وقت التوقف وهدر الطاقة وزيادة كفاءة الخدمة. أظهرت التجارب والمحاكاة فعالية النهج مقارنةً بالطرق التقليدية، مع تحسينات ملحوظة في إدارة الطاقة وتقديم الخدمات والأداء التشغيلي.

الكلمات المفتاحية: المركبات الكهربائية، سيارات الأجرة الكهربائية، التعلم المعزز، شبكة العميقة، تحسين الشحن، التنقل الحضري.

Résumé

Les véhicules électriques (VE), en particulier les taxis électriques, jouent un rôle crucial dans la mobilité urbaine durable. Ce mémoire traite des défis liés à l'intégration des taxis électriques dans les villes, notamment la capacité limitée des batteries, la planification optimale de la recharge, et la nécessité de servir un maximum de clients de manière efficace. Nous proposons une approche d'apprentissage par renforcement basée sur le réseau de neurones profond Q (DQN) pour optimiser les opérations des taxis électriques. Le modèle proposé apprend à décider entre servir les passagers ou se recharger, en visant à réduire le temps d'inactivité et la consommation inutile d'énergie tout en maximisant l'efficacité du service. Des simulations approfondies démontrent l'efficacité de notre approche par rapport aux stratégies classiques, avec des améliorations significatives en gestion d'énergie, service client et performance opérationnelle.

Mots-clés : Véhicules électriques, Taxis électriques, Apprentissage par renforcement, DQN, Optimisation de la recharge, Mobilité urbaine.

Table of Contents

List of Figures

List of Tables

List of Algorithms

List of Acronyms

AC Alternating Current.

BEV Battery Electric Vehicle.

BMS Battery Management System.

CCS Combined Charging System.

CHAdeMO CHArge de MOve.

DC Direct Current.

DQN Deep Q-Networks.

EV Electric Vehicle.

FCEV Fuel Cell Electric Vehicle.

HEV Hybrid Electric Vehicle.

HHV Hybrid Hydraulic Vehicle.

ICE Internal Combustion Engines.

ICEV Internal Combustion Engine Vehicle.

IEC 61851 IEC Standard 61851.

IEC 62196 IEC Standard 62196.

ISO 15118 ISO Standard 15118.

MSE Mean Squared Error.

PHEV Plug-in Hybrid Electric Vehicle.

PHV Pneumatic Hybrid Vehicle.

RL Reinforcement Learning.

SAE J1772 SAE Standard J1772.

V2G Vehicle-to-Grid.

Chapter 1

Introduction

1.1 Context

In recent years, EVs have become a pivotal solution in the global effort to reduce greenhouse gas emissions and transition toward sustainable transportation systems. Among these, electric taxis stand out as a viable alternative to traditional diesel or gasoline-powered cabs, especially in densely populated urban areas. These vehicles offer numerous environmental and operational advantages, such as zero tailpipe emissions, lower maintenance costs, and quieter operation. However, the widespread adoption of electric taxis faces several technical and logistical challenges that must be addressed to ensure their effectiveness in urban settings.

1.2 Problem Statement and Objective

Despite their benefits, electric taxis suffer from limited driving range, long charging times, and insufficient infrastructure support. These constraints directly affect their ability to serve customers efficiently and continuously throughout the day. Therefore, this thesis aims to explore the optimization of electric taxi operations from multiple perspectives:

- Efficient deployment and use of charging stations.
- Intelligent decision making regarding when and where to charge.
- Maximizing the number of customers served while managing limited energy reserves.
- Designing a smart, responsive system architecture that enables real-time decision making.

To address these issues, we develop a DQN-based reinforcement learning model that allows electric taxis to make autonomous, data-driven decisions about serving passengers and charging. This model is validated through extensive simulations and benchmarked against traditional optimization approaches.

1.3 Organization of the Thesis

This thesis is organized as follows:

- **Chapter 2** provides a comprehensive background on electric vehicle (EV) technologies, charging infrastructure, and electric taxi systems. It also surveys related work and presents a comparative analysis of existing optimization approaches applied in electric taxi operations..
- **Chapter 3** details the proposed system architecture, including the design of the DQN learning model and Q-learning, state-action space, and reward functions.
- **Chapter 4** presents the simulation setup, training results, testing outcomes, and a comparative performance evaluation.
- **Chapter 5** concludes the thesis with a summary of contributions and outlines potential future directions.

Chapter 2

Background and Related Works

2.1 Introduction

This chapter presents the necessary theoretical and contextual background on EVs, their charging infrastructure, and the specific characteristics of electric taxi systems. It also surveys the relevant literature on optimization approaches applied to EV taxi operations.

The first part of the chapter provides an overview of EV technologies, including their definition, technical components, and environmental implications. It further explores the electric vehicle charging ecosystem, discussing charging levels, connector types, smart charging solutions such as Vehicle-to-Grid (V2G), and the main challenges and trends in infrastructure deployment. This background establishes the foundational understanding required to address the operational characteristics and constraints of EV-based mobility services.

The second part of the chapter is devoted to a review of existing research in electric taxi optimization. The literature is categorized into three main strands: (i) reinforcement learning approaches for EV taxi dispatch and fleet coordination, (ii) heuristic and game-theoretic scheduling strategies, and (iii) metaheuristic algorithms focused on charging and routing optimization. Each category is analyzed regarding its methodological contributions, strengths, and relevance to energy-aware service management.

Building upon the insights from this literature, this thesis investigates the use of reinforcement learning—specifically Q-learning and DQN to improve decision-making in EV taxi systems. The detailed methodology and experimental setup are presented in the next Chapter 3.

2.2 Background on Electric Vehicles

2.2.1 Electric Vehicles

2.2.1.1 What is an Electric Vehicle (EV)?

EVs are transportation systems powered entirely or partially by electricity, using electric motors and rechargeable batteries instead of Internal Combustion Enginess (ICEs). Unlike traditional vehicles fueled by gasoline or diesel, EVs draw energy from the electrical grid, making them a cleaner and more energy-efficient alternative to conventional transportation modes [3].

2.2.1.2 History and Evolution of Electric Vehicles

Although often perceived as a recent innovation, EVs have existed since the late 19th century. Their development can be categorized into five distinct phases as illustrated the following Table 2.1.:

Table 2.1: Condensed Historical Evolution of Electric Vehicles (EVs)

Era	Period	Key Developments
Emergence of EVs	Late 1800s – Early 1900s	EVs gained popularity in urban areas for their quiet and clean operation. By 1900, they accounted for about one-third of U.S. vehicles. Edison contributed to early battery efforts [3].
ICE Dominance	Early 1900s – 1920s	The rise of gasoline cars, especially Ford's Model T, combined with battery limitations and fueling infrastructure growth, led to EVs' decline [3].
Stagnation Period	1930s – 1980s	EV progress stalled; limited use in niche markets. Oil crises in the 1970s revived interest, but policy and technology constraints hindered growth [3].
Technological Rebirth	1990s – 2000s	Environmental policies and battery innovation (e.g., lithium-ion) reignited R&D. Notable models included GM's EV1 under California's ZEV mandate [3].
Modern EV Revolution	2010s – Present	Tesla's success demonstrated EV viability. Widespread adoption followed with vehicles like the Nissan Leaf and Model S, supported by expanded charging infrastructure and smart technologies [3].

2.2.1.3 Types of an Electric Vehicle (EV)

EV can be broadly categorized into:

- **Battery Electric Vehicles (BEVs):**

  EVs are fully electric systems that offer significant environmental and operational advantages over Internal Combustion Engine Vehicles (ICEVs). They produce up to 50 % fewer CO_2 emissions even when powered by fossil fuels and are approximately three times more energy efficient [4]. EVs also reduce maintenance costs due to their simplified mechanical design and absence of traditional engine components [5].

However, current battery technology presents limitations, including long charging times, limited driving range, and high cost and weight [4].

- **Fuel Cell Electric Vehicles (FCEVs):**

  FCEVs address key limitations of EVs, offering faster refueling, longer driving range, and lower maintenance due to their continuous electricity generation from hydrogen fuel cells [4]. Unlike batteries, fuel cells produce energy on demand through electrochemical reactions, with hydrogen as the primary input. Various types exist, including PEMFCs, DMFCs, and SOFCs, each suited to different applications and conditions [4].

- **Hybrid Electric Vehicles (HEVs):**

  HEVs offer a practical interim solution to zero-emission transport, combining ICEs with electric motors to improve fuel efficiency and reduce emissions [4]. They feature regenerative braking and operate without reliance on grid electricity. HEVs reduce fuel use, noise, and operational costs, making them widely adopted in both private and commercial sectors [4].

- **Plug-in Hybrid Electric Vehicles (PHEVs):**

  PHEVs combine electric propulsion with an ICE, allowing short-range electric driving (typically 20–80 km) using grid-supplied power, with the ICE as backup [4]. They reduce reliance on fuel and charging infrastructure and offer lower production costs than EVs [4]. Emissions depend on electricity sources but are generally lower than ICEVs [4, 6].

- **Hybrid Hydraulic Vehicles (HHVs):**

  HHVs combine an ICE with a hydraulic motor and accumulator system for propulsion and regenerative braking [4, 6]. They outperform HEVs in regenerative braking efficiency—recovering over 70 % of kinetic energy—making them ideal for urban stop-and-go traffic [4]. HHVs are simpler, cost-effective, and efficient but face adoption barriers due to bulky components, noise, and limited support for electrical accessories when the ICE is off [4].

- **Pneumatic Hybrid Vehicles (PHVs):**

  PHVs combine fuel and compressed air to reduce emissions and fuel consumption, using the ICE as both power source and air compressor [4, 7]. They recover braking energy as compressed air, which powers accessories and assists propulsion under low-load or high-torque conditions. Unlike other hybrids, PHVs share the same ICEs for both combustion and air-hybrid functions, reducing system complexity [4].

However, challenges include limited energy storage due to air pressure constraints and the need for advanced variable valve systems [4, 7].

2.2.1.4 The Key Components of an Electric Vehicle (EV)

The core of an EV is its electric drivetrain, which replaces the traditional ICE. Key components include:

- **Electric Motor:**

  Electric motors in EVs convert electrical energy into mechanical propulsion and are mainly categorized as Direct Current (DC) or Alternating Current (AC) types [3].

  - **DC motors:** offer strong torque-speed performance and simple speed control, making them suitable for high-traction, cost-sensitive EV applications. However, their brush-based design increases maintenance needs and limits efficiency [4].
  - **AC motors:** in contrast, are more efficient, lightweight, and durable due to their brushless operation. Despite requiring complex and costly inverters, they dominate modern EV applications. Common types include Induction Motors (IM), ideal for large vehicles, and Permanent Magnet Synchronous Motors (PMSM), used in compact EVs like the Toyota Prius and Honda Insight [4].

- **Batteries:**

  Batteries serve as the dominant energy storage medium in BEVs due to their high energy density, technological maturity, and decreasing costs [4]. However, challenges remain in terms of power density, cycle life, cost, and charging time [8, 9]. The key advantages and limitations can be found in Table 2.2,

Table 2.2: Classification of EV Battery Technologies [2].

Battery Type	**Main Advantages**	**Main Disadvantages**
Lead-Acid Traditional, low-cost design [2]	Inexpensive, widely available High starting current (good for ICE)	Toxic and heavy Low energy density Short cycle life (600 cycles)
Nickel-Cadmium (Ni-Cd) [2]	Higher energy density ($80\,\mathrm{W\,h\,kg^{-1}}$) Affordable	Memory effect Cadmium toxicity, difficult to recycle

Battery Type	Main Advantages	Main Disadvantages
Lithium-Ion (Li-ion) [2,4]	High energy density ($150\,\mathrm{W\,h\,kg^{-1}}$ to $200\,\mathrm{W\,h\,kg^{-1}}$) Long cycle life (1000 cycles) Low self-discharge (<10%/year) No memory effect	Sensitive to deep discharges Risk of explosion without Battery Management System (BMS)
Lithium Iron Phosphate ($LiFePO_4$) [2]	High thermal stability and safety Non-toxic, cobalt-free Cost-effective	Lower energy density Requires more cells in series
Lithium-Polymer (Li-Po) [2]	Flexible form factor Lightweight (no metal casing) Improved safety vs. Li-ion	Less energy efficient Higher production cost Sensitive to charging conditions
Lithium-Metal-Polymer (LMP) [2]	Solid-state: explosion-resistant Environmentally friendly Very low self-discharge	Requires high operating temperature (85 °C) Limited practical usage data

- **Power Electronic Converters:**

  Power electronic converters are essential in BEVs for managing energy flow, converting voltages, and enabling bidirectional power transfer between the high-voltage battery, traction motor, and auxiliary systems [4]. They replace the alternator in ICEVs and ensure efficient operation across various voltage levels (12 V–120 V) [4]. DC–DC converters supply low-voltage loads and support regenerative braking through energy recovery [4]. DC–AC inverters power the traction motor, while AC–DC rectifiers assist with battery charging [4]. These converters must be compact, lightweight, and highly efficient [10, 11].

- **Controller:**

  The controller is the central unit in a BEV that regulates motor speed, torque, and power flow based on driver input, ensuring optimal performance and energy efficiency [3]. It interprets signals from the accelerator pedal via potentiometers and adjusts power delivery to the motor accordingly [12].

- **Power Bus:**

BEVs use DC power buses to manage energy flow between components through a centralized architecture [4]. Two main buses are employed: a high-voltage bus for traction systems and a low-voltage bus for auxiliary devices [4]. DC–DC converters enable bi-directional energy transfer between these buses, improving efficiency and supporting system safety by isolating voltage domains [4].

- **Regenerative Braking System:**

  In BEVs, regenerative braking enhances energy efficiency by converting kinetic energy into electrical energy during deceleration via the traction motor acting as a generator [4]. It works in tandem with hydraulic brakes to ensure safety and optimize energy recovery, particularly under varying speed and load conditions [4].

- **Charging Port:**

  The charging port enables external electrical energy replenishment of the vehicle's battery system [3].

- **Battery Management System (BMS):**

  The BMS ensures safe, efficient, and reliable battery operation by monitoring parameters, managing thermal conditions, balancing cells, and estimating states such as SoC, SoH, SoP, and SoE [13–15]. It plays a vital role in protecting the battery and optimizing performance.

- **Transmission (Reduction Gear):**

  The simplified transmission in BEVs transfers torque from the electric motor to the wheels, minimizing complexity and maximizing efficiency compared to traditional ICEVs [3].

2.2.1.5 How does an Electric Vehicle (EV) Work?

When the accelerator pedal is pressed in an EV, it closes an electrical circuit that allows current to flow through the motor windings, generating a rotating magnetic field. This field interacts with magnetic elements in the rotor, producing torque and causing the rotor to spin. In AC motors, the use of alternating current is essential, as it continuously switches the polarity of the magnetic fields, maintaining rotation. Most modern EVs use AC induction motors or Permanent Magnet Synchronous Motors (PMSMs), both of which offer efficient and reliable electric-to-mechanical energy conversion through electromagnetic induction and interaction [16].

2.2.1.6 Environmental Benefits and Drawbacks

Electric vehicles (EVs) offer substantial environmental advantages, particularly in reducing greenhouse gas emissions and urban air pollution [17]. They support sustainability through cleaner operation, enhanced energy efficiency, and alignment with renewable energy integration. Additionally, EVs benefit from lower operating costs and government incentives, reinforcing their appeal [17].

However, their widespread adoption faces challenges such as limited driving range and dependence on charging infrastructure, which may hinder practicality for long-distance users [17]. A balanced evaluation of these factors is essential to inform policy and guide future mobility strategies.

2.2.2 Electric Vehicle Charging Infrastructure

2.2.2.1 Overview of Charging Stations

Charging stations are a cornerstone of sustainable electric mobility, serving as the essential interface for energy replenishment in (BEVs) [2]. Their classification into Level 1, Level 2, and Level 3 (DC fast charging) reflects their role in meeting diverse user needs—ranging from slow, cost-effective home charging to high-speed public and highway-based solutions.

Beyond power delivery, modern charging infrastructure integrates smart features such as dynamic load control and grid communication, which are vital for optimizing energy usage and maintaining grid stability. As EV adoption accelerates, the strategic expansion of intelligent and accessible charging networks remains critical to ensuring operational efficiency and user convenience [2].

2.2.2.2 EV Charging Infrastructure: Importance, Classification, and Standardization

EV chargers are fundamental to enabling the regular replenishment of energy in high-voltage batteries (HVBs), facilitating the transition to sustainable transportation systems [18, 19]. Chargers operate by converting energy from the grid through multiple power stages and can be classified by architecture (on-board vs. off-board), directionality (unidirectional vs. bidirectional), integration (dedicated vs. traction-integrated), and connection method (conductive vs. wireless) [2, 20–23].

In terms of performance, EV charging systems are typically categorized into three levels:

- **Level 1 (120V AC):** Suitable for home use; offers low-cost, slow charging [2].
- **Level 2 (208–240V AC):** Balances speed and cost, prevalent in residential and public settings [24].
- **Level 3 (DC fast charging):** High-speed charging (up to 80 % in 30 min), ideal for commercial and highway use [25, 26].

EV plug types vary regionally, reflecting differing electrical and regulatory standards as illustrated in Figure. 2.1. Common connectors include:

- **Type 1 (SAE Standard J1772 (SAE J1772)):** in North America and Japan for AC charging [27].
- **Type 2 (Mennekes):** in Europe for AC and DC [28].
- **CHArge de MOve (CHAdeMO):** in Japan for DC fast charging [29].
- **Combined Charging System (CCS) (Combo 1/2):** for integrated AC/DC fast charging in the U.S. and Europe [30].

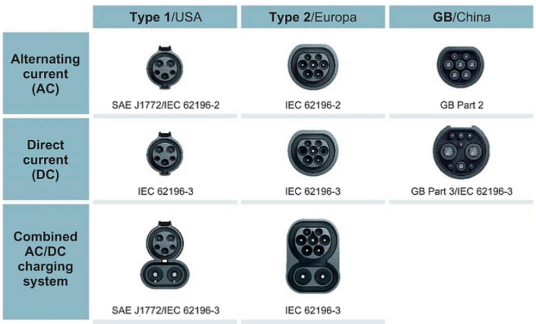


Figure 2.1: Plug types [1].

Globally, charging interoperability is governed by standards such as IEC Standard 62196 (IEC 62196) (hardware connectors), IEC Standard 61851 (IEC 61851) (charging system requirements), and ISO Standard 15118 (ISO 15118) (communication protocols) [31, 32]. These protocols enable smart functions like Plug & Charge, authentication, and V2G) integration [33, 34].

The development of a reliable and standardized charging infrastructure—featuring smart grid communication, dynamic load management, and universal compatibility is vital for large-scale EV deployment and long term grid stability [2].

2.2.2.3 Smart Charging, V2G, and Emerging EV Charging Technologies

The large-scale adoption of EVs, incentivized by government policies (e.g., UK's 2030 ban on ICEVs), raises concerns about grid stress due to uncoordinated charging [35,36]. Smart charging systems address this by shifting charging to off-peak hours, reducing costs, and supporting renewable integration [37,38]. V2G extends this by enabling EVs to discharge energy to the grid, enhancing load balancing and grid stability [39,40].

While these systems improve efficiency, they introduce cybersecurity and privacy risks due to their reliance on digital communication [39].

Infrastructure development varies globally: China leads in public deployment, while Europe and the US combine residential and public strategies. However, rural and low-income areas remain underserved [41,42].

2.2.2.4 Advanced Charging Technologies

To reduce charging times and improve accessibility, several technologies are emerging:

- **Inductive Charging:** Offers wireless, contactless charging but suffers from misalignment and lower efficiency [43–45].
- **Dynamic Charging:** Allows in-motion charging via embedded road coils, still under development [46,47].
- **Ultra-Fast Charging (UFC):** Delivers greater then 350 kW, enabling sub 30 minute charging, though with higher cost and potential battery wear [48,49].
- **DC Fast Charging (DCFC):** Provides 50–350 kW directly to batteries, common in highways [50,51].
- **Tesla Superchargers:** Proprietary fast chargers offering up to 250 kW, efficient but limited to Tesla models [52,53].
- **Bidirectional Charging (V2G):** Enables EVs to supply energy back to the grid, requires regulatory support and specialized hardware [54,55].
- **Battery Swapping:** Enables rapid replacement of depleted batteries, limited by cost and lack of standardization [56,57].

2.2.2.5 Environmental and Economic Impact

EV charging infrastructure reduces GHG emissions, especially when powered by renewables [58,59]. Smart charging can lower energy costs and improve grid efficiency, while V2G enhances energy flexibility [41,60,61].

2.2.3 Electric Taxis: Characteristics and Challenges

2.2.3.1 Introduction and Adoption Trends

Electric taxis (e-taxis) have emerged as a promising solution for decarbonizing urban transport. Their adoption is accelerating globally, supported by policy incentives, environmental regulations, and commitments from ride-hailing platforms to electrify their fleets. The global e-taxi market reached an estimated value of 21.8×10^9 USD in 2023 and is projected to grow at a CAGR of 12.6 % through 2032 [62].

2.2.3.2 Charging Behavior and Infrastructure

Data from large-scale operations in cities such as Shenzhen and Beijing indicate that most e-taxis undergo charging sessions lasting between 1.1 and 1.6 hours, with over 85 % completed within an hour [63]. These trends highlight the importance of deploying fast-charging stations in strategic locations, particularly at depots and high-demand urban zones, to minimize operational disruptions.

2.2.3.3 Economic and Environmental Considerations

While the initial purchase price of e-taxis exceeds that of conventional vehicles by 10 000 USD to 20 000 USD, their long-term economic viability is supported by lower fuel and maintenance costs [64]. Environmental analyses report reductions in urban CO_2 emissions of up to 5 %, particularly when paired with low-carbon electricity sources [65, 66].

2.2.3.4 Operational Challenges

Despite their benefits, e-taxis face challenges including limited range, infrastructure gaps, and charging-related downtime. High utilization rates necessitate efficient recharging strategies. Studies suggest that integrating real-time scheduling and demand-based pricing mechanisms can improve fleet efficiency by nearly 19 % [67].

2.3 Related Work on EV Taxi Optimization

This section categorizes the relevant literature into three main streams: (1) reinforcement learning for electric vehicle dispatching and charging, (2) heuristic and game-theoretic scheduling methods, and (3) metaheuristic optimization for EV route and energy planning. Each category presents a different perspective on solving the operational challenges faced by electric taxi systems, and is discussed below in relation to the proposed DQN-based approach.

2.3.1 Reinforcement Learning for EV Taxi or Fleet Dispatch

Reinforcement Learning (RL) has emerged as a powerful approach for optimizing decision-making in EV taxi systems, particularly under dynamic and uncertain conditions. Several studies have proposed RL-based models to address challenges such as passenger dispatch, route optimization, and energy management. This section reviews notable contributions in this domain and highlights how the proposed work advances the state of the art by applying both q-learning and DQN to train a single agent to intelligently decide between serving passengers and recharging. The system incorporates energy thresholds and scenario-based complexity to enable realistic, energy-aware behavior in electric taxi operations.

2.3.1.1 Rivière & Chung (2021)

Rivière and Chung introduced H-TD2, a hybrid temporal-difference learning framework designed for large-scale taxi dispatch in dynamic urban environments [68]. Their method combines local Q-learning updates with a centralized value function, enabling distributed agents to balance local autonomy with global coordination. The goal is to reduce passenger wait time and improve scalability in fleet-wide deployments.

Unlike their multi-agent and partially centralized setting, the current work focuses on a single electric taxi agent learning when to serve a passenger or charge based on energy availability. Using both Q-learning and DQN, the agent is trained to optimize its actions to minimize unnecessary recharging trips and maximize passenger service. The system models energy thresholds and charging stations but excludes agent coordination or distributed training elements.

2.3.1.2 Oda & Joe-Wong (2018)

Oda and Joe-Wong proposed MOVI, a decentralized DQN-based framework where each taxi independently learns ride assignment strategies [69]. Their approach reduces the number of unserved ride requests in a model-free manner, achieving up to a 76 % improvement over static dispatch baselines. The system assumes independent agents, and learning is performed per agent without centralized information.

In contrast, the present study addresses the decision-making of a single electric taxi agent choosing between charging or serving passengers. While both works utilize DQN, this work introduces energy-awareness, charging station constraints, and scenario-based simulation in both small and large grids. Unlike MOVI, the focus here is not on fleet-wide dispatch efficiency but on teaching a single agent to act intelligently under energy limitations.

2.3.1.3 Cao et al. (2021)

Cao et al. developed a smart charging algorithm for electric vehicles using a customized actor-critic reinforcement learning approach [70]. Their model aims to reduce operational costs and balance grid impact in uncertain demand settings. The framework employs continuous control over charging actions but is not designed for taxi operations or service decision-making.

This thesis adopts a discrete-action learning strategy, where an electric taxi chooses between two high-level actions—serve or charge—based on current energy levels and passenger proximity. Unlike Cao et al., who focus on grid management and cost, the current work emphasizes service efficiency in electric taxi operation. The model uses both Q-learning and DQN to train the agent in two grid scenarios with increasing complexity, without requiring continuous control or grid-level optimization.

2.3.2 Heuristic and Game-Theoretic Scheduling

Beyond reinforcement learning, many researchers have explored traditional optimization methods such as heuristic rules and game theory to address electric taxi coordination and charging challenges. These methods typically assume static or semi-static environments, focusing on predefined decision rules or equilibrium strategies rather than real-time learning. While they offer interpretability and tractability, such approaches are limited in adapting to dynamic passenger demand, variable energy availability, or agent-level interaction.

2.3.2.1 Zhu et al. (2017): Game-Theoretic Joint Transportation and Charging

Zhu et al. propose a game-theoretic model for joint transportation and charging scheduling in public vehicle (PV) systems [71]. The system is modeled as a cake-cutting game among PV groups, where a central cloud platform coordinates the allocation of energy and transportation resources. Each PV group aims to maximize its utility by balancing passenger service and charging costs, considering dynamic electricity prices. Their simulations, conducted on real New York City data, show improved energy cost efficiency without significantly degrading transportation performance.

However, the model assumes rational behavior and full observability, with limited adaptability to real-time uncertainty. In contrast, our DQN-based single-agent system allows an electric taxi to learn adaptive policies through direct interaction with the environment, handling stochastic energy levels, station congestion, and dynamic passenger distributions using a trial-and-error framework.

2.3.2.2 Zhang et al. (2024): Grid-Aware Coordinated Charging and Routing Strategy

Zhang et al. present a coordinated scheduling strategy that addresses the challenge of balancing regional power grid loads while managing EV charging behavior and travel routing [72]. Recognizing the increasing penetration of renewable energy and its contribution to grid instability, the authors model the transportation network and power grid as a cyber-physical system. They propose two strategies: one for individual EVs that aligns charging schedules with travel plans to reduce individual time cost, and another for collective EV flows that balances charging station loads during peak hours. By simulating travel patterns and charging behaviors under varying conditions, the study shows that appropriate joint charging and routing schedules can reduce power load imbalance and enhance grid stability, transforming EVs from being grid burdens into active load-balancing agents. The strategies focus on grid-level efficiency, especially under high EV penetration and renewable integration.

Unlike our DQN-based framework, which focuses on an individual electric taxi navigating a dynamic environment with passenger demand and energy constraints, Zhang et al.'s model emphasizes grid-aware system-wide scheduling under known traffic and energy demand patterns. While their methods improve long-term energy balance across regions, they do not support real-time, decentralized learning or dynamic passenger servicing, which are central to our work.

2.3.2.3 Ma et al. (2025): MILP-Based Coordinated Dispatching with Congestion Awareness

Ma et al. propose a dynamic dispatching and charging scheduling framework for electric ride-hailing fleets, explicitly accounting for charging station congestion and real-time energy pricing [73]. The authors introduce CongestionAware, a scheduling policy that anticipates vehicles' charging needs and optimizes station usage through a two-phase strategy: (1) day-ahead charging plans using a sequential Mixed Integer Linear Programming (MILP) model, and (2) intra-day adjustments responsive to real-time vehicle energy status and queue lengths at stations. Unlike prior models that assume unconstrained charging infrastructure or fixed prices, this approach models charging congestion and dynamic electricity costs, which are both critical for realistic urban deployment. Simulation results on a Manhattan-style grid using NYC yellow taxi data reveal significant improvements in profitability and service rate demonstrating a +15.06% profit and +19.16% higher service coverage compared to benchmark strategies.

While highly effective in structured environments with predictable demand, the approach is largely centralized and rule-driven, requiring prior knowledge of daily demand patterns and station behavior. In contrast, our DQN-based agent learns to adapt in real-time to uncertain passenger arrivals and dynamic station conditions without the need for handcrafted day-ahead plans. This enables flexible operation even under unforeseen demand surges or fluctuating energy availability, highlighting the benefits of decentralized learning in volatile environments.

2.3.3 Metaheuristic Algorithms for Charging Optimization

In addition to learning-based and game-theoretic approaches, metaheuristic algorithms such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Differential Evolution (DE) have been widely applied to EV systems, especially for large-scale optimization tasks involving charging, grid interaction, and infrastructure planning. These algorithms are often used in offline optimization scenarios where computational time is available and global optimality is prioritized over real-time adaptability.

2.3.3.1 Shaheen et al. (2024): Multi-Algorithm Metaheuristic Scheduling for V2G-Based Charging Optimization

Shaheen et al. present a metaheuristic optimization framework for scheduling EV charging and discharging using V2G technology to reduce user costs and support grid stability [74]. The study applies four bio-inspired metaheuristic algorithms, Differential Evolution (DE), Particle Swarm Optimization (PSO), Whale Optimization Algorithm (WOA), and Grey Wolf Optimizer (GWO) to manage EVs charging activities while parked, enabling bidirectional energy flow between EVs and the grid.

The objective is to minimize the daily charging costs incurred by EV owners while enhancing grid load management. The simulation results indicate that all four algorithms can effectively optimize EV scheduling; however, WOA demonstrated superior performance in terms of convergence speed and solution quality. The study confirms the feasibility of using metaheuristics to coordinate large-scale charging and discharging operations, especially under V2G-enabled systems. While this work successfully demonstrates offline optimization for cost-efficient and grid-supportive EV behavior, it does not address real-time adaptability or dynamic user interaction. In contrast, our DQN-based agent learns charging and serving strategies online, responding to uncertainties in passenger demand, energy levels, and station availability without requiring a full model of the grid or user behavior.

2.3.3.2 Piamvilai & Sirisumrannukul (2022): Genetic Algorithm for Load-Aware EV Charging with Queue Management

Piamvilai and Sirisumrannukul propose a smart charging strategy for managing spatially distributed and time-varying EV loads using a behavior-based simulation framework and a genetic algorithm (GA) [75]. The approach addresses key challenges posed by uncontrolled EV charging namely increased peak demand, line overloads, and voltage degradation in power systems.

To mitigate these effects, the authors introduce a charging event matrix, which models user behavior and time-dependent charging events, serving as input to the GA optimization process. The scheduling algorithm defines available charging slots that optimize the system load factor while maintaining user satisfaction. Additionally, a queuing management strategy is implemented to prioritize EVs based on prior charging durations and remaining state of charge (SoC), enhancing the fairness and efficiency of slot allocation. Simulation results on a modified IEEE-14 bus system confirm the method's capability to reduce peak demand and respect grid constraints. The framework integrates user behavior modeling with optimization, making it suitable for utility-scale demand management under controlled environments.

In contrast to this offline, operator-centered method, our DQN-based model emphasizes decentralized real-time learning. It allows an EV agent to adaptively decide when to charge or serve passengers based on local energy levels and environmental feedback—without relying on pre-generated matrices or centralized queuing heuristics.

2.3.3.3 Efthymiou et al. (2017): Genetic Algorithm for Charging Infrastructure Placement

Efthymiou et al. address the Electric Vehicle Charging Infrastructure Location Problem (EVCILP) by proposing a genetic algorithm (GA)-based optimization model for identifying optimal locations of public charging stations in urban areas [76]. In anticipation of increased EV adoption driven by decarbonization policies, the study focuses on equipping European cities with adequate charging infrastructure.

Due to limited real-world EV usage data, the authors use origin-destination (OD) matrices from conventional vehicle traffic and forecast future EV penetration to simulate charging demand. The GA model is implemented in R and applied to the city of Thessaloniki, Greece. The results reveal that 15 optimally located stations can satisfy 80% of the projected charging demand for 2020. The model outputs strategic charging locations across city zones and is offered as an open-source tool for public sector use.

This work contributes to long-term infrastructure planning by optimizing spatial deployment of chargers under demand uncertainty. However, it assumes static demand patterns and operates at a zonal level, without capturing temporal or behavioral variability. In contrast, our DQN-based system focuses on operational decision-making for individual agents interacting dynamically with charging infrastructure and demand, learning in real time to balance service goals and energy constraints.

2.4 Comparative Analysis Table

Table 2.3: Comparative Study of Electric Taxi Optimization Methods

Study / Approach	Model Type	Objectives	Advantages	Drawbacks
Rivière & Chung (2021) [68]	RL (H-TD2)	Fleet-wide dispatch via hybrid learning (local Q + global value)	Scalability through hybrid learning; balances global and local control	Requires multi-agent coordination; not energy-aware
Oda & Joe-Wong (2018) [69]	RL (DQN)	Minimize unserved rides via decentralized learning	High performance in large fleets; agent independence	Ignores energy and charging constraints; assumes ride-sharing focus
Cao et al. (2021) [70]	RL (Actor-Critic)	Cost-effective charging under uncertain demand	Handles continuous actions; reduces energy cost	Focus on grid-level charging; not tailored to taxi operations
Zhu et al. (2017) [71]	Game Theory	Joint transport and charging scheduling (cake-cutting)	Improves energy efficiency; tested on real data	Full observability assumption; no real-time adaptation
Zhang et al. (2024) [72]	Heuristic + Grid Model	Charging + routing to balance regional load	Reduces grid stress; models cyber-physical interaction	Offline, static planning; not responsive to agent-level variation

Study / Approach	**Model Type**	**Objectives**	**Advantages**	**Drawbacks**
Ma et al. (2025) [73]	MILP (2-phase)	Maximize profit + service rate with congestion-awareness	Considers dynamic pricing and congestion; realistic simulation	Centralized scheduling; lacks real-time adaptability
Shaheen et al. (2024) [74]	Metaheuristic (DE, PSO, WOA, GWO)	Cost reduction via V2G scheduling	Effective cost savings; comparative study of 4 algorithms	No real-time decision-making; assumes stationary EVs
Piamvilai & Sirisumrannukul (2022) [75]	Metaheuristic (GA)	Minimize peak load and maintain user satisfaction	Models behavior-based charging; considers queueing	Offline optimization; needs charging matrices and predefined profiles
Efthymiou et al. (2017) [76]	Metaheuristic (GA)	Optimal charging infrastructure location planning	Strategic long-term placement; open-source tool for cities	Operates at zonal level; assumes static demand patterns
Proposed Work	RL (Q-learning, DQN)	Maximize served passengers and minimize unnecessary charging	Learns adaptively; considers energy, congestion, and dynamic demand	Single-agent only; currently lacks multi-agent or city-scale coordination

2.5 Conclusion

This chapter has established a comprehensive foundation for understanding the core elements of EV taxi systems and reviewed the principal research streams addressing their operational challenges. First, the background section outlined the evolution of electric vehicles, the structure and constraints of charging infrastructure, and the specific requirements and limitations of electric taxis. These foundational insights highlight the need for intelligent decision-making in environments with limited energy, fluctuating demand, and constrained charging resources.

The review of related work classified the optimization strategies into three main categories. RL-based approaches have demonstrated promising results for dynamic decision-making under uncertainty, enabling agents to adapt in real-time to passenger demand and energy constraints. Heuristic and game-theoretic methods, while offering analytical tractability, often assume full system knowledge and are less flexible in dynamic contexts.

Metaheuristic algorithms, on the other hand, excel at solving complex, large-scale offline problems such as charging scheduling or infrastructure planning, but they lack the responsiveness needed for real-time applications.

The comparative analysis table has synthesized the strengths and limitations of these methods, illustrating a research gap at the intersection of adaptability, energy-awareness, and decentralized learning. In response, the proposed work introduces a DQN-based single-agent framework that trains an electric taxi to make intelligent serve-or-charge decisions. This model prioritizes real-time adaptability, minimal assumptions, and environment-driven learning, addressing the limitations observed in existing literature.

Overall, this chapter situates the proposed methodology within the broader academic landscape, justifying the use of reinforcement learning—particularly DQN as a suitable and scalable approach for optimizing electric taxi operations in dynamic urban environments.

Chapter 3

System Architecture and Reinforcement Learning Approaches

3.1 Introduction

Modern intelligent transportation systems increasingly rely on autonomous decision-making agents to improve operational efficiency, enhance user satisfaction, and minimize energy consumption [78, 79]. In this context, electric taxis face unique challenges such as limited battery capacity, dynamic passenger requests, and spatial constraints in urban environments.

This chapter presents the architectural and algorithmic foundations of a learning-based control framework designed to optimize electric taxi operations under energy constraints. The proposed system focuses on enabling agents to maximize the number of passengers served while minimizing unnecessary visits to charging stations.

To address the decision-making problem, the system initially employs *Q-learning*, a model-free reinforcement learning algorithm suitable for environments with discrete and relatively small state-action spaces [80]. However, as the problem complexity increases particularly in scenarios involving large grid sizes and multiple interacting agents, Q-learning becomes computationally infeasible due to its reliance on a tabular representation of the value function.

To overcome these limitations, the system integrates *DQN* [81], which combine reinforcement learning with deep neural networks to enable scalable policy learning in high-dimensional spaces. The DQN model generalizes across similar states and supports end-to-end learning from raw or structured state inputs.

The architecture is evaluated across two progressively complex scenarios:

- **Scenario 1 (S1)**: A single taxi, one charging station, and three passengers within a 5×5 grid environment. This baseline setting allows a direct comparison between Q-learning and DQN, emphasizing basic learning dynamics under limited spatial and operational complexity.
- **Scenario 2 (S2)**: A more complex configuration featuring a single taxi, three charging stations, and nine passengers in a 20×20 grid. This scenario extends the state and reward representations to account for greater environmental variability and operational decision-making under resource constraints.

These two scenarios are designed to benchmark the performance and adaptability of traditional Q-learning and deep reinforcement learning techniques under increasing levels of complexity. Both are trained using scenario-specific configurations of state representations, action spaces, and reward functions, enabling a structured evaluation of energy-aware decision-making and service efficiency in electric taxi systems.

This chapter is organized as follows: Section 3.2 outlines the motivation and system architecture. Section 3.3 describes the Q-learning and DQN model in detail. State, action, and reward representations are covered in Section 3.4. Sections 3.5 and 3.6 defines each scenario, discussing its structure, agent behavior, and algorithmic setup. The chapter concludes in Section 3.7, with hyperparameter tuning and evaluation metrics provided in the following chapter (Chapter 4).

3.2 Motivation and System Overview

The deployment of autonomous electric taxis in urban environments presents several challenges, including **limited battery capacity**, **dynamic passenger demand**, and the need for **real-time decision-making under uncertainty**. Traditional rule-based or static scheduling methods often lack the flexibility required to handle such complex and dynamic environments [82].

As the environment grows in scale—featuring more passengers, larger grids, and multiple charging stations—**heuristic approaches become less effective**. To overcome these limitations, we employ a *reinforcement learning* (RL) framework that allows the agent to learn optimal policies through interaction with the environment.

In this system, an electric taxi operates in a grid-based simulation. At each decision point, the agent must choose whether to *serve a passenger* or *recharge at a station*. This decision is based on its current energy level, the locations of passengers and charging stations, and recent outcomes. The objective is to **maximize the number of passengers served** while **minimizing energy waste and unnecessary charging**.

To support this learning process, we use a Deep Q-Network (DQN) [81], which enables the agent to approximate the optimal action-value function in a high-dimensional state space using a neural network.

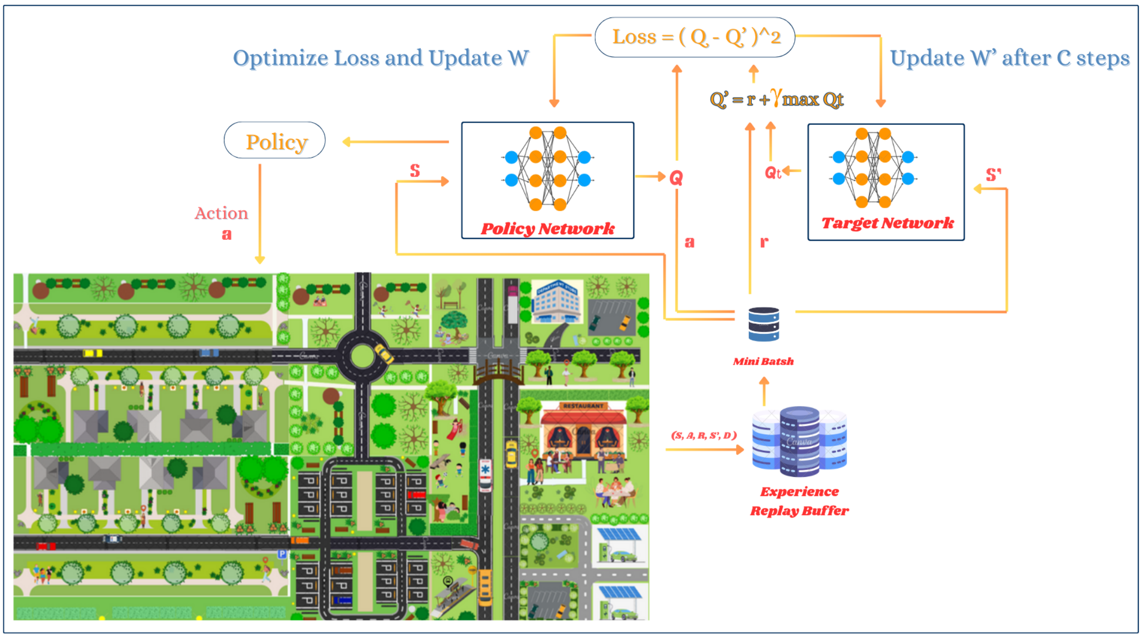


Figure 3.1: DQN architecture for single-agent electric taxi control.

Figure 3.1 illustrates the full DQN learning architecture used in this work. It shows how the agent interacts with the environment, stores experiences, and updates its networks through a structured learning process.

At each time step:

- The agent **observes the current state** of the environment.
- The **Q-network** estimates the value of each possible action.
- An action (serve or charge) is selected using an **ϵ-greedy strategy**.
- The selected action is **executed**, leading to a new state and a reward.
- The experience tuple (s, a, r, s', d) is stored in a **replay buffer**.
- A mini-batch is sampled from the buffer to **update the Q-network** using gradient descent.
- A separate **target network** is updated periodically to stabilize the learning process [83].

This cycle enables the agent to improve its policy over time and learn to make intelligent service and charging decisions based on experience.

The next section provides the technical design of both the DQN and Q-learning architectures, including the network structure, input representation, and exploration-exploitation strategies.

3.3 Learning Models Overview

3.3.1 Q-learning Algorithm

In this work, Q-learning was implemented in Scenario 1 (small grid) and Scenario 2 (larger grid), where environments are moderate in complexity. Q-learning is a model-free, value-based reinforcement learning algorithm that enables agents to learn optimal action-selection policies through experience without requiring prior knowledge of the environment's dynamics [80]. The core idea is to estimate the state-action value function, known as the Q-function, which captures the expected cumulative reward an agent will receive by taking an action a in a state s and following the optimal policy thereafter.

The Q-value update rule is defined by the Bellman equation:

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[r + \gamma \max_{a'} Q(s', a') - Q(s, a) \right] \tag{3.1}$$

where α is the learning rate, γ is the discount factor, r is the reward received after taking action a in state s, and s' is the resulting next state. A lookup table was used to store Q-values for each (s, a) pair, and the agent updated its values iteratively using the epsilon-greedy policy for action selection.

While effective in small environments, the tabular Q-learning approach exhibits critical limitations in scalability. As the number of state-action pairs increases, the memory and computational requirements grow exponentially, making it impractical for high-dimensional settings such as those involving continuous variables, multiple agents, or dynamic constraints. These challenges motivated the integration of DQN, which leverages deep learning to scale reinforcement learning to larger, more complex environments. This transition is elaborated in the next subsection.

3.3.2 Overview of DQN

DQN mark a pivotal advancement in RL, particularly for environments characterized by high-dimensional and continuous state spaces. Traditional tabular Q-learning methods, while effective in simple discrete settings, scale poorly when the number of states and actions increases [81]. DQN addresses this limitation by approximating the Q-value function using deep neural networks, enabling generalization across similar states and decision contexts.

In this framework, the Q-function $Q(s, a)$ is approximated using a neural network parameterized by θ, where s is the current state and a is an action. The output is a vector of Q-values, each corresponding to a discrete action in the action space. The goal of the training process is to minimize the temporal-difference error using the Bellman equation.

This architecture is particularly well-suited to the electric taxi dispatch problem, where the agent must make informed decisions based on variables such as remaining energy, agent status, and proximity to passengers or stations. Through continuous interaction with the environment, the agent learns to optimize its behavior for long-term cumulative reward.

3.3.3 DQN Model Architecture

The DQN used in this system is implemented as a fully connected feedforward neural network, designed to estimate Q-values over a discrete action space with two choices: *Charge* (action 0) or *Serve* (action 1). The input to the network is a structured, flattened state vector that encodes:

- Unique state and agent identifiers
- Current (x, y) grid position
- Remaining energy level
- Status of the agent (available, charging, occupied, etc.)
- Nearest passenger and station IDs
- Cumulative counts of clients served and charging events

To ensure numerical consistency, categorical features (e.g., agent status) are one-hot or label encoded. This preprocessing produces a fixed-size numerical input vector compatible with the neural network.

The architecture comprises the following layers:

- **Input Layer:** Receives the state vector (flattened representation).
- **Hidden Layers:** Two fully connected layers with 64 ReLU-activated neurons each.
- **Output Layer:** A linear layer with two neurons representing the Q-values for the two actions.

The model is trained using the Mean Squared Error (MSE) loss between predicted and target Q-values, and optimized using the Adam optimizer with a learning rate of 0.0005. A **target network** is maintained separately and synchronized with the main network every fixed number of steps (e.g., every 1000 iterations) to improve training stability [83].

An ϵ-greedy policy is used for exploration. The value of ϵ decays linearly from 1.0 to 0.01 over time, ensuring sufficient exploration in early training while gradually shifting toward exploitation of the learned policy [84].

This lightweight yet expressive architecture is designed to support efficient learning in constrained environments while maintaining generalization capability across a wide range of states.

3.3.4 Learning Mechanism

The learning mechanism in this work is based on the Deep Q-Learning framework, which integrates reinforcement learning principles with deep neural networks to approximate optimal action policies in environments with high-dimensional state spaces.

At the beginning of each episode, the taxi agent starts from a predefined initial state. The environment consists of a grid with charging stations and multiple passengers randomly placed. The agent interacts with the environment over a fixed number of steps or until all passengers are served or energy is depleted. Each interaction results in a transition tuple comprising the current state, selected action, received reward, the next state, and a done flag indicating episode termination. These transitions are stored in a replay buffer for subsequent training.

An epsilon-greedy policy is employed to select actions. Initially, the agent explores the environment by selecting actions at random. As learning progresses, the exploration rate (epsilon) decays, allowing the agent to increasingly exploit the learned Q-values to make decisions. This balance ensures sufficient exploration of the state-action space while guiding the agent toward optimal behaviors.

The agent is trained by sampling minibatches of transitions from the replay buffer and computing the target Q-values using the Bellman equation [85]:

$$Q_{\text{target}} = r + \gamma \max_{a'} Q_{\text{target}}(s', a') \tag{3.2}$$

where r is the reward received, γ is the discount factor (set to 0.9), s' is the next state, and Q_{target} refers to the target network. The loss is computed as the MSE) between the predicted Q-values and the target values, and the policy network is updated accordingly using gradient descent.

To stabilize learning, a separate target network is maintained, which is synchronized with the policy network at regular intervals (every 1000 steps) [81]. This helps reduce the moving-target problem during updates and improves convergence stability.

In each training episode, the following metrics are recorded: cumulative reward, pickup and dropoff rewards, remaining energy, steps taken, and the number of charging events. These metrics are used to monitor learning progression and evaluate the agent's behavior in balancing service efficiency with energy management.

This learning mechanism allows the agent to incrementally improve its decision-making policy across episodes, leading to effective energy-constrained navigation and task fulfillment in a multi-passenger grid environment.

3.4 Environment Design

3.4.1 State Space Representation

The state representation varies depending on the learning algorithm used. In the case of **Q-learning**, a mapping is maintained between a structured tuple (representing the full environment state) and a unique state identifier (`state_id`). This `state_id` is used to index rows in the Q-table, allowing efficient lookup and update of Q-values for discrete (s, a) pairs without needing to preprocess the entire state into a numerical vector.

In contrast, for the **DQN** model, the full state is explicitly transformed into a fixed-length numerical vector that can be fed into a neural network. All categorical and spatial variables are encoded numerically to ensure input consistency and neural compatibility. The following structure is adopted in Scenarios 1 and 2:

- **state_id**: Unique identifier for the state.
- **taxi_id**: Identifier for the taxi agent.
- **position**: Current position of the taxi, encoded as two separate values (x, y).
- **energy**: Remaining energy level.
- **passenger_id**: Nearest passenger identifier, encoded as an integer (0 = None, 1 = ID).
- **station_id**: Nearest charging station identifier, encoded as an integer (0 = None, 1 = ID).

- **status**: Encoded as an integer:
  - 0 = available
  - 1 = charging
  - 2 = occupied
  - 3 = searching_for_passenger
  - 4 = searching_for_station
- **clients_served**: Total number of passengers served.
- **charging_count**: Number of charging events performed.

To ensure compatibility with the neural network input, all categorical and structured variables are converted into fixed-length numerical representations. For example, string-based status values are mapped to integer codes, and coordinate tuples are separated into scalar values. This preprocessing guarantees that each state vector is fully numeric and of consistent dimensionality, suitable for use in the Deep Q-Network.

This representation enables the agent to learn context-aware policies based on location, energy levels, service status, and operational history. It strikes a balance between expressiveness and efficiency, allowing for meaningful decision-making while keeping the input dimension tractable.

3.4.2 Action Space Representation

The action space in this environment is defined as a discrete set of two high-level decisions available to the taxi agent at each decision point. These actions govern the agent's behavior and trajectory within the grid environment and are designed to reflect the primary operational choices in the service model. The discrete action space is defined as follows:

- **Action 0 – Charge:** Move to the nearest charging station using A*, recharge if needed, then return to available status.
- **Action 1 – Serve:** Move to the nearest passenger, pick them up, then transport them to their destination.

Each scenario involves a single agent choosing between only these actions (Charge or Serve) based on its energy and environment status. Q-learning and DQN both use this discrete action set during training.

At each decision point, the agent selects an action based on its learned policy, trained via reinforcement learning specifically Q-learning or its deep variant, the DQN. These methods allow the agent to explore the environment, gather experience, and iteratively refine its policy by updating Q-values according to observed rewards and state transitions.

By restricting the agent to these two essential yet strategic actions, the learning problem remains computationally tractable while still supporting complex, goal-oriented behavior. This binary action formulation facilitates focused learning and promotes faster convergence while effectively modeling the trade-off between energy management and passenger service in the dynamic grid environment.

3.4.3 Reward Function Design

The reward function plays a critical role in shaping the learning behavior of the taxi agent, guiding it to develop optimal policies that balance energy efficiency and service quality. In this environment, a carefully crafted reward structure is employed to reflect key performance objectives: ensuring efficient service delivery, minimizing unnecessary charging, and penalizing suboptimal behavior.

The total reward at each step is computed as a weighted combination of positive incentives and penalties, normalized by the total time taken. The reward function considers both immediate outcomes (e.g., successful pickup/drop-off) and long-term operational strategies (e.g., energy conservation and responsible charging behavior).

3.4.3.1 Reward and Penalty Components

The reward function is the same for both Q-learning and DQN for both scenarios. It incorporates the following components:

- **Energy Depletion Penalty:** If the taxi's remaining energy reaches zero, the agent incurs a penalty of **-10**. This discourages agents from depleting energy without planning to recharge.
- **Incorrect Action Penalty:** If the agent selects an action inconsistent with its current status (e.g., chooses to charge while searching for a passenger), a penalty of **-2** is applied. This enforces logical and context-aware decision-making.
- **Correct Action Reward:** If the chosen action aligns with the agent's status (e.g., charging when low on energy), a reward of **+2** is granted.
- **Pickup Reward:** When the agent picks up a passenger from the correct location while in the `searching_for_passenger` state, it receives a reward of **+100**.

- **Drop-off Reward:** When the agent drops off a passenger at the correct destination while in the `occupied` state and with energy > 0, it earns a reward of **+200**.
- **Charging Reward:** If the agent reaches a charging station in the `searching_for_station` state, it receives **+10**. By tracking the agent's charging history, if it initially arrives with energy below 80%, it gains **+5** for necessary charging. If charging was unnecessary (based on charging history), a penalty of **-5** is applied.
- **Idle Penalty:** If the agent remains idle in the same position while being `available`, a penalty of **-5** is imposed.
- **Unnecessary Charging Penalty:** If the taxi is fully charged and historical data shows it arrived without needing to charge, a penalty of **-10** is applied.
- **Excessive Charging Penalty:** If the number of charging events exceeds twice the number of passengers served, a penalty proportional to the number of charging events is applied:
$$-0.05 \times \text{Charging Events}$$
- **Energy Usage Penalty:** Each movement in the environment results in a consumption penalty:
$$-0.01 \times \text{Energy Consumed} - 0.1$$

3.4.3.2 Motivation Behind the Rewards Design

This reward formulation encourages:

- Strategic trade-offs between serving passengers and managing energy levels,
- Timely recharging without unnecessary station visits,
- Efficient movement across the environment with minimal energy waste,
- Avoidance of idle or illogical behavior,
- Maximization of service quality through high pickup and drop-off rewards.

Through continuous interaction with the environment, the agent uses Q-learning or DQN to update its value estimates and converge toward policies that maximize cumulative reward. The design ensures a balance between learning stability and realistic operational constraints, making it suitable for scalable deployment in electric taxi fleets.

3.4.3.3 Reward Function Formula

Based on the reward structures defined for each scenario in the preceding subsubsection, this part formalizes the overall reward computation used during training. While the core structure remains consistent across scenarios, as described earlier. The complete reward function used to guide learning is given by:

$$\text{Reward} = \frac{\text{Passengers}_{\text{Served}} \times (R_{\text{Pickup}} + R_{\text{Dropoff}}) + R_{\text{Action}} + R_{\text{Charge}} + \text{Penalty}}{\max(\text{Total Time Taken}, 1)}$$

Where:

- R_{Pickup} and R_{Dropoff} are the rewards for successful passenger service,
- R_{Action} is the reward for correct action selection,
- R_{Charge} includes rewards or penalties based on charging logic and history,
- Penalty accumulates all negative contributions (energy use, idleness, wrong actions, etc.),
- and Total Time Taken normalizes the reward to avoid large disparities across varying episode durations.

3.5 Scenario Definitions and Learning Configurations

This section presents two scenarios designed to evaluate the learning agents under varying operational conditions. Each scenario modifies aspects of the environment (e.g., grid size, number of passengers and stations), thereby altering the complexity and coordination required for successful task completion. Scenarios 1 and 2 involve single-agent configurations using both Q-learning and DQN.

3.5.1 Scenario 1: Small Grid – Q-learning & DQN

Scenario 1 is deployed within a compact 5×5 grid environment featuring a single taxi agent, one charging station, and three passengers. This scenario serves as a baseline to evaluate learning behavior in a simplified, low-dimensional setting conducive to early-stage policy development.

The state space, action space, and reward structure remain consistent between the Q-learning and DQN approaches. However, the implementation differs: Q-learning relies on a tabular method, maintaining a mapping between discrete state identifiers and Q-values, while DQN processes the full numerical representation of the state as input to a neural network.

At each decision step, the taxi begins from an *available* status in the state and chooses one of two high-level actions: (i) *Charge* or (ii) *Serve*. If the agent opts to charge, it locates the nearest operational station using Manhattan distance, updates its status to `searching_for_station`, and navigates to the station via an A* shortest path. Upon arrival, if its energy is not yet full, it enters the `charging` state, replenishing energy incrementally until fully charged. It then resets to `available`. If the taxi reaches the station already fully charged, it transitions immediately back to the `available` state.

Alternatively, if the *Serve* action is selected, the agent identifies the closest passenger using Manhattan distance, sets its status to `searching_for_passenger`, and moves toward the passenger via A*. Upon successful pickup, the taxi transitions to the `occupied` state and transports the passenger to their destination, after which the passenger is marked as `served` and the agent becomes `available` again.

After each action execution, the agent receives a reward computed based on the scenario-specific reward function, as discussed in Section **??**. The agent transitions to the next state, energy is decremented by 1 unit, and time is incremented. If Q-learning is used, the Q-table is updated using the Bellman equation via a state-ID-based mapping. If DQN is employed, the transition is stored in the replay buffer, and the network is trained using the steps outlined in the DQN pseudocode.

3.5.2 Scenario 2: Large Grid – Q-learning & DQN

Scenario 2 scales the environment to a 20×20 grid and introduces increased complexity by expanding the number of passengers to nine and deploying three charging stations with limited capacities. Unlike Scenario 1, stations in this setup may temporarily become unavailable when energy is depleted, thereby modeling realistic constraints on infrastructure resources.

The action and reward structures remain consistent with Scenario 1. However, the increased spatial complexity significantly enlarges the state space, posing additional learning challenges. This increase affects the tabular size and update frequency in Q-learning and results in a more diverse input distribution in DQN.

Q-learning remains applicable but begins to exhibit memory and convergence inefficiencies due to the curse of dimensionality. In contrast, DQN retains generalization capabilities, leveraging its neural representation to capture state-action correlations across broader configurations. The agent behaviors (e.g., serving, charging, movement logic) follow the same procedural flow as described in Scenario 1 but under increased decision pressure and environmental uncertainty.

3.6 Pseudocode of the DQN and Q-learning Algorithms

Both the Q-learning and DQN algorithms are employed in this study to train the taxi agent to learn optimal action-selection policies within the defined environment. While Q-learning uses a tabular approach for value estimation [80], DQN leverages neural networks to approximate Q-values in high-dimensional state spaces [81]. The following pseudocode summarizes the core learning steps used in both frameworks:

Algorithm 1 Q-learning Algorithm

```
Input: Learning rate α, discount factor γ, exploration rate ϵ, number of episodes N
Initialize: Q-table Q(s, a) arbitrarily
for each episode = 1 to N do
    Initialize state s
    while not terminal state do
        Choose action a using ϵ-greedy policy
        Take action a, observe reward r and next state s′
        Q(s, a) ← Q(s, a) + α[r + γ max_a′ Q(s′, a′) − Q(s, a)]
        Set s ← s′
    end while
end for
```

Algorithm 2 Deep Q-Network (DQN) for Taxi Agent

Initialize replay memory D to capacity N
Initialize Q-network with random weights θ
Initialize target network $\hat{Q}$ with weights $\theta^- \leftarrow \theta$
for each episode **do**
  Initialize environment and get initial state s
  for each step in the episode **do**
    Choose action a from s using ϵ-greedy policy derived from $Q(s, a; \theta)$
    Execute action a, observe reward r and next state s'
    Store transition (s, a, r, s') in D
    Sample mini-batch (s_j, a_j, r_j, s'_j) from D
    Compute target $y_j \leftarrow \begin{cases} r_j & \text{if terminal state} \\ r_j + \gamma \max_{a'} \hat{Q}(s'_j, a'; \theta^-) & \text{otherwise} \end{cases}$
    Perform gradient descent step on loss:
    $L(\theta) = \mathbb{E}[(y_j - Q(s_j, a_j; \theta))^2]$
    if mod(step, C) == 0 **then**
      Update target network: $\theta^- \leftarrow \theta$
    end if
    Set $s \leftarrow s'$
  end for
end for

3.7 Conclusion

This chapter provided a detailed overview of the proposed system architecture and its integration with reinforcement learning methods specifically, Q-learning and DQN for intelligent electric taxi optimization under energy constraints.

The environment was modeled using a structured state space, a binary action space (charging or serving), and a carefully designed reward function aimed at balancing short-term service performance with long-term operational efficiency. The reward structure was adapted across all scenarios to suit the corresponding settings under shared environmental conditions.

Two simulation scenarios of increasing complexity were defined to evaluate scalability and generalization:

- **Scenario 1** employed a small grid to validate early-stage learning and benchmark both Q-learning and DQN under minimal complexity.
- **Scenario 2** expanded the state space with a larger grid, additional stations, and more passengers, testing the limits of tabular Q-learning and demonstrating the scalability advantage of DQN.

Q-learning, with its tabular formulation, was leveraged effectively in Scenarios 1 and 2, offering interpretability and fast convergence in small-scale settings. However, it exhibited limitations in memory usage and generalization as the environment grew. DQN, in contrast, enabled value approximation via neural networks, allowing for robust learning even in high-dimensional, stochastic environments.

The learning pipeline for the agent was articulated through pseudocode, demonstrating how agent update policies through exploration, experience replay, and temporal-difference learning.

In sum, this chapter laid the foundation for intelligent decision-making in electric taxi systems by combining scalable learning models with environment aware system design. The following chapter will analyze the empirical performance of the proposed approach across all scenarios, comparing Q-learning and DQN in terms of efficiency, convergence, and behavioral effectiveness.

Chapter 4

Simulation and Performance Evaluation

4.1 Introduction

This chapter presents the simulation experiments and performance evaluation of the proposed control framework for autonomous electric taxis. The objective is to assess the learning efficiency, energy aware behavior, and service quality of agents trained using both Q-learning and DQN across multiple urban mobility scenarios.

Two deployment scenarios of increasing complexity were designed to evaluate and compare learning performance. Scenarios 1 and 2 are single-agent environments with small and large grid sizes, respectively, where both Q-learning and DQN are applied.

The simulation process is structured into two main phases:

- **Learning Phase:** Each algorithm is trained over 2000 episodes. Six key performance metrics are recorded throughout training to monitor convergence behavior and policy improvement.
- **Testing Phase:** A single episode is executed for each model using the learned policy. The same six metrics are evaluated to analyze performance in a static, real-time scenario without further learning.

To benchmark the effectiveness of DQN, two baseline approaches are included:

- A **Random Policy**, where actions are selected uniformly at random.
- A **Q-learning Policy**, using a tabular value function for discrete state-action pairs.

The resulting performance curves provide both quantitative and qualitative insights into the agents' abilities to make energy-efficient decisions, maximize passenger service, and adapt to dynamic grid environments across all scenarios.

4.2 Experimental Setup

This section describes the environment configuration, agent behaviors, and learning setups used to evaluate the proposed Q-learning and DQN approaches across two scenarios of increasing complexity.

4.2.1 Experimental Environment Settings

The simulation environments for Scenario 1 and Scenario 2, including grid dimensions, agent configurations, and passenger distributions, are comprehensively summarized in Table 4.1.

Table 4.1: Experimental Environment Settings

Parameter	**Scenario 1 (S1)**	**Scenario 2 (S2)**
Grid Size	5×5	20×20
Taxi Agents	1	1
Charging Stations	1 at (0, 0)	(1,5), (14,7), (4,14)
Charging Station Capacity	2000	1000 (each)
Initial Taxi Position	(1, 0)	(0, 0)
Initial Energy	15 units	75 units
Passengers	P1: (0,2) $\rightarrow$ (4,0) P2: (3,0) $\rightarrow$ (1,4) P3: (4,4) $\rightarrow$ (1,0)	P1: (19,19) $\rightarrow$ (3,0) P4: (5,17) $\rightarrow$ (17,10) P9: (19,0) $\rightarrow$ (0,19)
Complexity	Low	High

4.2.2 Agent Behavior and Movement Rules

The movement model and energy-related constraints governing agent dynamics are presented in Table 4.2.

Table 4.2: Agent Behavior and Movement Rules

Behavior Element	**Specification**
Movement Directions	Up, down, left, right (Manhattan grid)
Energy Consumption	-1 unit per move
Charging Mechanism	+1 unit per time step at station
Time	+1 per move

4.2.3 Q-Learning Configuration

The hyperparameters employed for Q-learning, including exploration strategy and learning rate, are detailed in Table 4.3.

Table 4.3: Q-Learning Configuration

Parameter	**S1**	**S2**
State Representation	Q-table indexed by state ID	Q-table indexed by state ID
Exploration Strategy	Epsilon-greedy	Epsilon-greedy
Epsilon Decay	1.0 $\rightarrow$ 0.01 (decay rate = 0.995)	1.0 $\rightarrow$ 0.01 (decay rate = 0.998)
Learning Rate (α)	0.01	0.01
Discount Factor (γ)	0.9	0.9
Episodes	2000	2000

4.2.4 DQN Configuration

The neural network architecture and training configurations adopted for the DQN approach are outlined in Table 4.4.

Table 4.4: DQN Configuration

Parameter	**S1**	**S2**
Input	Flattened state vector	Flattened state vector
Hidden Layers	2 $\times$ 64 neurons (ReLU)	2 $\times$ 64 neurons (ReLU)
Output	2 Q-values (Charge, Serve)	2 Q-values (Charge, Serve)
Replay Buffer Size	100,000	100,000
Batch Size	64	128
Target Network Update	Every 1000 steps	Every 5 episodes
Exploration Decay	1.0 $\rightarrow$ 0.01	1.0 $\rightarrow$ 0.05
Discount Factor (γ)	0.9	0.99
Optimizer	Adam	Adam
Learning Rate	0.0005	0.0005
Loss Function	Mean Squared Error (MSE)	Mean Squared Error (MSE)

4.2.5 Simulation Metrics Tracked

The performance metrics used for evaluating the agents' behaviors during training and testing phases are listed in Table 4.5.

Table 4.5: Simulation Metrics Tracked

Metric
Total cumulative reward
Cumulative pickup/dropoff reward
Total steps per episode
Charging events
Remaining energy

4.2.6 Hardware and Software Setup

The computational environment and libraries utilized for executing the experiments are specified in Table 4.6.

Table 4.6: Hardware and Software Setup

Component	**Specification**
Operating System	Windows 10
CPU	Intel Core i5
RAM	4 GB
GPU	NVIDIA GeForce MX110
Python Version	3.8.20
Libraries Used	NumPy, PyTorch, Tkinter
Visualization	Tkinter-based GUI

4.3 Learning Phase: Training Results

4.3.1 Introduction

In this section, we present the outcomes of the training phase, where agents were trained over 2000 episodes using two different strategies: DQN and Q-learning. The objective of this phase is to evaluate how each algorithm learns to optimize energy use, task completion (i.e., passenger pickups and drop-offs), and overall behavior in the simulated electric taxi environment.

To measure learning progress and agent performance, several cumulative and per-episode metrics were collected and analyzed. These include total reward, remaining energy, pickup rewards, charging events, drop-off rewards, and steps taken. The results provide insight into each algorithm's ability to learn effective policies that balance operational efficiency and task-oriented behavior. Figures presented in this section highlight the comparative performance of the two strategies and form the basis for selecting the most promising model for further testing and evaluation.

4.3.2 Scenario 1: A Small Grid (5×5) Single Agent Environments

4.3.2.1 Total Reward per Episode

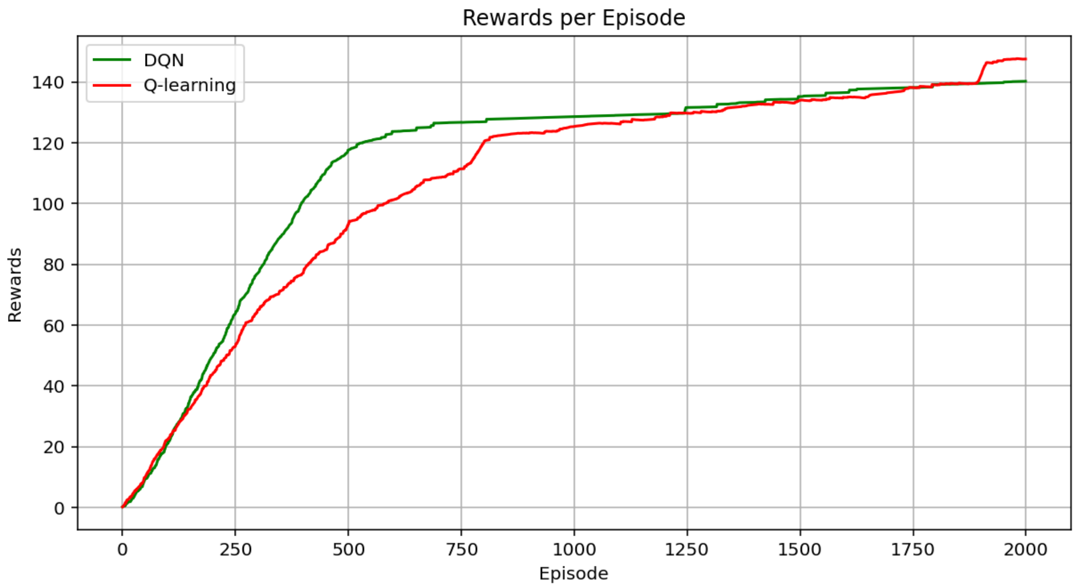


Figure 4.1: Total reward collected per episode for DQN (green) and Q-learning (red) over 2000 episodes..

Figure. 4.1 presents the episodic reward progression over 2000 training episodes for both the DQN and Q-learning agents in Scenario 1. The DQN agent exhibits a more rapid and smooth increase in total episodic rewards, reaching high performance levels earlier and converging more quickly. In contrast, the Q-learning agent demonstrates a slower rate of improvement with a more gradual learning curve. Although the Q-learning curve eventually approaches that of DQN in later episodes, the DQN's early advantage reflects its ability to efficiently extract meaningful patterns from the environment using neural function approximation.

This performance difference highlights DQN's capacity for better generalization and faster convergence in environments with moderately complex state spaces, such as a 5×5 grid with dynamic passenger and station locations. The improved reward trajectory suggests that the DQN agent learns to balance between serving and charging actions more effectively, leading to optimized task execution.

4.3.2.2 Cumulative Pickup Rewards

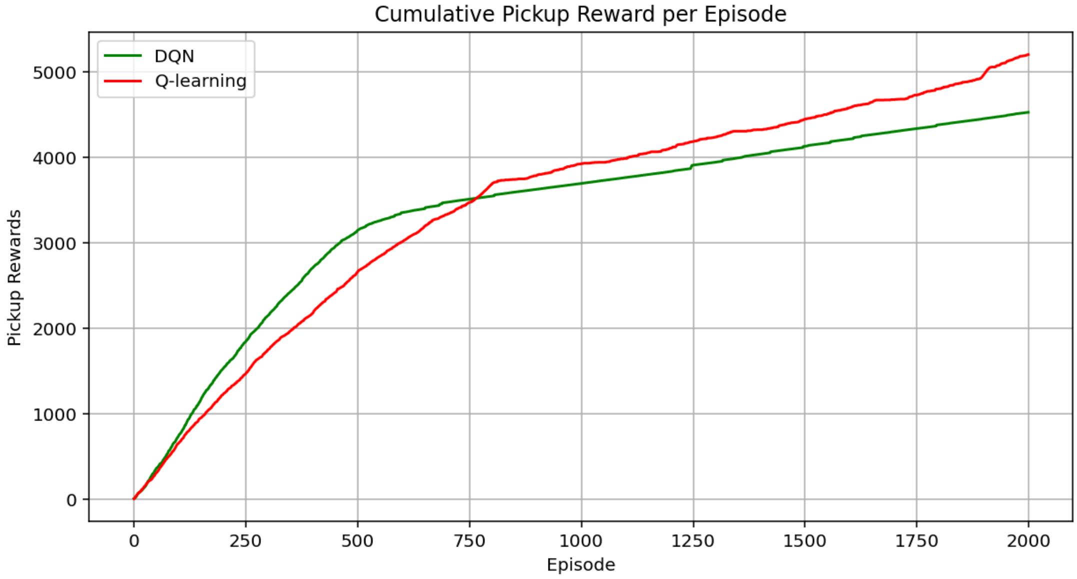


Figure 4.2: Cumulative pickup reward per episode for DQN (green) and Q-learning (red).

As shown in Figure. 4.2, the cumulative pickup rewards for the DQN agent increase sharply in the early stages of training but are later surpassed by the Q-learning agent beyond episode 750. This indicates that while DQN initially learns to identify and serve passengers more quickly, the Q-learning agent eventually catches up and exceeds DQN in terms of the total number of successful pickups.

This result implies that Q-learning, despite its slower convergence, becomes highly effective at executing pickup actions over time, possibly due to exhaustive exploration of the discrete state-action space. However, this improvement in pickup behavior does not necessarily reflect better overall efficiency, especially when considered in conjunction with the dropoff rewards and energy management behavior.

4.3.2.3 Cumulative Drop-off Rewards per Episode

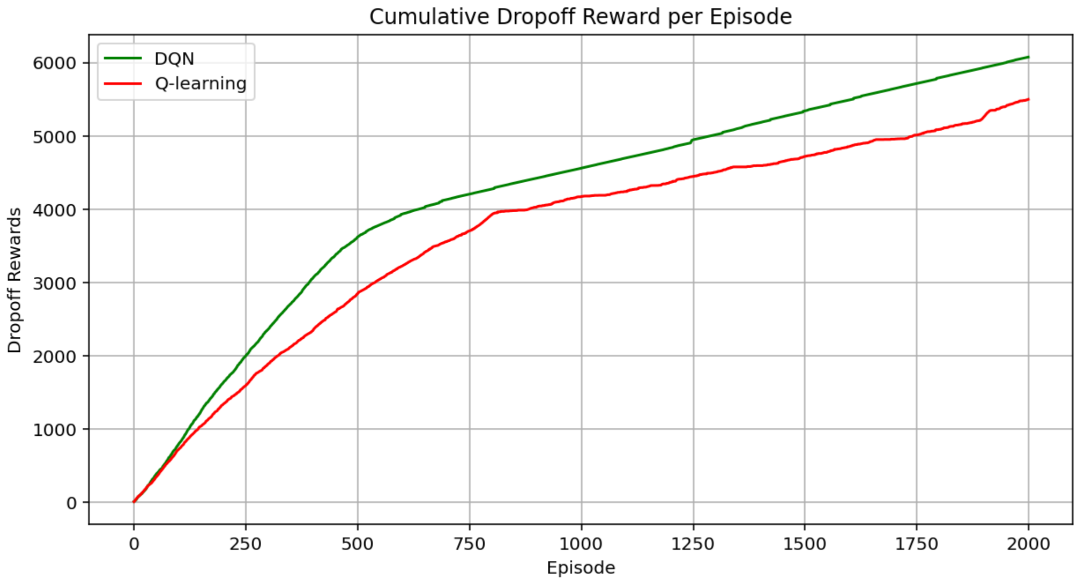


Figure 4.3: Cumulative reward for successful passenger drop-offs per episode.

Figure. 4.3 highlights the cumulative dropoff rewards over time, where the DQN agent significantly outperforms Q-learning throughout the entire training process. The divergence between the two curves grows steadily, indicating that the DQN agent consistently completes more end-to-end service tasks successfully picking up and dropping off passengers compared to Q-learning.

This sustained advantage demonstrates the DQN agent's ability to optimize long-term goals, such as completing full passenger service loops, rather than simply initiating pickups. The result underscores the effectiveness of deep reinforcement learning in capturing delayed rewards and managing multi-step tasks, which are critical for efficient operation in electric taxi dispatch systems.

4.3.2.4 Cumulative Remaining Energy per Episode

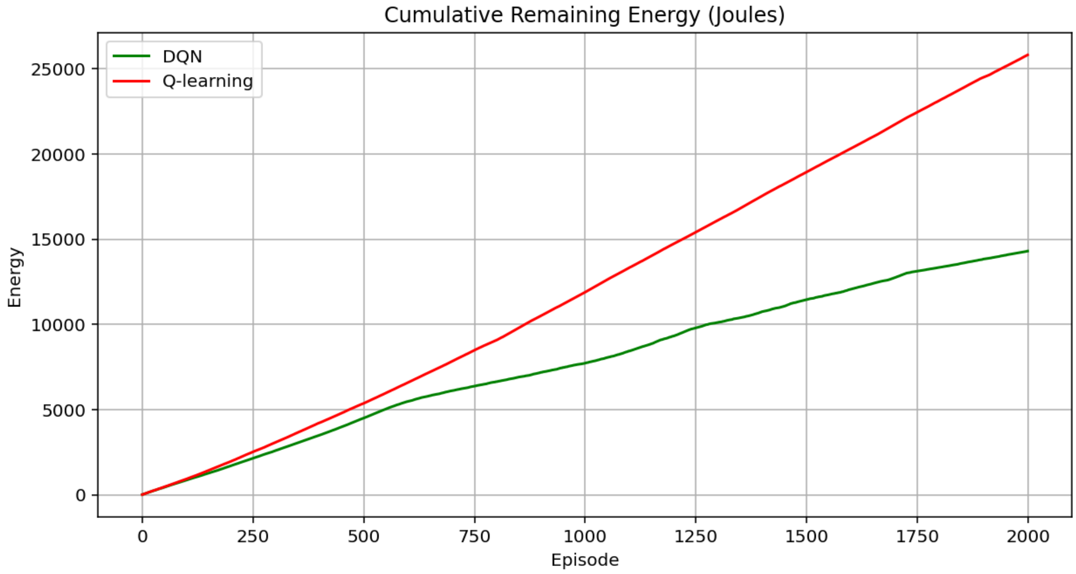


Figure 4.4: Cumulative remaining energy per episode for DQN (green) and Q-learning (red).

Figure. 4.4 indicate that the cumulative remaining energy at the end of episodes is generally higher for Q-learning than for DQN, as evidenced by the upward trajectory of the red curve. While surplus energy might suggest successful charge management, when viewed alongside excessive charging frequency, it points to overcharging behavior—the agent often charges more than necessary.

By contrast, the DQN agent maintains lower cumulative remaining energy, which indicates a balanced consumption pattern. This reflects an agent that plans its energy usage according to expected demand, avoiding wasteful recharging. The result supports the conclusion that DQN not only performs more tasks but does so while maintaining tighter control over energy resources.

4.3.2.5 Cumulative Steps Taken per Episode

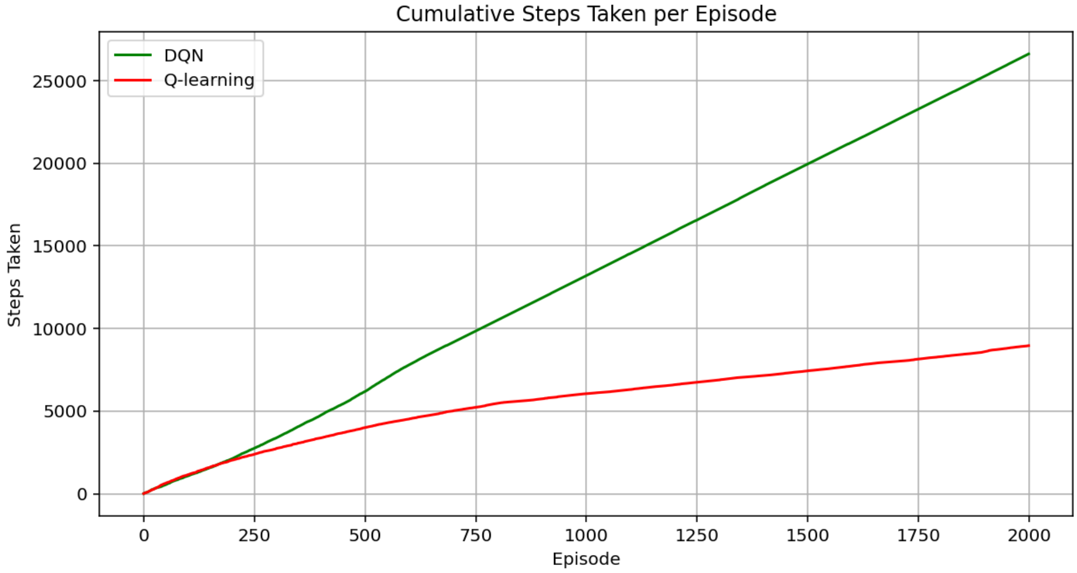


Figure 4.5: Total steps executed per episode for each learning strategy.

Figure. 4.5 shows that the cumulative number of steps per episode is consistently higher for the DQN agent compared to Q-learning. This suggests that the DQN agent performs more actions and interacts more actively with the environment, likely completing more service-related tasks such as pickups and dropoffs. In contrast, Q-learning's lower step count may result from frequent early charging returns, inefficient routing, or missed opportunities to serve passengers.

This metric further substantiates the productivity advantage of DQN. A higher number of executed steps, coupled with fewer charges and balanced energy usage, implies a more purposeful and efficient exploration strategy. The DQN agent appears to learn optimal movement policies that maximize operational coverage while minimizing energy waste, which is essential in real-world electric taxi operations.

4.3.2.6 Cumulative Charging Events per Episode

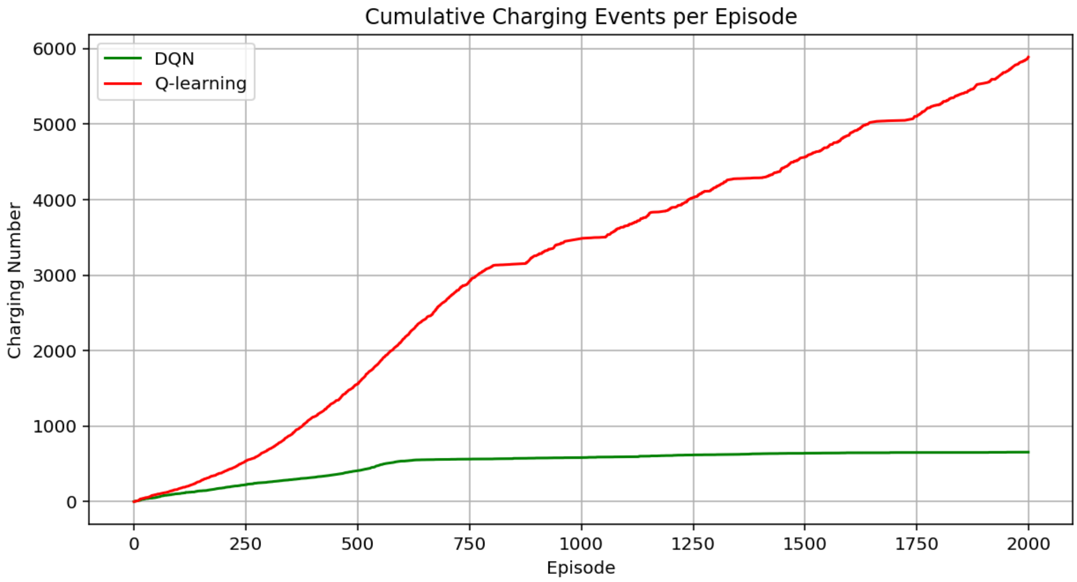


Figure 4.6: Total charging events accumulated per episode for Q-learning (red) and DQN (green).

This Figure. 4.6 depicts the cumulative number of charging events over 2000 episodes for both learning strategies. The Q-learning agent exhibits a consistently steep growth in charging frequency, indicating frequent returns to charging stations. In contrast, the DQN agent demonstrates a significantly flatter curve that begins to plateau around episode 1500, suggesting a decrease in charging reliance as learning progresses.

This disparity reveals important behavioral differences: Q-learning agents tend to overuse charging stations, likely due to their inability to generalize energy-aware policies efficiently. Meanwhile, the DQN agent learns to incorporate energy constraints into its decision-making, reducing redundant charging and prioritizing optimal routes. This reinforces DQN's superior energy efficiency and policy quality in long-term planning.

4.3.2.7 Conclusion

Collectively, these performance metrics underscore the enhanced adaptability and learning effectiveness of the DQN approach within Scenario 1. Although the Q-learning agent demonstrates occasional strengths in specific aspects, such as cumulative pickups or residual energy levels, these advantages are offset by its tendency toward inefficient charging behavior and lower overall task execution. In contrast, the DQN agent exhibits a more integrated and balanced strategy, effectively managing energy resources, fulfilling service demands, and navigating the environment—capabilities that are critical for efficient electric taxi operations in complex urban environments.

4.3.3 Scenario 2: A Large Grid (20×20) Single Agent Environments

4.3.3.1 Total Reward per Episode

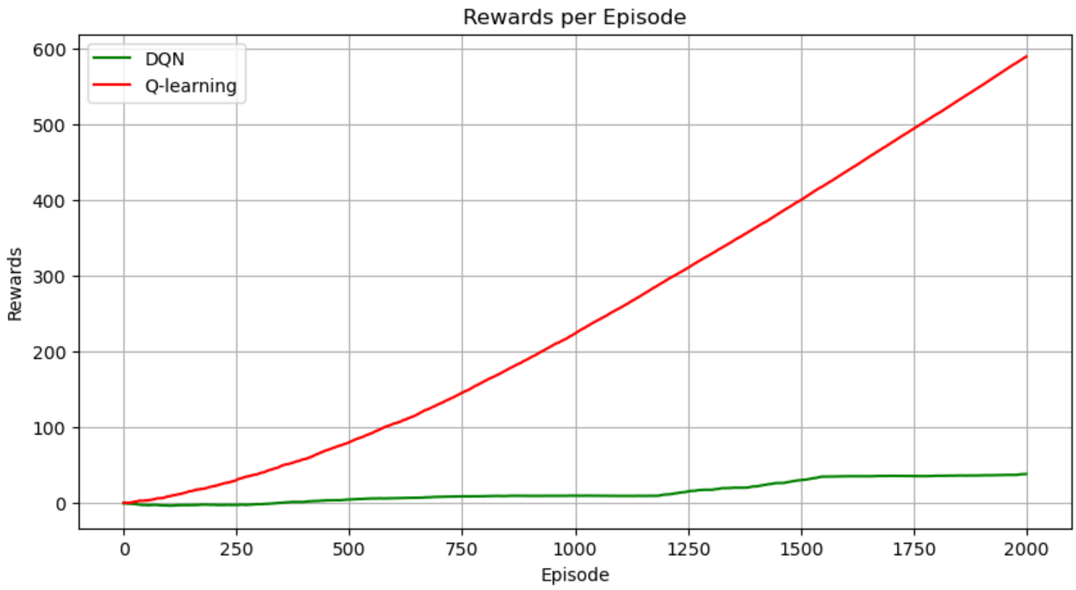


Figure 4.7: Total reward collected per episode for DQN (green) and Q-learning (red) over 2000 episodes..

In Figure. 4.7 the episodic reward trajectory indicates that the Q-learning agent exhibits a steep and consistent improvement in total rewards over time, suggesting effective policy acquisition and robust task performance. In contrast, the DQN agent shows slow and limited reward accumulation, indicating difficulty in adapting to the increased dimensionality and complexity of the environment. This outcome suggests that in larger state spaces, Q-learning's exhaustive tabular exploration may offer better short-term learning benefits, while DQN struggles to generalize effectively with limited data.

4.3.3.2 Cumulative Pickup Rewards

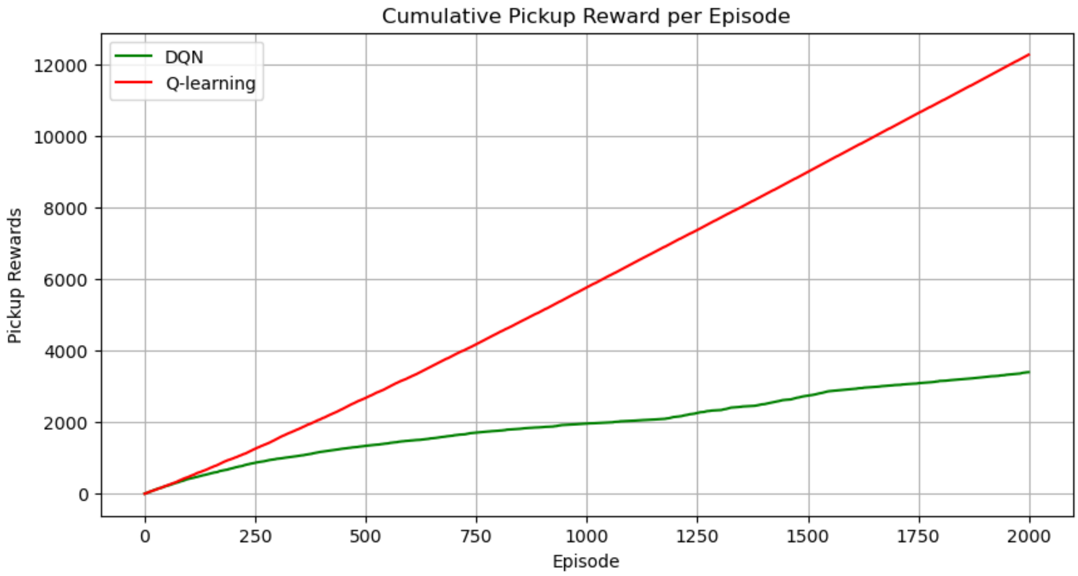


Figure 4.8: Cumulative pickup reward per episode for DQN (green) and Q-learning (red).

In Figure. 4.8, Q-learning outperforms DQN in cumulative pickup rewards, displaying a faster and more substantial growth curve. This result implies that Q-learning is more successful in identifying and serving passengers, which directly correlates with improved task execution. DQN's comparatively slower pickup rate further reinforces its difficulty in navigating and servicing within the more complex environment, possibly due to inadequate function approximation in high-dimensional state-action mappings.

4.3.3.3 Cumulative Remaining Energy per Episode

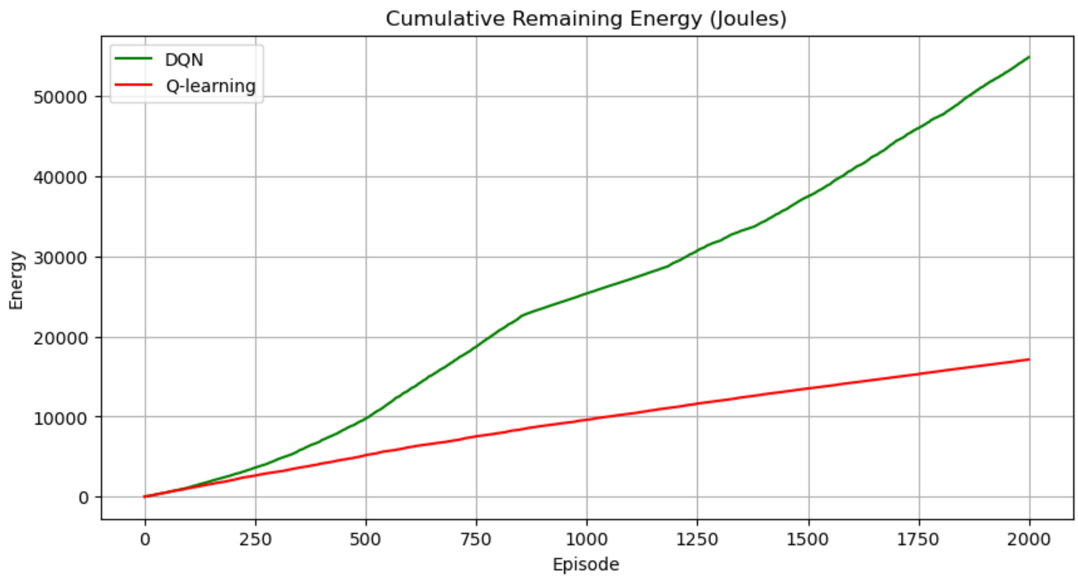


Figure 4.9: Cumulative remaining energy per episode for DQN (green) and Q-learning (red).

In Figure. 4.9 the DQN agent consistently maintains higher levels of remaining energy, suggesting a strategy focused on conserving resources and avoiding unnecessary energy expenditure. This contrasts with Q-learning, which tends to consume more energy in pursuit of higher task completion. While DQN's behavior reflects energy-aware planning, it may also signify overly cautious or conservative policies that hinder service productivity.

4.3.3.4 Cumulative Steps Taken per Episode

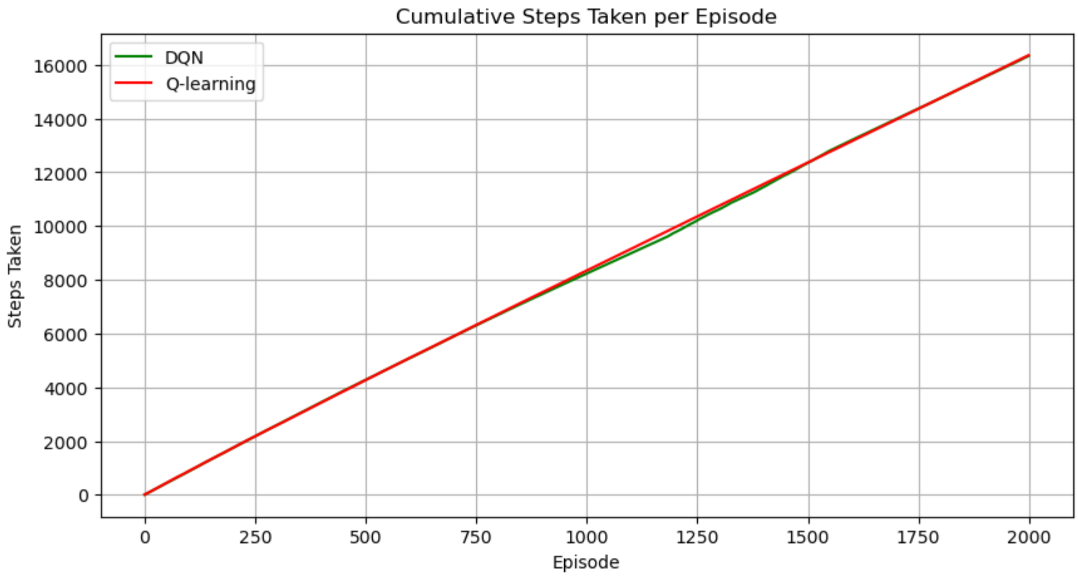


Figure 4.10: Total steps executed per episode for each learning strategy.

In Figure. 4.10 both agents exhibit nearly identical growth in cumulative steps, indicating a similar degree of environmental exploration. However, the similarity in movement does not equate to performance equivalence. Given Q-learning's superior results in rewards and pickups, it can be inferred that Q-learning utilizes its steps more effectively for productive outcomes, whereas DQN's actions may involve less efficient routing or redundant movements.

4.3.3.5 Cumulative Charging Events per Episode

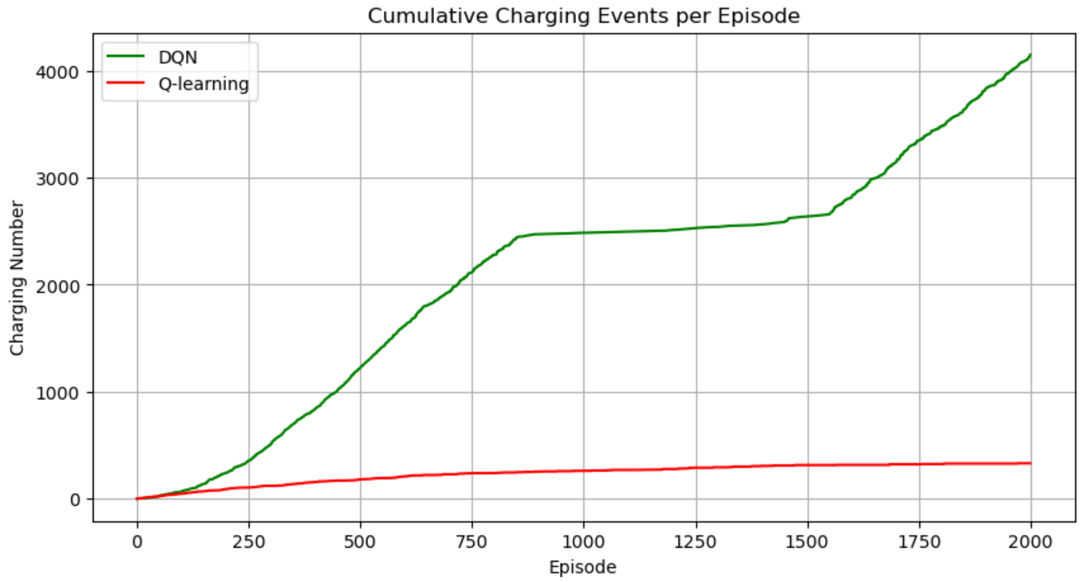


Figure 4.11: Total charging events accumulated per episode for Q-learning (red) and DQN (green).

Figure. 4.11 shows that DQN exhibits a significantly higher frequency of charging events, indicating a strong emphasis on maintaining energy reserves. This conservative energy management behavior contrasts with Q-learning's more aggressive task-oriented strategy, which involves fewer charging episodes but results in higher energy depletion. While DQN demonstrates robustness in energy sustainability, its frequent charging may detract from its ability to service more passengers within a given episode.

4.3.3.6 Conclusion

The evaluation of Scenario 2 highlights a distinct trade-off between task execution efficiency and energy management strategies. Q-learning demonstrates superior performance in terms of reward maximization and task completion, which can be attributed to its exhaustive exploration of discrete state spaces and its focus on short-term gains. In contrast, the DQN agent adopts a more conservative approach, emphasizing long-term energy sustainability and resource preservation, albeit resulting in reduced pickup frequency and overall rewards. These observations indicate that Q-learning may be more effective in contexts that prioritize high service throughput, whereas DQN appears better suited for applications where energy constraints and cautious decision-making are critical considerations.

4.4 Testing Phase: Evaluation Results

4.4.1 Introduction

4.4.2 Scenario 1: A Small Grid (5×5) Single Agent Environments

4.4.2.1 Total Rewards vs Number of Passengers

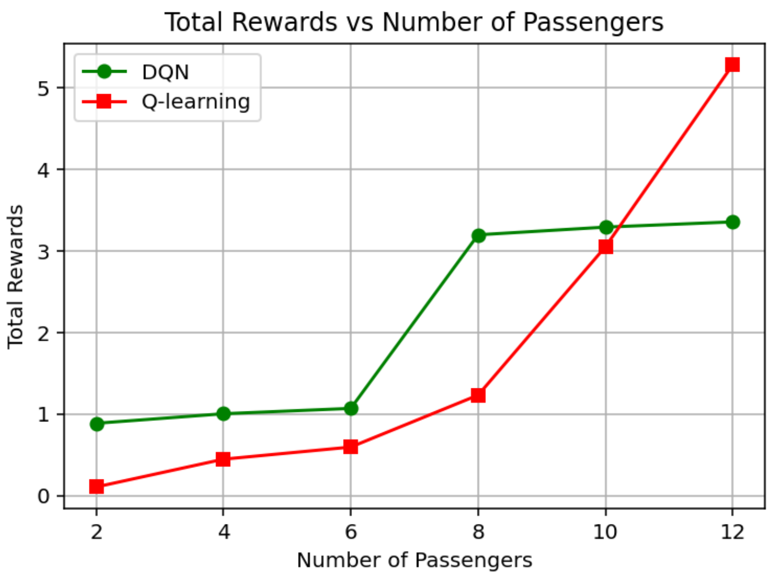


Figure 4.12: Total Rewards vs Number of Passengers in Scenario 1 (Training Phase)

As shown in Figure 4.12, Q-learning exhibits a steep and pronounced rise in cumulative rewards as the number of passengers increases, especially at higher loads (10–12 passengers). This suggests an aggressive pursuit of immediate reward signals. In contrast, DQN displays a more gradual, stable increase in total rewards, indicative of a conservative yet consistent strategy.

Although Q-learning achieves higher absolute rewards in high-demand settings, this behavior may reflect overfitting to the training environment or an overemphasis on short-term returns. DQN's smoother reward progression implies a more generalized policy with improved stability across varying conditions.

4.4.2.2 Charging Frequency vs Number of Passengers

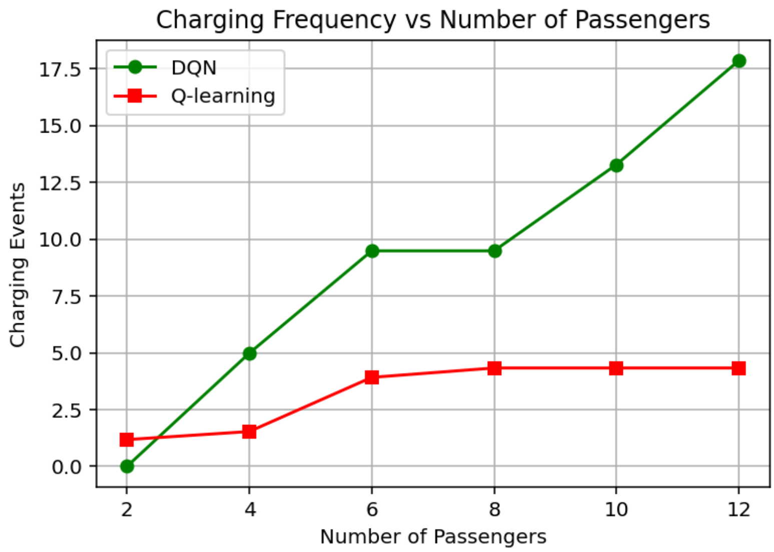


Figure 4.13: Charging Frequency vs Number of Passengers in Scenario 1 (Training Phase)

As shown in Figure 4.13, DQN engages in charging more frequently across all passenger levels, with a charging rate that scales with increasing demand. This behavior reflects a proactive and risk-averse approach to energy management. In contrast, Q-learning's relatively flat charging frequency suggests delayed or risk-prone charging behavior, which may compromise service continuity in high-load scenarios.

Frequent yet strategic charging by DQN enables uninterrupted task completion, reinforcing its robustness and operational reliability.

4.4.2.3 Steps Taken vs Number of Passengers

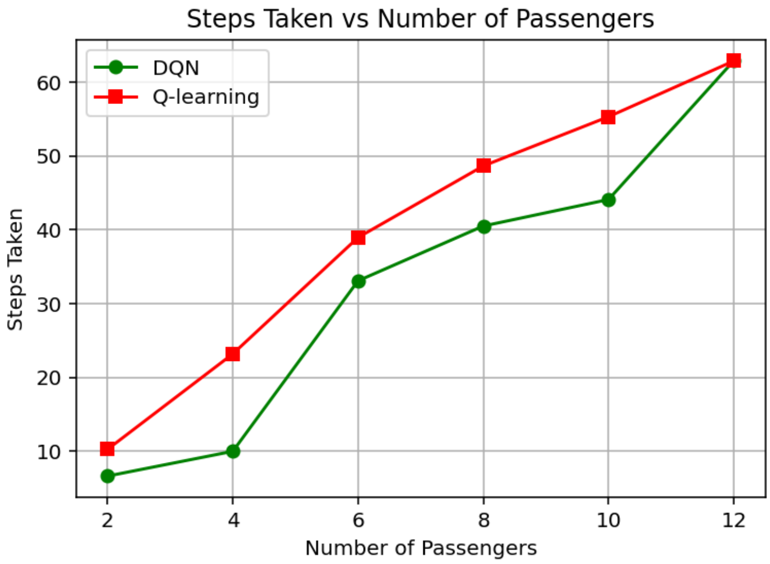


Figure 4.14: Steps Taken vs Number of Passengers in Scenario 1 (Training Phase)

As shown in Figure 4.14 the step count analysis reveals that Q-learning consistently takes more steps to serve passengers, particularly as the passenger count increases. This suggests less efficient routing and task execution, likely resulting from reactive planning and redundancy in path selection.

Conversely, DQN demonstrates greater spatial and temporal efficiency, completing tasks with fewer steps. Since each step carries energy and time costs, this finding further supports DQN's advantage in practical deployment.

4.4.2.4 Dropoff Rewards vs Number of Passengers

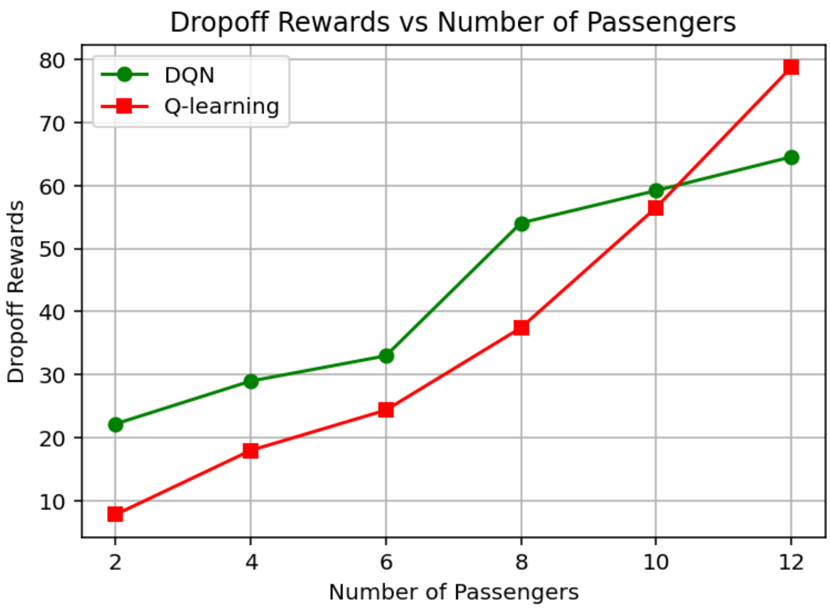


Figure 4.15: Dropoff Rewards vs Number of Passengers for DQN and Q-learning strategies.

As illustrated in Figure 4.15, Dropoff reward analysis reveals DQN's clear advantage across nearly all tested passenger counts, with the exception of the highest level (12 passengers), where Q-learning marginally surpasses it. This trend suggests that DQN is more effective at completing full service cycles—both pickup and dropoff—reinforcing its capacity for sequential planning under operational constraints.

The superior dropoff performance further confirms DQN's reliability in executing complete, high-value tasks, which are essential in real-world deployment scenarios where partial service (pickup without dropoff) is insufficient.

4.4.2.5 Remaining Energy vs Number of Passengers

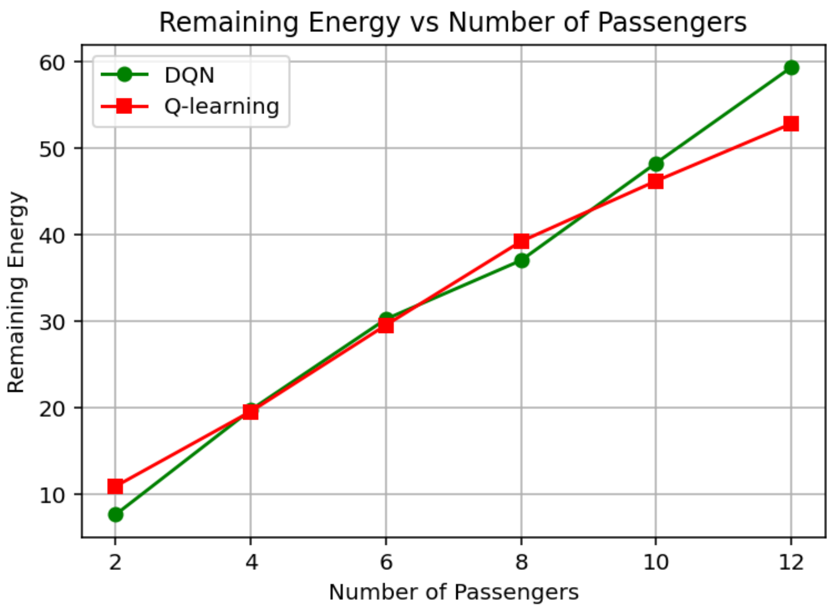


Figure 4.16: Remaining Energy vs Number of Passengers for DQN and Q-learning strategies.

As shown in Figure 4.16 as passenger load increases, DQN consistently maintains higher levels of remaining energy at the end of the episode. This energy conservation reflects more optimal routing, decision-making, and proactive energy management. In contrast, Q-learning depletes energy more rapidly, likely due to suboptimal planning or lack of foresight in energy budgeting.

From a systems perspective, energy efficiency is a critical operational constraint. DQN's superior performance in this metric makes it more suitable for sustained deployment in energy-constrained environments.

4.4.2.6 Pickup Rewards vs Number of Passengers

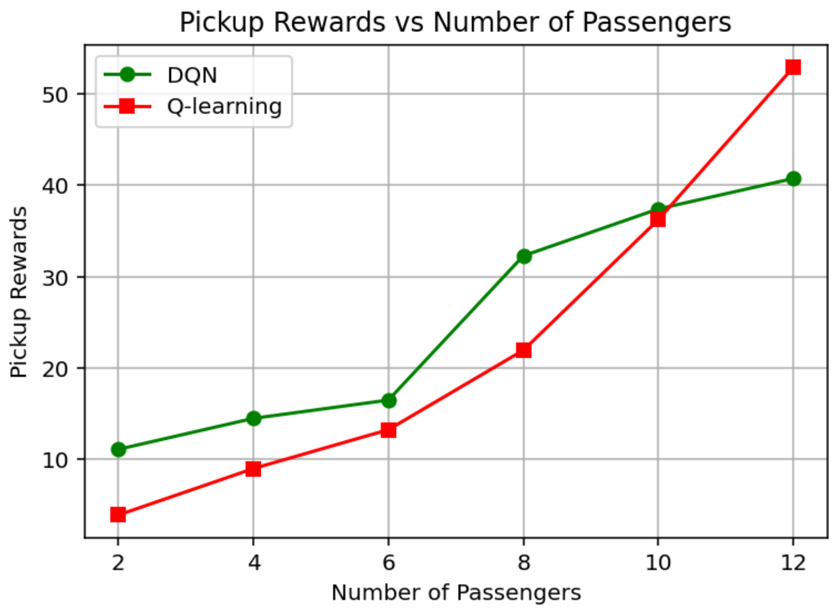


Figure 4.17: Pickup Rewards vs Number of Passengers for DQN and Q-learning strategies.

According to Figure 4.17, DQN achieves superior pickup rewards at lower passenger counts (2–8), demonstrating its capacity for effective passenger identification and task execution in simpler operational scenarios. Q-learning begins to outperform DQN only when the passenger count exceeds 10, likely due to its broader exploratory behavior.

However, the transition in dominance must be interpreted cautiously. The Q-learning agent's later advantage in pickups does not necessarily imply greater efficiency, especially considering its associated energy inefficiencies and higher step counts in those same conditions.

4.4.2.7 Conclusion

Although Q-learning achieves higher total rewards under heavy passenger demand, it does so through more energy-intensive and step-heavy behavior, reflecting short-term optimization rather than sustainable policy learning. In contrast, DQN maintains a balanced and robust strategy, optimizing energy usage, charging behavior, and movement efficiency. These characteristics make DQN more suitable for deployment in realistic, resource-constrained autonomous mobility systems where long-term viability and consistent performance are critical.

4.4.3 Scenario 2: A Large Grid (20×20) Single Agent Environments

4.4.3.1 Total Rewards vs Number of Passengers

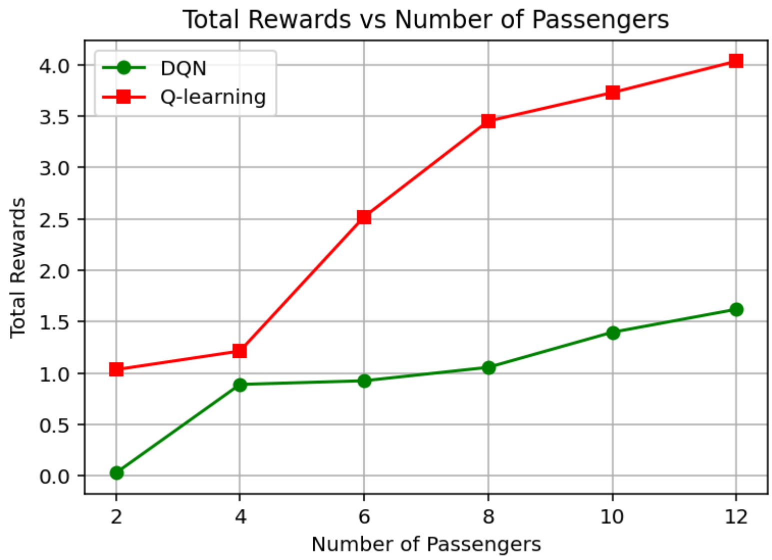


Figure 4.18: Total Rewards vs Number of Passengers in Scenario 1 (Training Phase)

According to Figure 4.18 across all tested passenger counts, Q-learning achieves substantially higher total rewards than DQN. The Q-learning agent demonstrates a strong ability to scale with increasing passenger demand, maintaining high reward accumulation even under heavier loads. In contrast, DQN's total rewards grow gradually and remain significantly lower, suggesting challenges in adapting to the increased complexity of the environment.

This disparity indicates that Q-learning is more effective at maximizing reward signals in single-episode deployments, likely due to its exhaustive tabular representation and exploitation of well-defined state-action mappings. DQN's relatively poor reward performance highlights its limitations in generalizing over large state spaces within limited interaction time.

4.4.3.2 Charging Frequency vs Number of Passengers

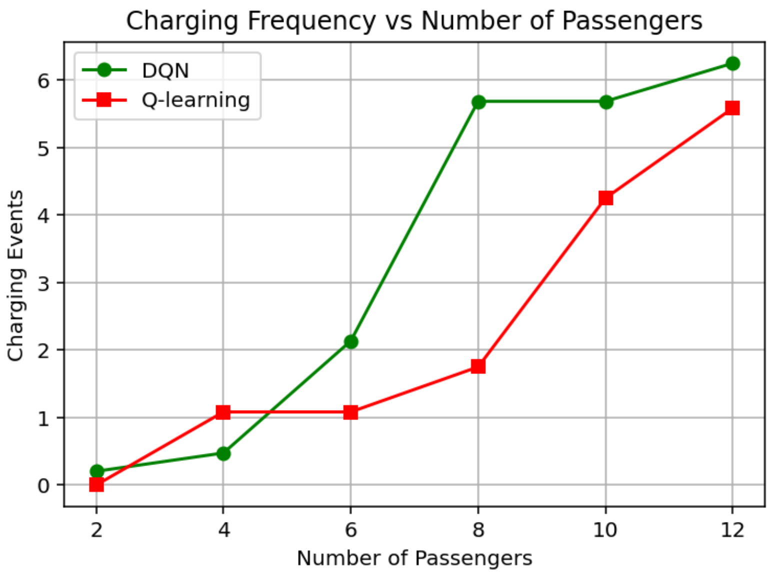


Figure 4.19: Charging Frequency vs Number of Passengers in Scenario 1 (Training Phase)

In Figure 4.19, DQN consistently exhibits a higher number of charging events than Q-learning. This behavior reflects its strategy of maintaining sufficient energy through proactive and frequent recharging, particularly as the number of passengers—and therefore task demand—increases. Conversely, Q-learning engages in fewer charging events, especially under lower passenger loads.

While DQN's frequent charging supports its energy conservation results, it may also divert attention from task completion. Q-learning, by contrast, prioritizes performance at the expense of energy robustness, revealing a trade-off between reliability and reward maximization.

4.4.3.3 Steps Taken vs Number of Passengers

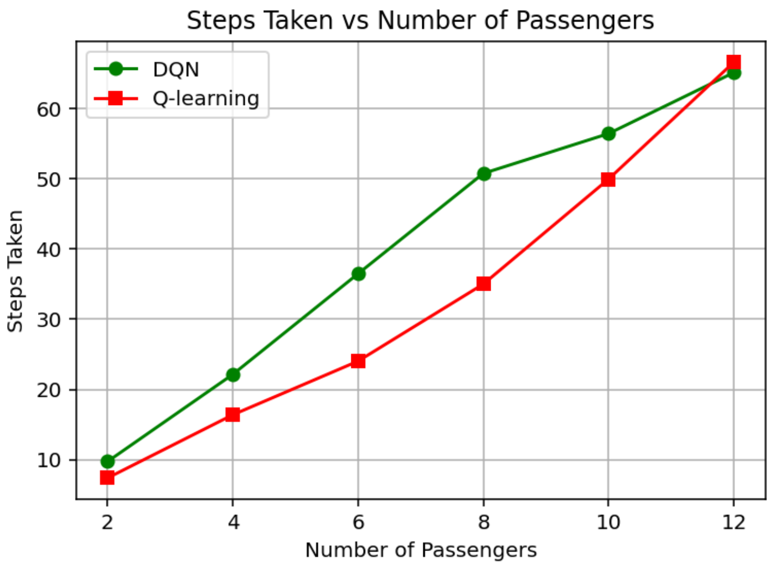


Figure 4.20: Steps Taken vs Number of Passengers in Scenario 1 (Training Phase)

Figure 4.20 shows that Q-learning completes its tasks using fewer steps compared to DQN at nearly every passenger count. This suggests that Q-learning agents follow more direct or efficient paths to serve passengers, which may contribute to their higher reward scores.

DQN's increased step count indicates more complex routing behavior, possibly due to more cautious planning or suboptimal decision sequences. While this may contribute to better energy management, it also limits the agent's ability to service multiple requests within a constrained time window.

4.4.3.4 Dropoff Rewards vs Number of Passengers

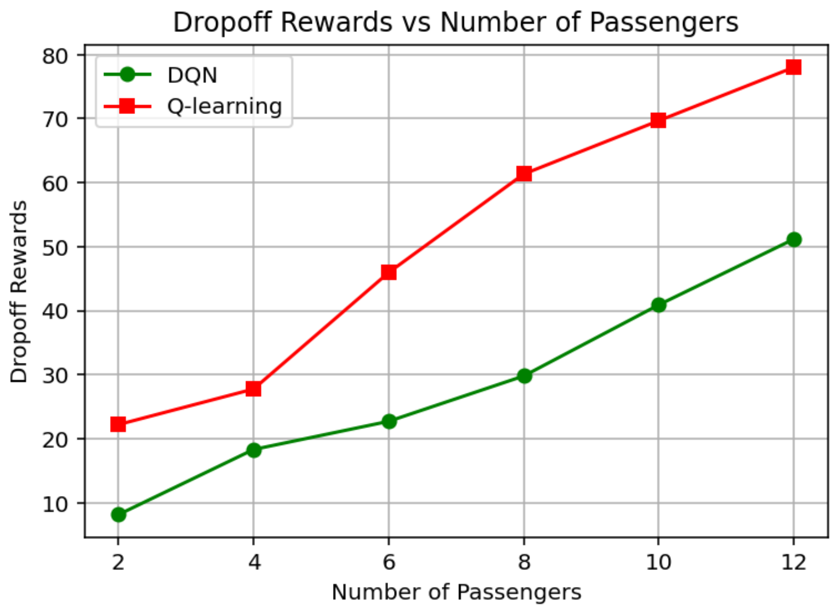


Figure 4.21: Dropoff Rewards vs Number of Passengers for DQN, Q-learning, and Random strategies.

Figure 4.21 illustrate that Dropoff reward trends closely mirror those observed in pickup performance. Q-learning again outpaces DQN significantly, demonstrating a stronger capacity for completing full passenger service cycles.

This superior performance may result from Q-learning's simplicity and effectiveness in exploiting short-term reward opportunities. DQN's gradual reward accumulation suggests difficulty in executing full pickup–dropoff sequences efficiently in this high-dimensional scenario, despite its long-term planning capabilities.

4.4.3.5 Remaining Energy vs Number of Passengers

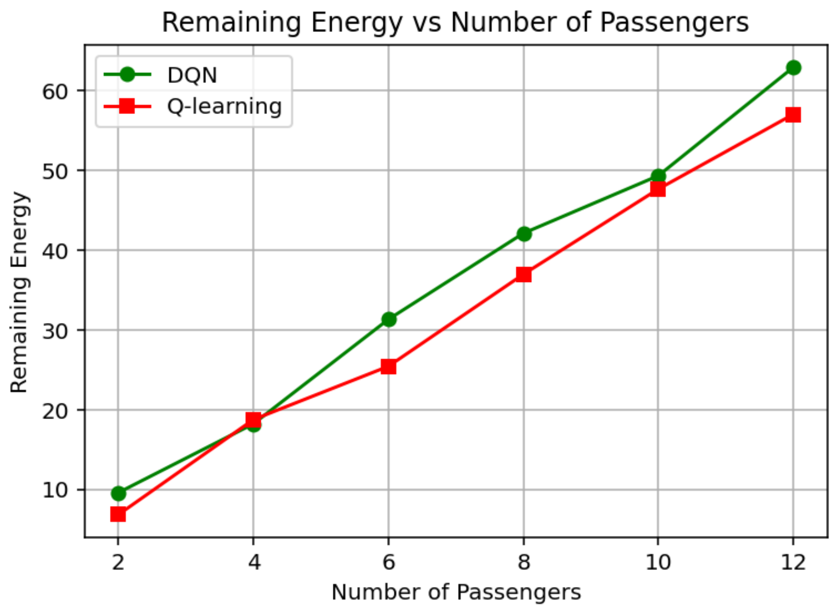


Figure 4.22: Remaining Energy vs Number of Passengers for DQN, Q-learning, and Random strategies.

Figure 4.22 shows that unlike the reward-based metrics, DQN maintains higher levels of remaining energy across most passenger levels, with the difference becoming more pronounced as demand increases. This indicates a more conservative operational style, likely characterized by frequent recharging and energy-aware routing.

From a sustainability perspective, this outcome highlights DQN's strength in managing energy constraints. Its cautious behavior may result in lower task performance but ensures continued operation under energy-limited conditions.

4.4.3.6 Pickup Rewards vs Number of Passengers

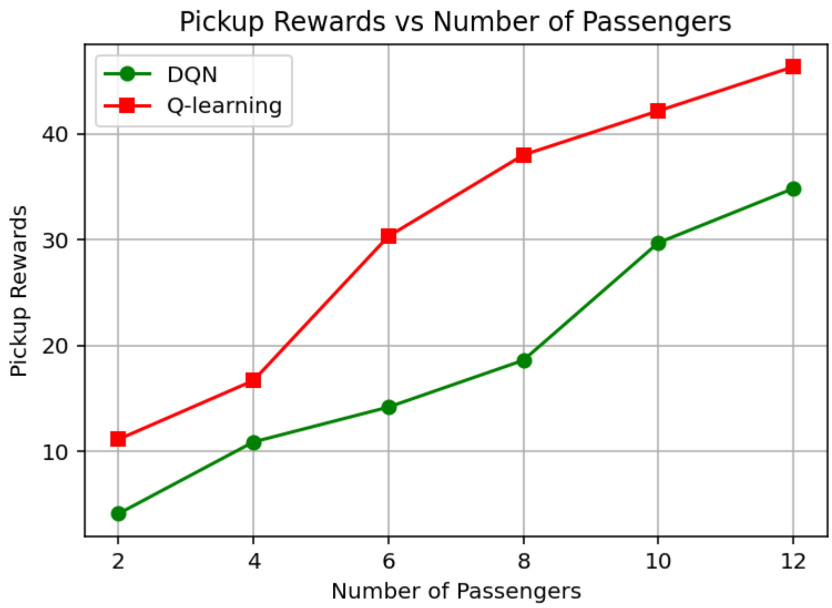


Figure 4.23: Pickup Rewards vs Number of Passengers for DQN, Q-learning, and Random strategies.

In Figure 4.23, Q-learning consistently outperforms DQN in cumulative pickup rewards across all passenger counts. Although both agents show a positive trend with increasing passenger numbers, DQN consistently lags behind.

This result suggests that Q-learning is more proficient at locating and reaching passengers under high-demand conditions. Its exploration strategy appears more effective in rapidly identifying service opportunities, while DQN may struggle to prioritize pickups when faced with a more complex spatial configuration.

4.4.3.7 Conclusion

The single-episode test results in Scenario 2 highlight a clear divergence in agent behavior and optimization strategies. Q-learning consistently outperforms DQN in key performance metrics related to task execution, such as total rewards, pickups, and dropoffs. Its aggressive and direct approach enables it to operate efficiently in environments demanding high throughput. However, this comes at the cost of reduced energy efficiency and charging frequency.

Conversely, DQN demonstrates superior energy management, engaging in more frequent charging and conserving energy across scenarios. While its conservative behavior limits immediate task performance, it suggests a higher degree of robustness in energy-constrained contexts. These findings reinforce the notion that Q-learning is more suitable for maximizing short-term performance, whereas DQN may be more appropriate in scenarios where long-term sustainability, energy constraints, and operational reliability are critical considerations.

4.5 Comparative Evaluation

Table 4.7: Comparative Analysis of DQN and Q-learning Across Training and Testing Phases

Metric / Criterion	**Scenario 1 (Small Grid)**	**Scenario 2 (Large Grid)**	**Preferred Model (Contextual)**
Learning Convergence	DQN converges faster and more smoothly	Q-learning improves steadily and overtakes DQN in rewards	DQN (S1), Q-learning (S2)
Total Reward	DQN performs better early on; Q-learning catches up later	Q-learning significantly outperforms DQN	Q-learning (S2), DQN (S1 early)
Pickup Rewards	DQN initially higher, Q-learning surpasses after episode 750	Q-learning consistently higher	Q-learning
Dropoff Rewards	DQN significantly higher across training and testing	Q-learning consistently better	DQN (S1), Q-learning (S2)
Energy Efficiency	DQN maintains lower energy surplus → better consumption balance	DQN maintains higher remaining energy → more sustainable behavior	DQN
Charging Frequency	DQN charges less but more strategically	DQN charges more frequently, supports energy robustness	DQN
Steps Taken	DQN takes more steps (linked to higher productivity)	Q-learning uses fewer steps more efficiently	Q-learning (S2), DQN (S1 context)
Scalability / Generalization	DQN generalizes well in small environments	Q-learning scales better in larger, deterministic grids	DQN (small), Q-learning (large)
Planning Quality	DQN better at full service cycles (pickup → dropoff)	Q-learning better at task density, less sequential planning	DQN (sequential), Q-learning (dense tasks)
Risk Management / Robustness	DQN uses cautious, proactive charging; handles uncertainty better	Q-learning favors aggressive task pursuit with less energy planning	DQN

4.6 Discussion and Comparative Insights

The comparative investigation of DQN and Q-learning throughout both the training and testing phases reveals distinct learning behaviors and adaptation mechanisms reflective of each algorithm's underlying design. In Scenario 1 (5×5 grid), characterized by a smaller and more tractable state space, DQN demonstrates superior performance relative to Q-learning in several key areas—namely faster convergence, higher cumulative drop-off rewards, and more effective energy management. These advantages are primarily attributable to DQN's capacity for function approximation, which allows for rapid generalization and more stable policy formation. Q-learning, although initially slower, gradually attains competitive pickup performance due to its comprehensive exploration of discrete state-action spaces.

In contrast, in the more complex Scenario 2 (20×20 grid), Q-learning outperforms DQN across the majority of service-oriented metrics, including total rewards, pickups, and drop-offs during both training and single-episode evaluation. This improvement can be linked to Q-learning's exhaustive tabular approach, which facilitates thorough state exploration in environments with high passenger demand. However, this reward-driven strategy tends to compromise energy efficiency, as evidenced by reduced charging behavior and more rapid energy depletion—factors that could limit its practicality in real-world deployments where operational reliability is critical.

DQN, meanwhile, maintains a more conservative policy in such environments, prioritizing frequent recharging and careful energy usage. While this results in fewer completed service tasks, it enhances long-term operational robustness, making it better suited for energy-constrained applications. The results thus reveal a fundamental trade-off between reward maximization (favored by Q-learning) and sustainable resource management (favored by DQN).

In summary, Q-learning is more effective in maximizing short-term performance, especially in deterministic, large-scale environments, whereas DQN excels in maintaining system stability and efficiency under resource constraints. These complementary strengths suggest that algorithm selection should be context-dependent, informed by the operational goals and constraints of the target system.

4.7 Conclusion

This section has provided a detailed performance comparison between DQN and Q-learning within simulated electric taxi dispatch tasks of increasing complexity. The experimental outcomes confirm that neither approach consistently outperforms the other across all performance dimensions. Instead, each method displays specific advantages:

- **DQN** demonstrates strong performance in smaller-scale environments and under energy constraints, showing faster convergence, greater energy efficiency, and higher task completion consistency.
- **Q-learning** proves more effective in larger, high-demand settings, achieving higher cumulative rewards and superior task throughput through exhaustive exploration.

These findings highlight an inherent tension between **operational efficiency** and **energy resilience**, with each algorithm favoring one at the expense of the other. Future research may benefit from exploring **hybrid strategies** or **context-aware frameworks** that dynamically adapt the learning policy based on environmental complexity and energy availability. Such approaches could integrate the generalization capacity of DQN with the precision and aggressiveness of Q-learning, thereby improving the robustness and scalability of autonomous mobility systems in diverse real-world scenarios.

Chapter 5

General Conclusion

5.1 Summary of the Problem

Urban electric taxi systems face challenges related to limited energy resources, sparse charging infrastructure, and inefficient dispatching. This thesis investigated how intelligent decision-making algorithms can improve energy utilization and task performance in such constrained environments.

5.2 Research Contributions

The main contributions of this thesis are as follows:

- Proposed a DQN-based framework for intelligent dispatch of electric taxis.
- Developed a simulation environment modeling realistic energy constraints, charging stations, and passenger demand.
- Conducted a comparative analysis between DQN and Q-learning across multiple operational scenarios (e.g., small vs. large grid).
- Demonstrated that DQN provides energy-efficient and sustainable policies, while Q-learning excels in maximizing short-term rewards.

5.3 Key Findings

The experiments revealed that:

- DQN outperforms Q-learning in energy management, charging frequency, and full trip completion in smaller and energy-constrained environments.
- Q-learning achieves higher cumulative rewards and better pickup/drop-off ratios in larger grids but at the cost of energy inefficiency.

- A clear trade-off exists between short-term reward maximization and long-term energy sustainability.

5.4 Limitations

While the results are promising, the simulation is limited in scope and does not incorporate real-time traffic conditions, time-varying demand, or coordination among multiple agents.

5.5 Future Work

Future research could explore:

- Multi-agent extensions involving collaboration or competition among electric taxis.
- Integration with real-world mobility and energy datasets.
- Use of hybrid models combining DQN with A2C, PPO, or metaheuristic strategies for dynamic adaptation.
- Testing in real-time, distributed simulations or hardware-in-the-loop (HIL) environments.

5.6 Final Remarks

Electric taxis offer a sustainable vision for future mobility. This thesis contributes a foundation for intelligent, energy-aware dispatching using reinforcement learning. Continued advancements in learning algorithms, simulation fidelity, and system integration will be crucial for realizing scalable, efficient, and reliable electric mobility solutions in smart cities.

Bibliography

[1] Bruno Scrosati, Jürgen Garche, and Werner Tillmetz. *Advances in Battery Technologies for Electric Vehicles*. Woodhead Publishing, 2015.

[2] Aziz Rachid, Hassan El Fadil, Khawla Gaouzi, Kamal Rachid, Abdellah Lassioui, Zakariae El Idrissi, and Mohamed Koundi. Electric vehicle charging systems: Comprehensive review, 2023.

[3] Olubusayo Aina. *Electric Vehicles: The Future of Transportation*. n.p., n.d.

[4] Amir Khajepour, M. Saber Fallah, and Avesta Goodarzi. *Electric and Hybrid Vehicles: Technologies, Modeling and Control - A Mechatronic Approach*. Wiley, 2014.

[5] I. Besselink, P. Van Oorschot, E. Meinders, and H. Nijmeijer. Design of an efficient, low weight battery electric vehicle based on a vw lupo 3l. In *Proceedings of EVS-25*, Shenzhen, China, 2010.

[6] United States Environmental Protection Agency. Epa – united states environmental protection agency. https://www.epa.gov/, 2024. Accessed: 2025-06-11.

[7] Ali Fazeli. *Development of a Novel Air Hybrid Engine*. Phd thesis, University of Waterloo, Canada, 2011.

[8] C. C. Chan and K. T. Chau. *Modern Electric Vehicle Technology*. Oxford University Press, 2001.

[9] A. Khaligh and Z. Li. Battery, ultracapacitor, fuel-cell, and hybrid energy storage systems for electric, hybrid electric, fuel-cell, and plug-in hybrid electric vehicles: State of the art. *IEEE Transactions on Vehicular Technology*, 59(6):2806–2814, 2010.

[10] D. M. Bellur and M. K. Kazimierczuk. Dc-dc converters for electric vehicle applications. In *Proceedings of the Electrical Insulation Conference and Electrical Manufacturing*, 2007.

[11] A. Emadi, S. S. Williamson, and A. Khaligh. Power electronics intensive solutions for advanced electric, hybrid electric, and fuel-cell vehicular power systems. *IEEE Transactions on Power Electronics*, 21(3):567–577, 2006.

[12] Seth Leitman and Bob Brant. *Build Your Own Electric Vehicle, Third Edition*. McGraw-Hill Education TAB, 2013.

[13] L. Lu, X. Han, J. Li, M. Ouyang, and G. Lu. A review on the key issues for lithium-ion battery management in electric vehicles. *Journal of Power Sources*, 226:272–288, 2013.

[14] K. Lu. Hardware architecture development of battery management system in electric vehicles. *Electronic Engineering and Product World*, 25(05):28–31, 2018.

[15] Rui Xiong and Weixiang Shen. *Advanced Battery Management Technologies for Electric Vehicles*. Wiley, 2019.

[16] Brian Culp. *Electric Cars For Dummies*. For Dummies, 2022.

[17] F.R. Spellman. *The Science of Electric Vehicles: Concepts and Applications*. CRC Press, 2023.

[18] X. Xiao, H. Molin, P. Kourtza, A. Collin, G. Harrison, S. Djokic, J. Meyer, S. Müller, and F. Möller. Component-based modelling of ev battery chargers. In *Proceedings of the 2015 IEEE Eindhoven PowerTech*, pages 1–6, Eindhoven, The Netherlands. IEEE.

[19] M. C. Kisacikoglu. *Vehicle-to-Grid (V2G) Reactive Power Operation Analysis of the EV/PHEV Bidirectional Battery Charger*. Ph.d. thesis, University of Tennessee, Knoxville, TN, USA, 2013.

[20] P. Kumar, S. Nikolovski, and Z. Y. Dong, editors. *Internet of Energy Handbook*. CRC Press, Boca Raton, FL, USA, 1 edition, 2021.

[21] D. Ronanki, A. Kelkar, and S. S. Williamson. Extreme fast charging technology—prospects to enhance sustainable electric transportation, 2019.

[22] T. J. C. Sousa, D. Pedrosa, V. Monteiro, and J. L. Afonso. A review on integrated battery chargers for electric vehicles. *Energies*, 15(8):2756, 2022.

[23] B. E. Lebrouhi, Y. Khattari, B. Lamrani, M. Maaroufi, Y. Zeraouli, and T. Kousksou. Key challenges for a large-scale development of battery electric vehicles: A comprehensive review. *Journal of Energy Storage*, 44:103273, 2021.

[24] Y. Wang, Y. Huang, Y. Wang, and X. Song. A review of electric vehicle charging infrastructure development and standards in china, 2020.

[25] F. Mwasilu, J. J. Justo, E. K. Kim, T. D. Do, and J. W. Jung. Electric vehicles and smart grid interaction: A review on vehicle to grid and renewable energy sources integration, 2014.

[26] P. Ghosh, S. Sinha, A. Palit, and S. Sharma. A comprehensive review on electric vehicle charging infrastructure and power grid challenges, 2022.

[27] M. Rarbach. Electric mobility standards overview, 2013. Internal publication referenced in EV standards documentation.

[28] J. Anthony. Charging infrastructure in the european ev market, 2013. Report on EV connector types and EU policy.

[29] CHAdeMO Association. Chademo technical specifications, 2013.

[30] PHOENIX CONTACT. Combined charging system (ccs) overview, 2013. White paper on CCS implementation by leading automakers.

[31] International Electrotechnical Commission. Iec 62196: Plugs, socket-outlets, vehicle connectors and vehicle inlets – conductive charging of electric vehicles, 2020. Available via IEC Standards Database.

[32] International Organization for Standardization. Iso 15118: Road vehicles — vehicle to grid communication interface, 2020. Available via ISO Standards Catalogue.

[33] E. Narassimhan, C. Johnson, and E. Elkind. Standardizing electric vehicle charging infrastructure: Challenges and opportunities, 2022.

[34] International Energy Agency. Global ev outlook 2023: Catching up with climate ambitions, 2023.

[35] Lewis Pickett et al. Electric vehicles and infrastructure, 2021. UK House of Commons Library Briefing.

[36] Myriam Neaimeh et al. A probabilistic approach to combining smart meter and electric vehicle charging data to investigate distribution network impacts, 2015.

[37] Department for Business, Energy & Industrial Strategy. Powering our net zero future, 2020. White Paper.

[38] Daryl Flack. The cyber security of smart metering and how it can be applied to ev charging, 2020. YouTube video.

[39] Roberto Metere, Myriam Neaimeh, Charles Morisset, Carsten Maple, Xavier Bellekens, and Ricardo M. Czekster. Securing the electric vehicle charging infrastructure. https://arxiv.org/abs/2105.02776, 2021. Last revised 6 Jul 2022.

[40] P. B. Andersen et al. Parker project final report, 2019. Parker Project, Denmark.

[41] International Energy Agency. Global ev outlook 2025, 2025.

[42] National Renewable Energy Laboratory. Electric vehicle charging infrastructure trends from the alternative fueling station locator, 2025.

[43] B. A. Rayan, U. Subramaniam, and S. Balamurugan. Wireless power transfer in electric vehicles: A review on compensation topologies, coil structures, and safety aspects, 2023.

[44] S. S. Williamson, U. Madawala, and D. Kumar. Guest editorial: Advances in wireless power transfer technologies, 2022.

[45] H. Wen, P. Wang, J. Li, J. Yang, K. Zhang, L. Yang, Y. Zhao, and X. Tong. Improving the misalignment tolerance of wireless power transfer system for auv with solenoid-dual combined planar magnetic coupler, 2023.

[46] C. Panchal, S. Stegen, and J. Lu. Review of static and dynamic wireless electric vehicle charging system, 2018.

[47] K. Zhou, Y. Wu, X. Wu, Y. Sun, D. Teng, and Y. Liu. Research and development review of power converter topologies and control technology for electric vehicle fast-charging systems, 2023.

[48] G. Rajendran, C.A. Vaithilingam, N. Misron, K. Naidu, and M.R. Ahmed. A comprehensive review on system architecture and international standards for electric vehicle charging stations, 2021.

[49] A. Saadaoui, M. Ouassaid, and M. Maaroufi. Overview of integration of power electronic topologies and advanced control techniques of ultra-fast ev charging stations in standalone microgrids, 2023.

[50] K.L. Lim, S. Speidel, and T. Bräunl. A comparative study of ac and dc public electric vehicle charging station usage in western australia, 2022.

[51] A. Mahdy, H.M. Hasanien, R.A. Turky, and S.H.E. Abdel Aleem. Modeling and optimal operation of hybrid wave energy and pv system feeding supercharging stations based on golden jackal optimal control strategy, 2023.

[52] J.Y. Yong, W.S. Tan, M. Khorasany, and R. Razzaghi. Electric vehicles destination charging: An overview of charging tariffs, business models and coordination strategies, 2023.

[53] S. Park and S. Choi. Integrating vehicle-to-grid technologies in autonomous electric vehicle systems, 2022.

[54] K. Taghizad-Tavana, A. Alizadeh, M. Ghanbari-Ghalehjoughi, and S. Nojavan. A comprehensive review of electric vehicles in energy systems: Integration with renewable energy sources, charging levels, different types, and standards, 2023.

[55] Z. Huang, Z. Guo, P. Ma, M. Wang, Y. Long, and M. Zhang. Economic-environmental scheduling of microgrid considering v2g-enabled electric vehicles integration, 2022.

[56] S. Schmidt. Use of battery swapping for improving environmental balance and price-performance ratio of electric vehicles, 2021.

[57] X. Sun, Z. Li, X. Wang, and C. Li. Technology development of electric vehicles: A review, 2023.

[58] Office of Scientific and Technical Information. Grid integration of electric vehicles, 2025.

[59] Green Car Congress. Economic impact of smart ev charging, 2025.

[60] Xin Yong and Wei Lin. Review on smart charging and grid challenges in electric vehicle infrastructure, 2022. Accessed: 2025-06-11.

[61] National Association of Regulatory Utility Commissioners. Ev charging and utility rate design, 2025.

[62] Global Market Insights. Electric taxi market size, share and growth report 2023-2032, 2024. Accessed: 2025-06-11.

[63] Tian Lei, Shuocheng Guo, Xinwu Qian, and Lei Gong. Understanding charging dynamics of fully-electrified taxi services using large-scale trajectory data, 2021.

[64] Market.us. Global ev taxi market size, share, demand with a cagr of 12.8%. Online report, 2023.

[65] Hong Tao, Yan Li, and Jin Feng. Battery swapping vs. plug-in charging: Economic and environmental trade-offs in urban electric taxi fleets, 2024.

[66] Yujing Chen, Lin Zhang, and Hao Wu. Evaluating the environmental benefits of electric taxis under different grid scenarios: A case study of beijing and hong kong, 2022.

[67] Zhu Lei, Li Wang, and Ming Zhang. Charging behavior analysis and infrastructure planning for electric taxis in china: A case study of beijing and shenzhen, 2021.

[68] Benjamin Rivière and Soon-Jo Chung. H-td2: Hybrid temporal difference learning for adaptive urban taxi dispatch. *arXiv preprint arXiv:2105.02138*, May 2021. Submitted on 5 May 2021.

[69] Takuma Oda and Carlee Joe-Wong. MOVI: A model-free approach to dynamic fleet management. *arXiv preprint arXiv:1804.05066*, April 2018. Submitted on 13 Apr 2018.

[70] Yongsheng Cao, Hao Wang, Demin Li, and Guanglin Zhang. Smart online charging algorithm for electric vehicles via customized actor-critic learning. *arXiv preprint arXiv:2106.00447*, June 2021. Submitted on 1 Jun 2021.

[71] Ming Zhu, Xiao-Yang Liu, and Xiaodong Wang. Joint transportation and charging scheduling in public vehicle systems – a game theoretic approach, 2017. 13 pages. Categories: Systems and Control (cs.SY).

[72] Jingyi Zhang, Wenpeng Jing, Zhaoming Lu, Haotian Wu, and Xiangming Wen. Collaborative strategy for electric vehicle charging scheduling and route planning. *IET Smart Transportation Systems*, 2024. First published: 26 April 2024.

[73] Tai-Yu Ma, Richard D. Connors, and Francesco Viti. Coordinated vehicle dispatching and charging scheduling for an electric ride-hailing fleet under charging congestion and dynamic prices, 2024. Categories: Optimization and Control (math.OC); Systems and Control (eess.SY).

[74] Husam I. Shaheen, Ghamgeen I. Rashed, Bo Yang, and Jun Yang. Optimal electric vehicle charging and discharging scheduling using metaheuristic algorithms: V2g approach for cost reduction and grid support. *Journal of Energy Storage*, 78:111816, 2024.

[75] Nattavit Piamvilai and Somporn Sirisumrannukul. Optimal scheduling of movable electric vehicle loads using generation of charging event matrices, queuing management, and genetic algorithm. *Energies*, 15(10):3827, 2022.

[76] Dimitrios Efthymiou, Katerina Chrysostomou, Maria Morfoulaki, and Georgia Aifantopoulou. Electric vehicles charging infrastructure location: A genetic algorithm approach. *European Transport Research Review*, 9(1):27, 2017. Published online: 5 May 2017.

[77] A. Author. A new perspective on v2g. *ScienceDirect*, 2024.

[78] Todd Litman. Autonomous vehicle implementation predictions, 2019.

[79] Tao Chen et al. Intelligent transportation systems and artificial intelligence: A review, 2021.

[80] Christopher JCH Watkins and Peter Dayan. Q-learning, 1992.

[81] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, et al. Human-level control through deep reinforcement learning, 2015.

[82] Xuesong Li et al. Efficient and fair dynamic assignment for ridesourcing platforms. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 5971–5978, 2019.

[83] Hado Van Hasselt, Arthur Guez, and David Silver. Deep reinforcement learning with double q-learning, 2016.

[84] Long-Ji Lin. Self-improving reactive agents based on reinforcement learning, planning and teaching, 1992.

[85] Richard S Sutton and Andrew G Barto. *Reinforcement Learning: An Introduction.* MIT press, 2018.