AI-Supported Explanation of Optimization Decision in Smart Charging of Electric Commercial Vehicles

Bachelor thesis in the department of Mechanical Engineering by Aziz Abouda
Date of submission: April 07, 2026

1. Review: Paul Heckelmann
2. Review: Tim Schneider
Darmstadt

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Mechanical Engineering
Department
IMS

Thesis Statement pursuant to §22 paragraph 7 of APB TU Darmstadt

I herewith formally declare that I, Aziz Abouda, have written the submitted thesis independently. I did not use any outside support except for the quoted literature and other sources mentioned in the paper. I clearly marked and separately listed all of the literature and all of the other sources which I employed when producing this academic work, either literally or in content. This thesis has not been handed in or published before in the same or similar form.

I am aware, that in case of an attempt at deception based on plagiarism (§38 Abs. 2 APB), the thesis would be graded with 5,0 and counted as one failed examination attempt. The thesis may only be repeated once.

In the submitted thesis the written copies and the electronic version for archiving are identical in content.

Darmstadt/Germany, 06.04.2026

Aziz Abouda

Contents

1 Introduction

1.1 Background and Motivation

The ongoing energy transition is fundamentally transforming the structure of modern energy systems. With the increasing integration of renewable energy sources such as wind and solar power, electricity generation is becoming more volatile and less predictable. As a result, balancing electricity supply and demand has become a significant challenge for grid operators. One promising approach to address this challenge is the flexibilization of electricity consumers, allowing energy demand to adapt dynamically to changes in electricity generation.

In this context, the electrification of the transport sector plays an increasingly important role. The number of electric vehicles (EV)(s) is steadily growing, not only in private transportation but also in commercial fleets such as buses, delivery vehicles, and logistics trucks. These electric commercial vehicles require regular charging and therefore represent a considerable electricity demand within the power system. If managed effectively, the charging processes of these vehicles can be shifted in time to better align with electricity market conditions and the available capacity of the power system.

Smart charging systems have been developed to optimize the charging behavior of electric vehicles. These systems typically rely on algorithmic approaches to determine when and how much each vehicle should charge within a given time horizon, taking into account factors such as electricity prices, vehicle availability, battery state of charge, grid capacity limits, and operational schedules. By considering these parameters, smart charging systems can reduce operational costs and improve the overall efficiency of energy usage.

1.2 Problem Statement

A central challenge in the practical use of smart charging systems lies in the limited interpretability of their decision-making processes. While such systems generate concrete charging actions, they do not sufficiently reveal the reasoning behind these decisions. As a result, operators are often unable to understand how and why specific charging strategies are derived.

This lack of insight becomes particularly problematic in operational contexts, where users are responsible for monitoring system behavior and ensuring that decisions align with practical requirements. Without a clear understanding of the reasoning behind charging actions, operators cannot reliably assess whether decisions are appropriate, nor can they confidently justify them.

In practice, this can lead to situations where system behavior appears unintuitive. For example, vehicles may not start charging immediately after arriving at a depot, charging may be postponed despite available capacity, or charging may occur during periods that seem unusual from an operational perspective. Although such decisions may be valid within the system, their underlying rationale is not readily accessible to users.

This lack of transparency may reduce the usability and acceptance of smart charging systems. If operators cannot interpret the reasoning behind system behavior, it becomes more difficult to verify decisions, detect potential issues, and build trust in automated processes.

Therefore, there is a need to explain why specific charging decisions are made by the optimization model. In particular, methods are required that can interpret complex optimization outputs and provide explanations that are accessible to human operators, supporting them in understanding and validating system behavior.

1.3 Research Objectives and Research Questions

Based on the problem described in the previous section, the objective of this thesis is to develop and evaluate an AI-based method that generates understandable explanations for optimization decisions in smart charging systems for electric commercial vehicles. In order to structure this investigation, the following research question and sub-questions are defined.

Main Research Question

How can AI-based methods be used to generate understandable explanations for optimization decisions in smart charging systems for electric commercial vehicles?

Sub Research Questions

- **RQ1:** Which factors influence charging decisions in optimization-based smart charging systems?
- **RQ2:** How can optimization results and contextual system information be transformed into human-understandable explanations using AI-based techniques?
- **RQ3:** How understandable and useful are the generated explanations from the perspective of users?

2 Theoretical Foundations and Related Work

This chapter provides the theoretical foundations and related work relevant to this thesis. First, the concept of smart charging for electric vehicles and the role of optimization in determining charging decisions are introduced. Subsequently, the field of explainability in AI-based and optimization-based decision systems is examined in depth, covering key methods, human-centered perspectives, and trust-related aspects. Finally, different approaches for generating explanations of algorithmic decisions are discussed, with a focus on their applicability to optimization-based systems.

2.1 Smart Charging of Electric Vehicles

The increasing adoption of electric vehicles introduces significant challenges for power systems, particularly when charging processes are not coordinated. Uncontrolled charging behavior can lead to peak load amplification, voltage instability, and overloading of grid components, thereby reducing grid reliability and increasing operational risks [6]. As the penetration of electric vehicles continues to grow, these effects become more pronounced, making the management of charging demand a critical issue for modern power systems [32].

Smart charging has emerged as a key solution to mitigate these challenges by enabling controlled and adaptive charging processes. Instead of charging vehicles immediately at maximum power, smart charging allows shifting charging activities over time based on grid conditions, electricity prices, and user requirements [24]. This temporal flexibility enables better alignment between electricity demand and supply, particularly in systems with increasing shares of renewable energy and fluctuating generation patterns [21].

A central idea behind smart charging is that charging demand can be treated as a controllable and flexible load. By dynamically adjusting charging power, electric vehicles can support grid operations by contributing to peak shaving, load balancing, and improved utilization of existing infrastructure [20]. This transforms electric vehicles from passive consumers into active participants in the energy system, enabling new forms of interaction between mobility and energy sectors.

From an economic perspective, smart charging enables cost optimization by exploiting temporal variations in electricity prices. Since electricity prices are typically lower during off-peak periods, shifting charging demand to these periods can significantly reduce operational costs for both users and system operators [24]. This introduces an inherent trade-off between immediate charging convenience and economic efficiency, as users may need to delay charging to benefit from lower costs.

To realize these benefits, smart charging systems require coordination mechanisms that manage interactions between multiple vehicles, charging infrastructure, and the power grid. Charging decisions cannot be made independently, as the aggregated demand of many vehicles directly affects grid conditions [7]. Therefore, coordinated control strategies are necessary to ensure that individual charging requirements are satisfied while maintaining overall system stability.

In practice, smart charging approaches differ in their level of sophistication. Simple rule-based strategies rely on predefined heuristics, such as restricting charging to off-peak hours, and are easy to implement but lack adaptability to dynamic system conditions. More advanced approaches formulate the charging process as an optimization problem, where charging schedules are computed by considering multiple objectives and constraints simultaneously [11]. These approaches enable more efficient and flexible solutions but require accurate system models and computational resources.

Optimization-based methods are particularly powerful because they can capture the interdependencies between different system components and time steps. However, they also introduce significant complexity, as charging decisions are influenced by a combination of technical constraints, economic signals, and operational requirements [32]. This complexity increases further in large-scale systems with many vehicles and heterogeneous user behavior.

A particularly relevant context for this thesis is the charging of commercial electric vehicle fleets operating from a shared depot. Unlike private vehicle charging, fleet charging is characterized by predictable route schedules, fixed return times, and centralized energy management. These properties make fleet charging well suited to optimization-based control, as departure times and minimum required state-of-charge levels are known in advance. At the same time, the shared nature of depot infrastructure introduces coupling constraints across vehicles: the transformer capacity limit applies to the aggregate charging power of the entire fleet, meaning that decisions for individual vehicles cannot be made in isolation. This interdependency between vehicles significantly increases the complexity of the scheduling problem and, as a consequence, makes the resulting decisions more difficult to interpret [28].

The typical formulation used for depot-level fleet charging is a Mixed-Integer Linear Program (MILP), in which binary decision variables enforce mutually exclusive operating modes such as charging and discharging, while continuous variables represent power flows and energy levels. The MILP structure allows the model to simultaneously satisfy vehicle-level constraints, such as minimum state-of-charge requirements by departure, and system-level constraints, such as transformer capacity limits, while optimizing an economic objective over the full planning horizon [33]. This intertemporal coupling — the fact that each time step's decision affects all future states through the state-of-charge dynamics — is a key source of complexity and a primary reason why individual decisions can appear unintuitive to human operators.

Another important challenge in smart charging is the presence of uncertainty. Charging demand depends on stochastic factors such as vehicle arrival and departure times, user preferences, and state-of-charge levels, which are difficult to predict accurately [24]. In addition, external factors such as fluctuating electricity prices and renewable energy generation introduce further variability into the system. The occurrence of negative electricity prices — situations where excess renewable generation causes market prices to drop below zero — represents a particularly counterintuitive condition that requires explicit handling in the optimization model and in the subsequent explanation of decisions.

Overall, smart charging represents a complex, multi-objective decision-making problem that involves balancing grid stability, economic efficiency, and user requirements. The resulting charging schedules emerge from the interaction of numerous dynamic factors, making the underlying decision-making process difficult to interpret. This complexity highlights the need for approaches that improve the transparency and explainability of smart charging decisions.

2.2 Explainability in AI-Based Decision Systems

Modern AI-based decision systems often operate as complex computational models whose internal reasoning is not directly accessible to users. Such systems are therefore frequently perceived as black boxes, where the relationship between inputs and outputs remains opaque. This lack of transparency reduces trust and limits the practical usability of these systems, especially in high-stakes domains [5].

This challenge is not limited to machine learning models but also applies to optimization-based systems. Although optimization models are mathematically defined, their decisions often emerge from complex interactions between multiple variables, constraints, and time-dependent relationships. As a result, users may still struggle to understand why a specific decision has been made.

Dimensions of Explainability Explainability can be analyzed along different dimensions depending on how explanations are generated and what aspect of the system they target. A common distinction is between global and local explainability.

Global explainability aims to describe the overall behavior of a system by analyzing general relationships between inputs and outputs. In contrast, local explainability focuses on explaining individual decisions by examining the specific conditions under which they occur. These two perspectives are complementary, as global explanations provide an overview of system behavior, while local explanations enable detailed interpretation of specific decisions [5]. In the context of this thesis, local explainability is the primary focus, as fleet operators are typically interested in understanding why a specific vehicle is or is not charging at a particular point in time, rather than understanding the general properties of the optimization model.

Categories of Explainability Methods Existing literature classifies explainability approaches into several categories based on their methodological characteristics. These include inherently interpretable models, model-agnostic methods such as LIME or SHAP, example-based explanations, model-specific techniques, and visualization-based approaches [5].

Among model-agnostic methods, SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) are the most widely used [18, 23]. SHAP assigns a contribution value to each input feature based on the Shapley value concept from cooperative game theory, distributing the output prediction across all features in a theoretically consistent way. LIME, by contrast, approximates the behavior of a complex model in the neighborhood of a specific input by fitting a simpler, locally linear surrogate model. Both methods produce feature-level importance scores that indicate which inputs most strongly influenced a given output [25].

However, applying SHAP or LIME directly to optimization-based decision systems raises fundamental conceptual difficulties. Both methods treat the underlying system as a black-box function and perturb its inputs to infer feature relevance. In optimization models, however, the decision is not a statistical prediction but the outcome of constraint satisfaction and objective minimization across coupled time steps. Perturbing a single input — for example, an electricity price at one time step — does not yield a meaningful local approximation of the decision function, because the intertemporal dependencies between time steps mean that the optimization response to a perturbation cannot be linearized in any principled way. As a result, SHAP and LIME explanations applied to optimization outputs may produce misleading attributions that do not reflect the true causal structure of the decision [25]. This limitation motivates the need for explanation approaches that are specifically designed for optimization contexts, rather than adapted from machine learning explainability.

Counterfactual and Contrastive Explanations A particularly important class of explanations for decision-support systems is counterfactual and contrastive explanations. A counterfactual explanation answers the question: what is the minimum change to the system inputs or conditions that would have led to a different decision? A contrastive explanation answers the related question: why was decision A made rather than decision B? Both forms of explanation are well aligned with the natural questions posed by human operators, such as “why is this vehicle not charging right now?” or “under what conditions would it charge earlier?” [31].

Counterfactual explanations have received substantial attention in the machine learning literature. Verma et al. [30] provide a comprehensive review of counterfactual methods, noting that they are particularly effective because they are contrastive — they focus on differences rather than exhaustive feature lists — and because they are inherently action-oriented, pointing toward conditions under which an alternative outcome would occur. In the context of optimization, Schleich et al. [15] extend the counterfactual framework to linear optimization problems, proposing formal definitions of counterfactual explanations based on parameter changes in the objective function or constraints. Their work demonstrates that in linear programs, such explanations can often be computed efficiently, making them practically relevant for real-time decision-support. Although this thesis does not implement counterfactual explanations directly, the concept informs the design of the explanation pipeline: the system is designed to identify the conditions that are actively shaping the decision, which is functionally equivalent to identifying the active constraints or binding price signals that, if changed, would produce a different outcome.

Multi-Level Perspective on Explainability A more comprehensive view of explainability distinguishes between different functional levels that address distinct user needs. A commonly used framework separates explainability into three levels.

The first level focuses on model reasoning, addressing the question of which input factors influenced a decision. This includes techniques such as feature attribution and dependency analysis.

The second level concerns cognitive explanations, which aim to explain why a decision occurred. This involves causal reasoning, contrastive explanations, and counterfactual analysis.

The third level relates to explanation delivery, which focuses on how explanations are communicated to users. This includes visualizations, interactive interfaces, and natural language explanations.

Together, these levels correspond to key user questions, namely what influenced a decision, why it occurred, and how it can be understood [2]. Miller further grounds this framework in insights from cognitive science and philosophy, arguing that human explanations are inherently selective, contrastive, and social [19]. People do not expect complete causal accounts; they expect focused answers to implicit comparison questions. This observation has direct implications for the design of automated explanation systems: an explanation that lists all relevant factors is not necessarily more useful than one that highlights the two or three factors that most strongly differentiate the actual decision from the most plausible alternative.

Explainability in Energy Systems The application of explainability methods to energy management systems has received increasing attention in recent years. Most existing work focuses on explaining predictive models, such as load forecasting or fault detection classifiers, rather than optimization decisions. Faiq et al. [29] propose EI-Build, an explainable AI framework for building energy management systems that integrates multiple XAI techniques including SHAP, LIME, and counterfactual analysis within a unified architecture. Their work concludes that while these methods offer complementary perspectives, a key limitation common to all of them is the inherent trade-off between fidelity and simplicity: detailed explanations improve precision but can overwhelm non-specialist users, while simpler summaries risk omitting critical information [29]. The authors further note that tools such as SHAP and LIME rely on surrogate approximations that may diverge from the actual model logic in complex systems, reducing their operational value in real-world energy management contexts. Similarly, Altiary and Hussain [2] review explainability approaches in the context of carbon price forecasting, identifying the lack of causal and contrastive explanation capabilities as a key gap in existing methods. Importantly, however, the explanation of scheduling and optimization decisions in energy systems

— as opposed to the explanation of predictive models — remains an underexplored area. This gap directly motivates the work presented in this thesis.

Human-Centered Explainability and Trust A growing body of research argues that the technical quality of an explanation is insufficient on its own. Explanations must also be evaluated in terms of how human users actually perceive and use them [14]. This human-centered perspective distinguishes between multiple dimensions of explanation quality from the user's viewpoint, including clarity, completeness, usefulness for decision-making, and the degree to which the explanation increases or appropriately calibrates trust in the system.

The relationship between explainability and trust is complex. Afroogh et al. [1] identify transparency, performance reliability, and alignment with user expectations as key components of trustworthy AI systems, noting that trust can be undermined both by opaque systems and by systems that provide explanations users cannot verify. Perrig [27] further distinguishes between subjective trust — measured via questionnaires — and objective trust-related behavior such as reliance on system recommendations, finding that empirical evidence on whether explanations reliably increase trust is mixed. These findings suggest that explanation quality cannot be reduced to a single dimension and that evaluations must account for both the technical accuracy of explanations and their perceived usefulness from the user's perspective.

A particularly relevant concept in this context is overtrust, which occurs when users accept system decisions without critical scrutiny because the explanation appears convincing, even if it is incomplete or imprecise. In automated decision-support systems, this risk is especially pronounced when explanations are generated by language models, which may produce fluent and confident-sounding text regardless of its factual grounding [1]. The design of the explanation pipeline in this thesis addresses this risk by grounding all explanations in structured, verified data derived directly from the optimization output, rather than relying on the language model to infer relevant factors independently.

Miller [19] also emphasizes that explanations are fundamentally social artifacts: they are produced in a communicative context and must be adapted to the knowledge, expectations, and goals of the intended audience. This implies that a single explanation format is unlikely to serve all users equally well, and that explanation systems should ideally support different levels of detail and technical language depending on the user's background. While the current system does not yet implement adaptive explanations, this design principle informed the evaluation methodology described in Chapter 5.

Limitations of Existing Approaches Despite significant progress, current explainability methods exhibit several limitations. First, most approaches focus on feature attribution while neglecting deeper reasoning mechanisms such as causal or counterfactual explanations. As a result, higher-level explainability requirements are often not addressed [2].

Second, many systems fail to provide user-centered explanations. Explanations are frequently presented as technical outputs rather than meaningful insights tailored to user needs [19]. Third, there is a lack of systematic evaluation of explanation quality, meaning that it is often unclear whether explanations are actually useful for decision-making [14].

Finally, a key limitation is the gap between technical system behavior and human understanding, which continues to hinder the practical adoption of explainable systems. This gap is particularly pronounced in optimization-based systems, where decisions emerge from mathematical constraints that have no direct counterpart in everyday reasoning.

Relevance for Optimization-Based Smart Charging These limitations are particularly relevant for optimization-based systems such as smart charging. In such systems, decisions result from complex interactions between electricity prices, operational constraints, and scheduling requirements. Although these systems are mathematically grounded, their outputs are often not directly interpretable for users.

Therefore, there is a need for approaches that go beyond traditional feature attribution and provide human-centered explanations of optimization decisions. The explanation approach proposed in this thesis directly addresses this need by combining structured data extraction, constraint analysis, and natural language generation into a pipeline that produces locally grounded, human-readable accounts of individual charging decisions.

2.3 Approaches for Generating Explanations

To address the limited transparency of smart charging systems, different approaches can be used to generate human-understandable explanations of algorithmic decisions. These approaches aim to transform structured system outputs, such as charging schedules, constraints, and system states, into natural language descriptions that can be interpreted by human operators.

It is important to distinguish this task from explainable artificial intelligence (XAI). XAI primarily focuses on explaining the behavior of machine learning models. In contrast, this thesis does not aim to explain an AI model itself, but rather to explain the decisions of an optimization-based system using AI techniques. The focus is therefore on making the reasoning behind algorithmic decisions accessible, rather than interpreting the internal structure of a predictive model. This distinction has important methodological consequences: whereas XAI methods can be evaluated for fidelity — how well they approximate the model they explain — optimization explanation systems should be evaluated for correctness, completeness, and comprehensibility, as the ground truth is defined by the mathematical structure of the optimization model rather than by a learned function [19].

This task is closely related to the field of Natural Language Generation (NLG), which focuses on transforming structured data into human-readable text. NLG systems take non-linguistic input and generate coherent textual output based on linguistic and domain knowledge [22]. However, in the context of this thesis, explanation generation goes beyond pure text generation. While NLG defines how information can be expressed, it does not determine which information should be communicated or why a specific decision was made. Therefore, an additional step is required to identify the relevant factors influencing a decision before generating the explanation.

Several approaches can be considered for generating explanations, differing in how they balance control, flexibility, and complexity.

2.3.1 Template-Based Explanation Generation

A first approach is template-based explanation generation, in which predefined sentence structures are filled with values derived from the system output. In this approach, explanations are generated by mapping structured data, such as decision variables, constraints, and system states, to linguistic templates that define how the information is expressed in natural language.

From a methodological perspective, template-based systems follow a deterministic pipeline consisting of content selection and surface realization [9]. First, relevant information is extracted from the system output, for example electricity prices, charging schedules, or constraint satisfaction. Second, this information is inserted into predefined sentence templates, where placeholders are replaced with concrete values. For instance, a template of the form

> "The vehicle is charged after {time} because electricity prices are {price_trend} and the departure constraint is satisfied"

can be instantiated using values obtained from the optimization model.

Template-based approaches have been widely used in explanation systems due to their simplicity, transparency, and reliability, as they ensure that generated explanations remain factually consistent with the underlying data [17, 26]. Since both the structure of the explanation and the inserted values are explicitly defined, the resulting output is fully deterministic and directly traceable to the system state. This property is particularly valuable in contexts where explanations must be verifiable and auditable, as every statement in the output can be traced back to a specific data value or rule [3].

In addition, this approach allows domain knowledge to be explicitly encoded into the explanation process. For example, rules can be defined to prioritize certain factors, such as cost minimization or constraint satisfaction, ensuring that explanations reflect the intended interpretation of the system behavior.

However, since templates are predefined and fixed, this approach suffers from limited flexibility and expressiveness [17, 4]. It is difficult to represent more complex relationships between multiple interacting factors, especially in scenarios where decisions result from trade-offs between competing objectives. As a result, explanations may oversimplify the underlying reasoning or fail to capture important dependencies between system variables.

Furthermore, as system complexity increases, the number of required templates grows significantly, which reduces scalability and makes the system harder to maintain [9]. The generated explanations may also become repetitive and overly rigid, as they are restricted to predefined sentence structures [4]. For optimization systems with many interacting constraints and time-dependent variables, the combinatorial space of possible decision contexts makes full template coverage practically infeasible.

2.3.2 Prompt-Based Large Language Models

Another approach is the use of prompt-based large language models (LLMs). In this setting, the output of the optimization model is provided to an LLM together with an instruction describing the explanation task, allowing the model to generate a natural-language explanation. This approach is appealing because it enables fluent, human-readable, and context-sensitive explanations without requiring explicitly defined rules or templates. In particular, LLMs are capable of combining multiple influencing factors, such as electricity prices, operational constraints, and temporal dependencies, into a single coherent explanation, thereby improving the accessibility of complex system behavior for human users.

However, when applied directly to optimization outputs, several fundamental limitations arise. Optimization models typically produce large volumes of high-dimensional, time-dependent data, including decision variables, constraints, and system states across multiple time steps and entities. Such outputs are not inherently structured for linguistic interpretation and often exceed the input capacity of LLMs. More importantly, they do not explicitly distinguish between relevant and irrelevant information. As a result, when raw optimization data is provided as input, the model must implicitly infer which aspects are important for explaining a decision. This implicit selection process is not guided by the actual optimization logic and can therefore lead to incomplete, inconsistent, or misleading explanations.

A key limitation of LLM-based approaches is the lack of grounding in the underlying optimization model. While optimization models rely on well-defined mathematical relationships and constraints, LLMs generate text based on statistical correlations learned from training data [12]. Consequently, the generated explanations are not guaranteed to faithfully represent the true reasoning process of the system. This discrepancy is particularly problematic in technical decision-support contexts, where explanations are expected to accurately reflect the factors that influenced a decision.

Closely related to this issue is the phenomenon of hallucination, a known limitation of large language models. Hallucinations occur when the model generates plausible-sounding but factually incorrect or unsupported statements [13]. In the context of optimization-based smart charging, this may result in explanations that incorrectly attribute decisions to certain factors, such as electricity prices, while omitting or misrepresenting other critical aspects, including grid constraints, vehicle availability, or scheduling dependencies. Huang et al. [12] categorize hallucinations into intrinsic errors, where the generated output contradicts the provided context, and extrinsic errors, where the output introduces information that cannot be verified from the available input. Both types are particularly dangerous in explanation systems, since users may not be able to detect factual errors in technically complex outputs. Since LLMs prioritize linguistic coherence over factual correctness, such errors may not be immediately detectable by users, thereby increasing the risk of misinterpretation and overtrust [1].

Another important limitation concerns the controllability and reproducibility of the generation process. Unlike template-based approaches, LLM outputs are not fully deterministic. Even when the same input and prompt are used repeatedly, the model may generate different explanations for the same underlying decision. This inherent variability is a consequence of the probabilistic nature of the generation process and cannot be fully eliminated.

In addition, small variations in input representation, prompt formulation, or model parameters can further amplify these differences, leading to significantly different explanations for identical system states. This lack of stability makes it difficult to ensure that explanations consistently include all relevant decision factors and adhere to predefined quality criteria. As a result, evaluating and validating LLM-generated explanations becomes challenging, particularly in scenarios where explanations must meet strict requirements for correctness, completeness, and reproducibility [14].

Furthermore, the quality of generated explanations is strongly dependent on the training data of the model. Since LLMs are typically trained on large-scale and heterogeneous data sources, they may reflect general language patterns, biases, or dominant narratives that are not specific to the system being explained. This can lead to explanations that are linguistically plausible but insufficiently aligned with the actual system behavior.

Overall, while prompt-based LLMs provide a flexible and powerful mechanism for generating natural-language explanations, their direct application to raw optimization outputs lacks reliability, transparency, and control. These limitations indicate that LLMs should not be used in isolation, but rather integrated into structured explanation pipelines that include preprocessing, context selection, and validation steps. Such an approach enables the extraction of relevant information from complex optimization results and provides a grounded basis for explanation generation. This requirement directly motivates the approach proposed in this thesis, which aims to transform optimization outputs into structured, interpretable representations before generating explanations.

2.3.3 Fine-Tuning Language Models on Domain-Specific Data

A further approach is fine-tuning language models on domain-specific datasets. In this setting, a pre-trained language model is trained on examples consisting of structured system data and corresponding natural-language explanations. The model learns to generate explanations by identifying patterns between the input data and the associated textual descriptions.

This enables more flexible and natural explanations compared to template-based approaches. In particular, the model can combine multiple influencing factors and adapt to domain-specific terminology. For example, a fine-tuned model could generate:

> "Charging is scheduled during the low-price period to minimize costs while meeting the required state of charge by 06:00."

However, this approach has several important limitations. A key issue is that fine-tuning does not reliably teach the model new knowledge. Instead, models primarily rely on knowledge acquired during pre-training and struggle to integrate new information introduced during fine-tuning. Empirical results show that examples containing new knowledge are learned significantly more slowly than those aligned with the model's existing knowledge [10].

Furthermore, introducing new information during fine-tuning can negatively affect the model's behavior. The study demonstrates that learning such new knowledge is associated with an increased tendency to produce hallucinations, meaning outputs that are not grounded in the model's actual knowledge [10]. In the context of explanation generation, this implies that the model may generate plausible-sounding explanations that are not fully consistent with the underlying optimization logic.

Another limitation is that fine-tuning can degrade overall performance when the model overfits to the training data. The results show that as the model increasingly fits examples containing unfamiliar information, its performance decreases, indicating reduced reliability [10]. This behavior highlights that the learning process is not only limited but can also introduce unintended side effects.

In addition, the generation process remains difficult to control. Since the model learns implicit patterns rather than explicit rules, it cannot be guaranteed that all relevant decision factors are consistently included in the generated explanations. This lack of controllability is particularly problematic in optimization contexts, where explanations must accurately reflect constraints, objectives, and trade-offs [19].

Overall, while fine-tuning enables more natural and flexible explanations, it does not provide the level of transparency, reliability, and control required for explaining optimization-based decisions. These limitations motivate the need for approaches that combine structured reasoning with language generation rather than relying solely on fine-tuned models.

2.3.4 Retrieval-Augmented Generation

A fourth approach, and the one adopted in this thesis, is Retrieval-Augmented Generation (RAG). RAG was introduced by Lewis et al. [16] as a framework that enhances language model generation by first retrieving relevant documents or data from an external knowledge base and then conditioning the generation process on the retrieved content. The key architectural distinction from pure prompt-based generation is that the model does not have to rely on its parametric knowledge alone; instead, it receives explicitly selected, context-specific information that directly informs the output.

In the context of explanation generation for optimization decisions, RAG offers several important advantages. Rather than providing the full optimization output to the language model — which would be high-dimensional, noisy, and exceed the model's effective context window — the pipeline selects a focused subset of relevant information based on the decision being explained. This retrieved context includes electricity price trajectories, vehicle availability schedules, constraint activity indicators, and state-of-charge dynamics over a temporal window centered on the time step of interest. By restricting the model's input to this curated subset, the explanation remains grounded in verifiable data and is less susceptible to hallucination [16, 8].

Gao et al. [8] provide a comprehensive survey of RAG architectures, distinguishing between naive RAG, which simply concatenates retrieved passages with the user query, and advanced RAG, which incorporates additional preprocessing, reranking, and context compression steps to improve retrieval quality. The approach in this thesis corresponds to advanced RAG: the retrieval step is not a simple keyword search but a structured extraction process that reconstructs the relevant decision context from multiple data sources before passing it to the language model.

An important limitation of RAG must be acknowledged: retrieval-augmented generation does not fully eliminate hallucination. As noted in the literature, LLMs can still misinterpret, overgeneralize, or fabricate information even when accurate context is provided [12]. This residual risk is addressed in the proposed pipeline through prompt engineering constraints that explicitly instruct the model to avoid unsupported assumptions and to base all claims on the provided data. The combination of structured retrieval and constrained generation thus provides a more reliable explanation mechanism than either approach would achieve in isolation.

2.3.5 Comparison of Approaches

These approaches differ in how they balance control, flexibility, and reliability. The approach selected in this thesis: a RAG-based pipeline combining structured preprocessing, context retrieval, and LLM-based generation, is positioned to overcome the key limitations of each individual method: it avoids the rigidity of templates, the grounding failures of raw LLM prompting, and the controllability issues of fine-tuned models. The approach is presented in detail in Chapter 4.

The main characteristics and limitations of the discussed approaches are summarized in Table 1.

Table 1: Comparison of explanation approaches

Approach	**Strengths**	**Limitations**	**Suitability for This Work**
Template-based	High control, fully deterministic, factually accurate	Rigid structure, limited expressiveness, cannot capture complex relationships	Too limited for explaining complex optimization decisions
LLM-based (prompt)	Fluent, natural language, handles complex context	Risk of hallucination, low controllability, inconsistent outputs	Lacks reliability for critical decision explanations
Fine-tuned models	Consistent outputs, domain-specific language, more stable than generic LLMs	Limited transparency, difficult to control content, hard to debug or validate	Limited interpretability and control over explanation content
RAG-based pipeline	Grounded in retrieved data, reduces hallucination, flexible generation	Retrieval quality affects output, residual hallucination risk	Selected approach: combines structured grounding with natural language generation

3 System Description and Optimization Model

This chapter provides a detailed description of the considered charging system and the underlying optimization model. In contrast to a purely abstract formulation, the focus is on explaining how the model operates, how decisions are derived, and how different system components interact. The presented model is implemented using the Pyomo optimization framework and represents a depot-level energy management system for electric commercial vehicles.

The overall workflow of the optimization model is illustrated in Figure 1.

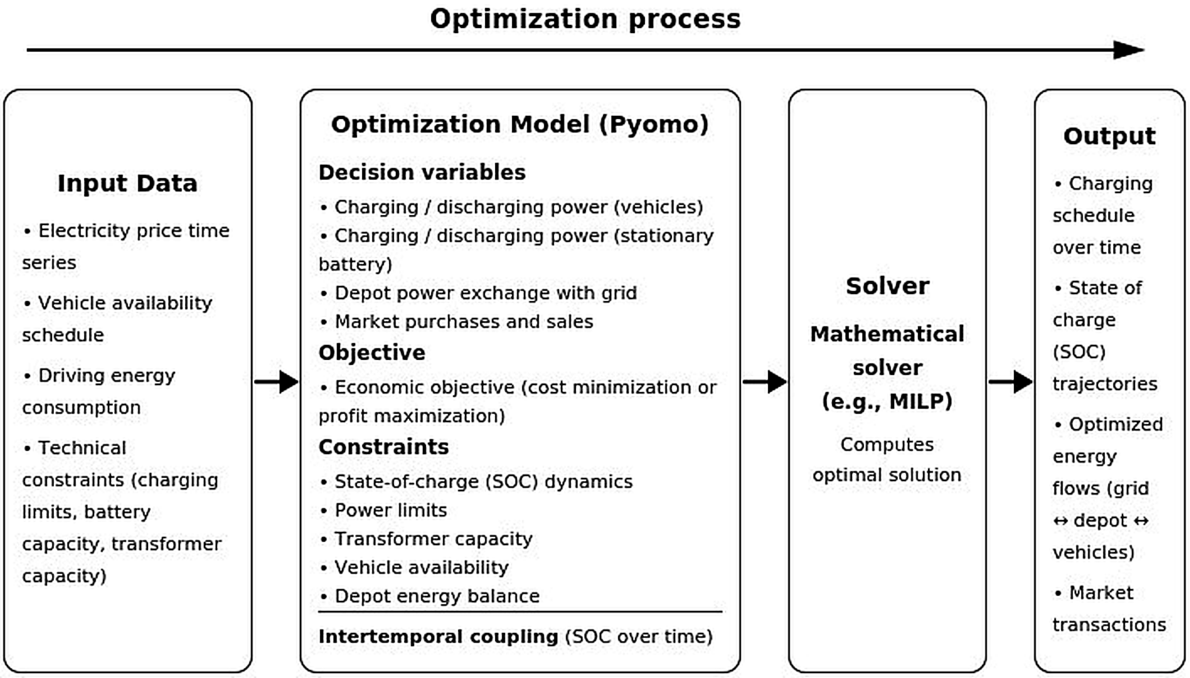

Figure 1: Overview of the smart charging optimization workflow (AI-generated)

3.1 System Description

The system represents a depot at which a fleet of electric commercial vehicles is operated. Each vehicle is equipped with a traction battery and can be connected to the charging infrastructure during the time it is physically present at the depot. Alongside the vehicle batteries, the system includes a stationary battery storage unit that can be used to buffer energy and increase operational flexibility by decoupling energy procurement from vehicle charging demand.

The depot acts as an interface between the electricity market and the internal energy consumers. Electrical energy can be purchased from different market products and is distributed among the available storage units. Depending on the scenario configuration, the system may also feed energy back into the grid through vehicle-to-grid (V2G) discharging, or participate in balancing service markets to generate additional revenue.

Time Resolution and Planning Horizon The optimization is performed over a planning horizon consisting of one or more representative days. Within each day, three time resolutions are used simultaneously to represent different market products and operational decisions:

- **15-minute resolution** (`i_96`): 96 time steps per day, used for all operational variables including vehicle charging and discharging, stationary battery operation, and intraday market products.
- **60-minute resolution** (`i_24`): 24 time steps per day, used for day-ahead electricity market products (DA60), which are traded in hourly blocks.
- **240-minute resolution** (`i_6`): 6 time steps per day, used for frequency containment reserve (FCR) capacity products, which are offered in four-hour blocks.

A combined index set `i_t3` links these three resolutions by associating each 15-minute interval with its corresponding 60-minute and 240-minute parent blocks. This multi-scale time structure enables the model to represent both fine-grained operational decisions and higher-level market commitments within a single coherent formulation, ensuring that energy quantities are consistent across all time scales.

Vehicle Characterization Each vehicle in the fleet is identified by a unique identifier from the set `i_ev` and is characterized by the following parameters:

- Battery capacity $p\_EV\_STO^{v}$ [MWh]
- Maximum charging power $p\_EV\_POW\_ch^{v}$ [MW], defaulting to 0.3 MW
- Maximum discharging power $p\_EV\_POW\_dis^{v}$ [MW], defaulting to 0.3 MW (only active when V2G mode is enabled)
- Availability parameter $p\_FL\_MAN^{v,d,t} \in \{0,1\}$, indicating whether vehicle v is present at the depot during time step t of representative day d
- Driving consumption $p\_FL\_CON^{v,d,t}$ [MWh per 15-min interval], representing the energy drawn from the battery while the vehicle is on a route

When a vehicle is not at the depot ($p\_FL\_MAN^{v,d,t} = 0$), charging and discharging power is forced to zero through big-M constraints, and the driving consumption parameter reduces the state of charge to reflect energy consumed during the route.

Stationary Battery The stationary battery is characterized by its capacity p_ST_STO [MWh], maximum charging and discharging power $p\_ST\_POW\_ch$ [MW], and efficiency parameters shared with the vehicle batteries. Its state of charge is bounded between a lower fraction $P\_ST\_SOC\_LB$ and an upper fraction $P\_ST\_SOC\_UB$ of its total capacity. Unlike vehicles, the stationary battery is always available and has no driving consumption term.

Depot Infrastructure The depot's connection to the electricity grid is characterized by a maximum power capacity $p\_DP\_POW\_ch$ [MW], representing the transformer limit. This capacity constrains the total power that can flow between the depot and the grid at any time step in either direction. Because all vehicles and the stationary battery share this single grid connection point, the transformer capacity introduces a hard coupling constraint across all charging and discharging decisions within the depot.

3.2 Optimization Model

The optimization model determines the charging and energy management strategy that maximizes economic performance while satisfying all technical and operational constraints. It is formulated as a Mixed-Integer Linear Program (MILP) using the Pyomo abstract modelling framework and is solved using the Gurobi solver. The model structure is organized into blocks (`b_rday`) over representative days, allowing the formulation to span multi-day planning horizons while maintaining intertemporal continuity between days.

Dual-Variable Energy Flow Structure

A distinctive feature of the model's implementation is the use of paired energy flow variables to represent charging and discharging efficiency losses. Rather than applying a scalar efficiency factor to a single power variable, the model defines two separate variables for each flow direction:

- $v\_EV\_ENE\_IN1^{v,t}$ and $v\_EV\_ENE\_IN2^{v,t}$: the energy entering the vehicle battery from the grid side and the battery side, respectively, during charging
- $v\_EV\_ENE\_OUT1^{v,t}$ and $v\_EV\_ENE\_OUT2^{v,t}$: the energy leaving the battery on the battery side and the grid side, respectively, during discharging

The relationship between paired variables is given by the efficiency link constraints. For charging:

$$v\_EV\_ENE\_IN1^{v,t} = v\_EV\_ENE\_IN2^{v,t} \cdot \eta_{ch} \tag{1}$$

For discharging:

$$v\_EV\_ENE\_OUT1^{v,t} \cdot \eta_{dch} = v\_EV\_ENE\_OUT2^{v,t} \tag{2}$$

where η_{ch} and η_{dch} are the charging and discharging efficiency factors taken from lookup tables (`EV_EFF_TABLE`). The grid-side variables (`IN2`, `OUT2`) represent the energy exchanged at the depot level and are used in the energy balance constraints, while the battery-side variables (`IN1`, `OUT1`) are used in the SOC update equations. An identical paired structure is applied to the stationary battery (`ST`) and the depot connection (`DP`), ensuring efficiency losses are consistently accounted for at every level of the system.

Binary Mode Variables and Mutual Exclusivity

To prevent simultaneous charging and discharging — which would be mathematically feasible but physically impossible — binary activation variables are introduced for each storage unit at each time step:

- $\delta_{ch}^{v,t} \in \{0,1\}$: equals 1 if vehicle v is in charging mode at time step t
- $\delta_{dch}^{v,t} \in \{0,1\}$: equals 1 if vehicle v is in discharging mode at time step t

Mutual exclusivity is enforced by:

$$\delta_{ch}^{v,t} + \delta_{dch}^{v,t} \leq 1 \quad \forall v \in \mathcal{V},\ t \in \mathcal{T} \tag{3}$$

The effective charging and discharging powers are then obtained by multiplying the continuous power variables with their respective binary activation variables, so that power can only flow in the active direction. Identical binary variables and mutual exclusivity constraints are applied to the stationary battery and the depot grid connection.

State-of-Charge Dynamics

For each vehicle v and each time step t, the state of charge evolves according to:

$$e_t^v = e_{t-1}^v + v\_EV\_ENE\_IN1^{v,t-1} \cdot \delta_{ch}^{v,t-1} + v\_EV\_ENE\_OUT1^{v,t-1} \cdot \delta_{dch}^{v,t-1} - p\_FL\_CON^{v,d,t-1} \tag{4}$$

The discharging variable `OUT1` is defined as non-positive in the model (bounded between $-P_{dch,max}$ and 0), so it reduces the SOC when active. The driving consumption term $p\_FL\_CON^{v,d,t}$ is non-zero only during periods when the vehicle is on a route and directly reduces the SOC independently of any control decision.

The first time step of each day is treated separately: the SOC at the beginning of a representative day is linked to the SOC at the end of the previous day through the inter-day continuity constraint described below. For the stationary battery, an analogous equation applies without the driving consumption term.

Inter-Day Continuity Constraints

When the planning horizon spans multiple representative days, the model enforces continuity constraints to ensure that the state of charge at the end of one day matches the initial state of the next day. For vehicles:

$$\begin{align} e_{t=1}^{v,d} &= e_{t=96}^{v,d-1} \tag{5} \\ &\quad + v\_EV\_ENE\_IN1^{v,t=96,d-1} \cdot \delta_{ch}^{v,t=96,d-1} \tag{6} \\ &\quad + v\_EV\_ENE\_OUT1^{v,t=96,d-1} \cdot \delta_{dch}^{v,t=96,d-1} \tag{7} \\ &\quad - p\_FL\_CON^{v,d-1,t=96} \tag{8} \end{align}$$

For the stationary battery, the continuity condition requires that the SOC at the start of each day equals the SOC at the end of the previous day. Additionally, an overall horizon boundary constraint enforces that the total stored energy across all vehicles and the stationary battery at the beginning of the first day lies within a defined range $[DP\_ENE\_SOC\_0\_lb, DP\_ENE\_SOC\_0\_ub]$ of the total system capacity, and equals the total energy stored at the end of the last day. This cyclic boundary condition prevents the optimizer from artificially depleting all storage at the end of the horizon to improve the objective value.

Market Products and Energy Transactions

The model supports participation in multiple electricity market segments, which are activated through scenario-specific condition flags in the code. The following market products are implemented:

- **Day-ahead market (DA60)**: Hourly energy blocks $v\_EL\_ENE\_DA60\_a\_PE^{d,h}$ [in 0.1 MWh units] purchased in advance. The corresponding prices $p\_EL\_ENE\_DA60\_a\_pr$ are provided as input parameters for each hour of each representative day.
- **Intraday market, 15-minute blocks (ID15)**: Two separate intraday products are modelled — an auction product (`ID15_a`) and a continuous trading product (`ID15_t`) — each with 15-minute resolution. These products allow the model to adjust its energy procurement after the day-ahead market has closed, responding to intraday price changes.
- **Frequency Containment Reserve (FCR)**: Balancing capacity offered in four-hour blocks $v^{d,b}_{BA\_CAP\_FCR\_OC}$ [MW], compensated at the FCR capacity price $p\_BA\_CAP\_FCR\_pr$. FCR participation requires the depot to hold reserved capacity available for activation, creating a competition between energy procurement and balancing service provision.

The total energy procured from the electricity market at each 15-minute time step is computed as the sum of contributions from all active market products, scaled to the appropriate time resolution. Specifically, DA60 quantities are divided by four to convert from hourly to 15-minute equivalents, while ID15 quantities are used directly.

Energy quantities in the market variables are stored as integers in units of 0.1 MWh (i.e., in steps of 100 kWh), which introduces a discretization of market transactions. This integrality requirement is enforced through the `NonNegativeIntegers` domain declaration in Pyomo and contributes to the MILP structure of the problem.

Energy Balance Constraints

Two levels of energy balance are enforced at each 15-minute time step. The first balance (c_DP_ENE_bal1) ensures that the net power exchanged between the depot and the grid equals the total energy procured from all active market products:

$$v\_DP\_ENE\_IN2^t \cdot \delta^t_{ch,DP} + v\_DP\_ENE\_OUT2^t \cdot \delta^t_{dch,DP} = e\_DP\_ENE\_EL^t \tag{9}$$

The second balance (c_DP_ENE_bal2) ensures that the net power entering or leaving the depot is fully distributed among the vehicle batteries and the stationary storage:

$$\begin{aligned}
v^t_{DP\_ENE\_IN1} \cdot \delta^t_{ch,DP} + v^t_{DP\_ENE\_OUT1} \cdot \delta^t_{dch,DP} = \sum_v \Big(& v^{v,t}_{EV\_ENE\_IN2} \cdot \delta^{v,t}_{ch} && (10)\\
& + v^{v,t}_{EV\_ENE\_OUT2} \cdot \delta^{v,t}_{dch} \Big) && (11)\\
& + v^t_{ST\_ENE\_IN2} \cdot \delta^t_{ch,ST} && (12)\\
& + v^t_{ST\_ENE\_OUT2} \cdot \delta^t_{dch,ST} && (13)
\end{aligned}$$

The distinction between the two balance levels reflects the dual-variable efficiency structure: the first balance operates on grid-side quantities after the depot transformer efficiency is applied, while the second balance distributes energy at the battery-side level after vehicle and stationary battery efficiencies are applied.

Power Capacity Constraints

The model enforces two types of power capacity limits. The first limits the total power procured from all markets to the aggregated charging capacity of available vehicles plus the stationary battery:

$$e\_DP\_CAP\_IN^t \leq \sum_v p\_FL\_MAN^{v,d,t} \cdot p\_EV\_POW\_ch^v + p\_ST\_POW\_ch \tag{14}$$

The second limits the total market power to the depot transformer capacity:

$$e\_DP\_CAP\_IN^t \leq p\_DP\_POW\_ch \tag{15}$$

Both constraints apply symmetrically to power outflows when energy selling is enabled. Constraint 15 is particularly significant for explanation purposes: when it is binding, the depot cannot charge all available vehicles simultaneously, forcing the optimizer to prioritize some vehicles over others based on economic and operational criteria that may not be immediately apparent to operators.

Vehicle availability is enforced through big-M constraints that bound individual vehicle charging and discharging variables by the product of the availability parameter and a large constant:

$$0 \leq v\_EV\_ENE\_IN2^{v,t} \leq p\_FL\_MAN^{v,d,t} \cdot M_{big} \tag{16}$$

$$-p\_FL\_MAN^{v,d,t} \cdot M_{big} \leq v\_EV\_ENE\_OUT2^{v,t} \leq 0 \tag{17}$$

where $M_{big} = 100$ is the big-M constant defined in the model. When $p\_FL\_MAN^{v,d,t} = 0$, these constraints force both charging and discharging variables to zero regardless of other decisions.

State-of-Charge Bounds

Vehicle battery SOC is bounded between a lower fraction $P\_EV\_SOC\_LB$ and an upper fraction $P\_EV\_SOC\_UB$ of the vehicle's total battery capacity:

$$P\_EV\_SOC\_LB \cdot p\_EV\_STO^v \leq e_t^v \leq P\_EV\_SOC\_UB \cdot p\_EV\_STO^v \quad \forall v,\ t \tag{18}$$

These bounds prevent overcharging and deep discharging, reflecting both battery safety specifications and operational requirements for battery longevity.

Objective Function

The model supports multiple optimization goals selectable through the `OPT_GOAL` parameter. The primary objective used in this work is economic maximization:

$$\max \sum_{d \in \mathcal{D}} e\_RE\_EUR\_rday^{d} + 10^{5} \tag{19}$$

where $e\_RE\_EUR\_rday^{d}$ is the net revenue expression for representative day d, defined as the sum of all revenue streams minus all procurement costs across the active market products. The constant 10^{5} is a numerical offset that ensures the objective value remains positive regardless of total costs, which simplifies convergence monitoring.

The revenue expression for the economic objective is scenario-dependent. In the baseline scenario with only day-ahead market participation (`condition_ref`), it reduces to:

$$e\_RE\_EUR\_rday^{d} = -\sum_{h} v\_EL\_ENE\_DA60\_a\_PE^{d,h} \cdot p\_EL\_ENE\_DA60\_a\_pr^{d,h} \cdot 0.1 \tag{20}$$

In more complex scenarios with intraday trading and FCR participation, additional revenue terms for electricity sales and balancing capacity are added to the expression. When battery aging is enabled (`BATT_AGING = True`), a penalty term is subtracted from the objective:

$$e\_RE\_EUR\_rday^{d} -= \sum_{v,t} \left(v\_EV\_ENE\_IN2^{v,t} - v\_EV\_ENE\_OUT2^{v,t} \right) \cdot p\_EV\_age\_factor \tag{21}$$

This penalty term discourages excessive battery cycling by assigning a cost to each unit of energy throughput, even when the price spread would otherwise make frequent charge-discharge cycles economically attractive.

Beyond economic optimization, the model also implements two alternative objectives: a CO2 minimization goal that weights energy procurement by time-varying carbon intensity factors, and a peak shaving goal that minimizes the quadratic sum of depot power imports. While these alternatives are not the focus of this thesis, their presence in the code reflects the model's flexibility and its applicability to a broader range of energy management use cases.

Sources of Decision Complexity

The interaction of all model components leads to complex decision patterns that are not immediately intuitive to human operators. Several structural properties of the model are particularly relevant in this regard.

The intertemporal coupling introduced by the SOC dynamics means that every charging or discharging decision at any time step propagates through all subsequent time steps via the state variable. The optimizer exploits this structure globally, often producing schedules that appear locally suboptimal — for example, charging at a moderately priced period rather than waiting for the cheapest period — but are globally cost-efficient given the constraints on available time windows and capacity.

The transformer capacity constraint introduces implicit competition between vehicles. When depot capacity is fully utilized, the model allocates available power among competing vehicles based on the global optimization objective rather than any explicit priority rule. A vehicle may therefore remain idle during a low-price period not because the price is unfavorable but because its battery is already sufficiently charged relative to other vehicles with more urgent energy needs, and the transformer capacity is fully assigned to those vehicles.

The multi-scale market structure adds a further layer of complexity. Day-ahead commitments made in hourly blocks constrain the intraday degrees of freedom at the 15-minute level, as the energy balance must be satisfied across all time scales simultaneously. A charging decision at a specific 15-minute interval may therefore be partly driven by the need to consume a pre-committed day-ahead quantity rather than by the current intraday price alone.

Finally, the integrality of market variables introduces discretization effects: the optimizer may round energy quantities to the nearest 100 kWh step, which can cause charging behavior to deviate slightly from what a continuous formulation would produce. These quantization effects are generally small but can occasionally be visible in charging schedules.

Together, these factors produce the non-trivial, context-dependent decision patterns that motivate the explanation approach developed in this thesis.

Electricity Market and Price Signals

A central input to the optimization model is the time-dependent electricity price vector, which reflects the cost of buying or selling energy on the electricity market at each time step. The model distinguishes between day-ahead auction prices ($p\_EL\_ENE\_DA60\_a\_pr$) provided at hourly resolution, and intraday prices ($p\_EL\_ENE\_ID15\_a\_pr$, $p\_EL\_ENE\_ID15\_t\_pr$) provided at 15-minute resolution. These prices vary over time and are influenced by factors such as total system demand, the available generation mix, and the instantaneous share of renewable energy in the grid.

Of particular significance is the possibility of negative electricity prices, which the model explicitly accommodates. Negative prices occur when electricity supply exceeds demand — for example, during periods of high renewable energy feed-in combined with low overall consumption. In such conditions, the objective function assigns a positive contribution to purchasing energy (since the price $\lambda_t < 0$ makes the cost term $-\lambda_t \cdot p_{buy,t}$ positive), creating a strong incentive to charge at maximum available power subject only to battery capacity and transformer limits.

From the perspective of the explanation system, the price trajectory across the full planning horizon is one of the most important contextual inputs. Since the optimizer considers all future prices simultaneously when constructing the schedule, a charging decision at any given time step reflects not only the current price but also the anticipated price evolution over the remainder of the planning horizon. A vehicle may not charge at a moderately low price point if the optimizer anticipates an even cheaper or negative price period later in the day, provided that the available time window and battery capacity allow deferral. Conversely, a vehicle may charge at a relatively high price if its departure constraint leaves insufficient time to wait for a cheaper window, or if the transformer capacity will be occupied by other vehicles during the cheaper period.

These interactions between price signals, battery dynamics, market commitments, and operational constraints produce the non-trivial charging behavior that operators find difficult to interpret and that the explanation pipeline developed in Chapter 4 is designed to communicate.

4 Proposed AI-Based Explanation Method

4.1 Overview of the Explanation Approach

The optimization model described in Chapter 3 determines charging strategies for a fleet of electric commercial vehicles by solving a complex decision problem. While the resulting schedules are optimal, the reasoning behind individual charging decisions is not directly observable. The interaction of multiple factors, such as electricity prices, vehicle availability, and system constraints, makes it difficult to understand why specific decisions are made.

To address this challenge, this work proposes an AI-based explanation approach that reconstructs and interprets the decision-making process of the optimization model. The method analyzes both the input data and the resulting optimization outputs in order to explain why specific charging decisions are generated. Rather than modifying the optimization algorithm itself, the approach operates as a post-processing layer that generates human-readable explanations based on the computed results.

The overall workflow of the proposed approach is illustrated in Figure 2.

4.2 Explanation Pipeline Design

The proposed explanation approach is implemented as a modular pipeline that transforms unstructured optimization outputs into interpretable information. The design of the pipeline closely follows the structure of the implemented system and consists of several sequential processing steps. Each step is responsible for extracting, enriching, or transforming information in order to reconstruct the decision context and enable explanation generation.

Step 1: Loading and Variable Extraction

The first step of the pipeline processes the raw optimization output generated by the solver. These outputs are stored in serialized formats (e.g., `.pkl` files) and contain decision variables in a compact internal representation that is not directly interpretable.

The module `convert_results.py` provides the foundation for this step. It enables loading serialized Pyomo results, parses solver variable keys into structured components (such as variable names, indices, and time steps), and reconstructs timestamps.

Building on this, the variable extraction module `extract_vars_to_csv.py` converts the processed data into a structured tabular representation. This involves extracting vehicle identifiers, mapping time indices to timestamps, and organizing the results into a format suitable for analysis. The resulting dataset provides a clear and accessible representation of optimization variables, such as charging power and energy flows.

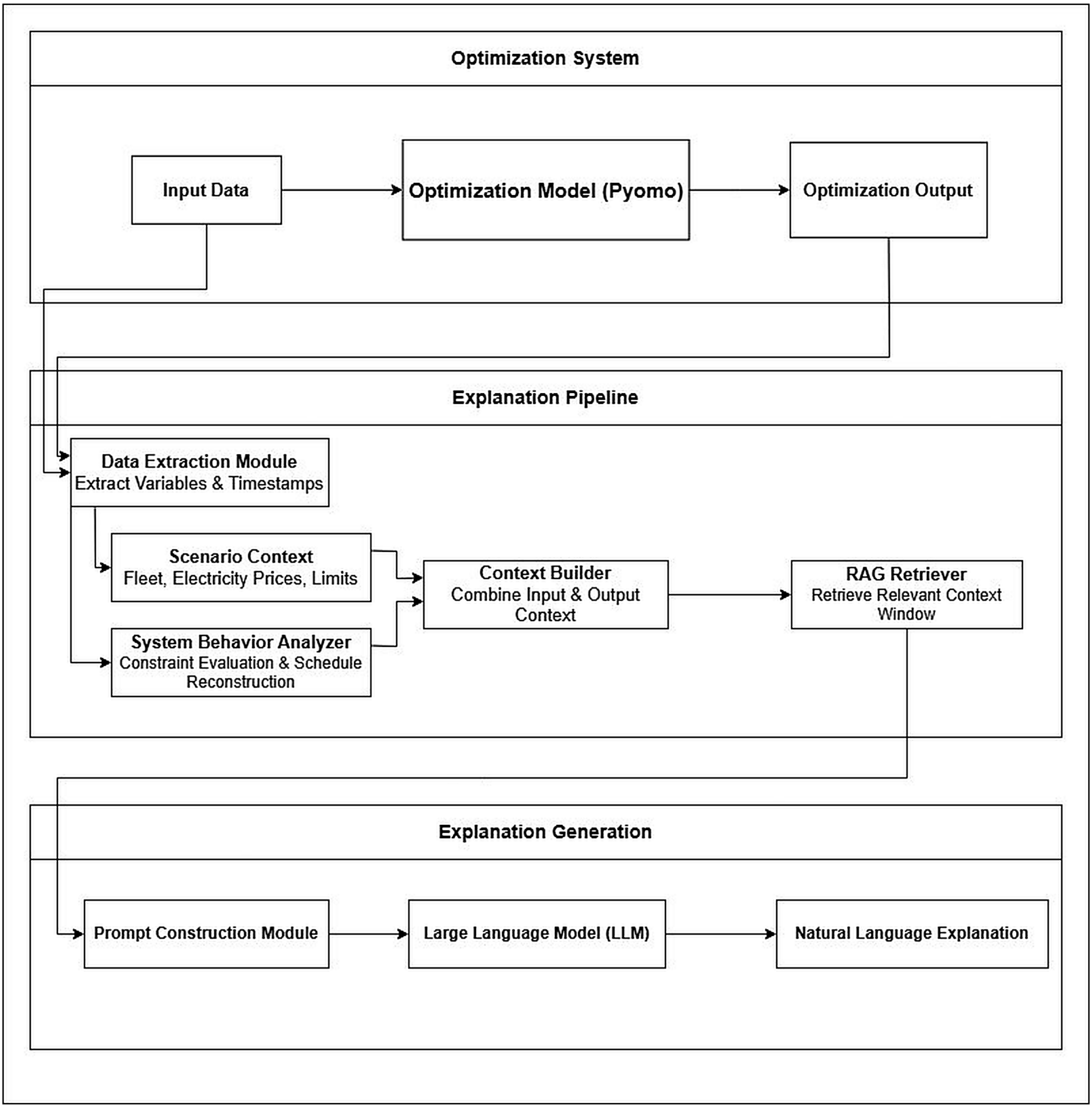

Figure 2: Overall workflow of the proposed explanation approach

Step 2: Scenario Context Construction

In parallel to processing the optimization outputs, the system constructs a structured representation of the input data. This step is implemented in `scenario_context.py` and forms the basis of the system's background knowledge.

The scenario context includes information about the fleet, battery characteristics, charging limits, and electricity prices. By organizing this information into a structured format, the pipeline creates a consistent representation of the system environment in which the optimization decisions are made.

This component serves as a knowledge base that provides the necessary context for interpreting the behavior observed in the optimization results.

Step 3: System Behavior Analysis

The system behavior analysis step investigates how the optimization model is influenced by its constraints and how these constraints shape the resulting decisions.

The optimization model includes a large number of constraints that ensure physical feasibility and system consistency. These include state-of-charge linking constraints, energy balance constraints, and intertemporal consistency constraints. While these constraints are essential for ensuring correct system behavior, they are typically always satisfied in feasible solutions and therefore provide limited insight into why specific decisions are made.

In contrast, the explanation pipeline focuses on constraints that actively influence decision-making. In particular, the depot charging capacity constraint is evaluated using the module `constraint_analyzer.py`. The aggregated charging power of all vehicles is compared to the available depot capacity at each time step. A constraint is considered binding if the total charging power approaches this limit within a predefined tolerance.

This allows the system to detect situations in which charging decisions are restricted by limited resources and to explicitly attribute observed behavior to these limitations.

In addition, the module `fleet_schedule_loader.py` reconstructs the temporal behavior of each vehicle. Based on the extracted variables, it determines whether a vehicle is charging, idle, or consuming energy due to driving. This transformation assigns semantic meaning to numerical optimization results and enables the system to describe behavior in interpretable terms.

Step 4: Temporal and Behavioral Reconstruction

In addition to constraint evaluation, the pipeline reconstructs the temporal behavior of the system based on the extracted optimization data. This step is implemented in `fleet_schedule_loader.py` and transforms raw variable values into interpretable system states over time.

The module derives meaningful behavioral representations such as charging, idle, and driving states for each vehicle. It also reconstructs the evolution of the state of charge (SOC), enabling a time-continuous view of the system dynamics.

A key aspect of this step is the introduction of a fixed temporal context window used for explanation. For a given time step of interest, the system considers a 24-hour window centered around that point in time.

Specifically, when explaining a decision at time t, the pipeline includes data from 12 hours before and 12 hours after t.

This temporal window is essential because charging decisions are not determined solely by the current system state but depend on both past and future conditions. In particular, electricity price trajectories over the entire day play a central role in shaping charging behavior. Charging is often shifted toward periods with lower or negative prices, which requires visibility of the full price curve rather than a single time point.

Furthermore, charging decisions are influenced by interactions between vehicles within the fleet. Since all vehicles share limited charging capacity, the charging behavior of one vehicle depends on the activity of others. The availability of grid capacity and the transformer limit can lead to competition for power, requiring the optimization model to prioritize certain vehicles over others.

This prioritization is typically influenced by factors such as upcoming departure times, required energy levels, and current state of charge. As a result, a vehicle may not charge at a given time even if electricity prices are favorable, because other vehicles have higher priority or the available capacity is already fully utilized.

Within the 24-hour window, the pipeline therefore includes:

- electricity price trajectories over the full period,
- the charging and operational schedules of all vehicles in the fleet,
- detailed behavior of the selected vehicle,
- and relevant system conditions such as capacity limitations.

By incorporating both temporal information and system-wide interactions, this step ensures that the explanation model has access to all relevant factors influencing the decision. This is particularly important in optimization problems with strong intertemporal and multi-agent dependencies, where decisions emerge from the interaction of economic signals, system constraints, and competing demands.

As a result, the reconstructed temporal context forms the basis for meaningful and causally grounded explanations in the subsequent stages of the pipeline.

Step 5: Prompt Construction and Explanation Generation

In the final step, the selected context is transformed into a natural language explanation using a large language model. This functionality is implemented in `run_llm.py`.

The module constructs a structured prompt that combines the retrieved data with explicit instructions. These instructions enforce constraints on the generated output, such as avoiding unsupported assumptions, focusing on causal reasoning, and maintaining a concise and analytical style. Domain-specific rules are also included to guide the interpretation of charging behavior and electricity prices.

The language model then generates a coherent explanation that describes why charging occurs at specific times, why it does not occur at others, and how system conditions influence the decision. The model is used strictly as an interpretation component and does not influence the optimization results.

Step 6: Pipeline Orchestration

The overall pipeline execution is coordinated by the script `run_explanation_pipeline.py`. This component ensures that all processing steps are executed in the correct order and that dependencies between modules are respected.

It sequentially triggers variable extraction, scenario context construction, behavior analysis, context retrieval, and explanation generation. It also ensures that intermediate results are available before subsequent steps are executed.

By organizing the explanation process into clearly defined modules, the pipeline ensures modularity, transparency, and extensibility of the system.

4.3 Retrieval-Augmented Explanation Generation

The use of retrieval-augmented generation (RAG) is a central component of the proposed explanation approach. The optimization model produces a large amount of time-dependent and high-dimensional data, which cannot be directly processed by a language model in its entirety. Providing the full dataset would introduce unnecessary complexity and reduce the clarity and reliability of the generated explanations.

RAG addresses this challenge by restricting the input to a carefully selected subset of relevant information. Instead of operating on the complete dataset, the language model receives only the context that is directly related to the decision being explained. This ensures that the explanation remains focused and grounded in the most important aspects of the system state.

A key advantage of this approach is the reduction of noise and irrelevant information. By limiting the input to a meaningful context, the system improves both the interpretability and the consistency of the generated explanations. At the same time, the retrieval mechanism preserves all necessary information required to understand the decision, ensuring that no critical factors are omitted.

Furthermore, the use of RAG helps to mitigate common issues associated with large language models, such as hallucination or the generation of unsupported statements. Since the model operates exclusively on retrieved, structured data, the explanation remains closely aligned with the actual optimization results.

Overall, the integration of retrieval-augmented generation enables a scalable and robust explanation process, allowing complex optimization decisions to be translated into clear and reliable natural language descriptions.

4.4 Design Considerations and Flexibility

The proposed explanation pipeline is designed with a strong emphasis on modularity and flexibility. Each component of the pipeline operates independently and is responsible for a clearly defined task, which allows individual modules to be extended or replaced without affecting the overall system.

In particular, the prompt construction process is designed to be easily adaptable. The structure and content of the prompt can be modified to incorporate additional rules, domain knowledge, or explanation strategies without requiring changes to the underlying data processing pipeline. This flexibility is essential for refining explanation quality and adapting the system to different use cases.

Furthermore, the extraction of all relevant optimization variables into a structured and accessible format provides a comprehensive representation of the system state. This design choice ensures that all available information can be reused for different types of analysis and explanation tasks. It also enables future extensions, such as incorporating additional explanation methods or performing deeper analytical evaluations.

Overall, the modular architecture of the pipeline supports extensibility, transparency, and reuse. This makes the approach suitable not only for the current application but also for further development and integration into more advanced decision-support systems.

4.5 End-to-End Example of the Explanation Pipeline

To illustrate the functionality of the proposed explanation approach, a complete end-to-end example is presented. The goal is to demonstrate how raw optimization data is transformed into a human-readable explanation.

Figure 3 shows the operational context of vehicle EV 2 over a 24-hour period. The figure includes both the vehicle status (on route, at the depot without charging, charging) and the corresponding electricity price trajectory.

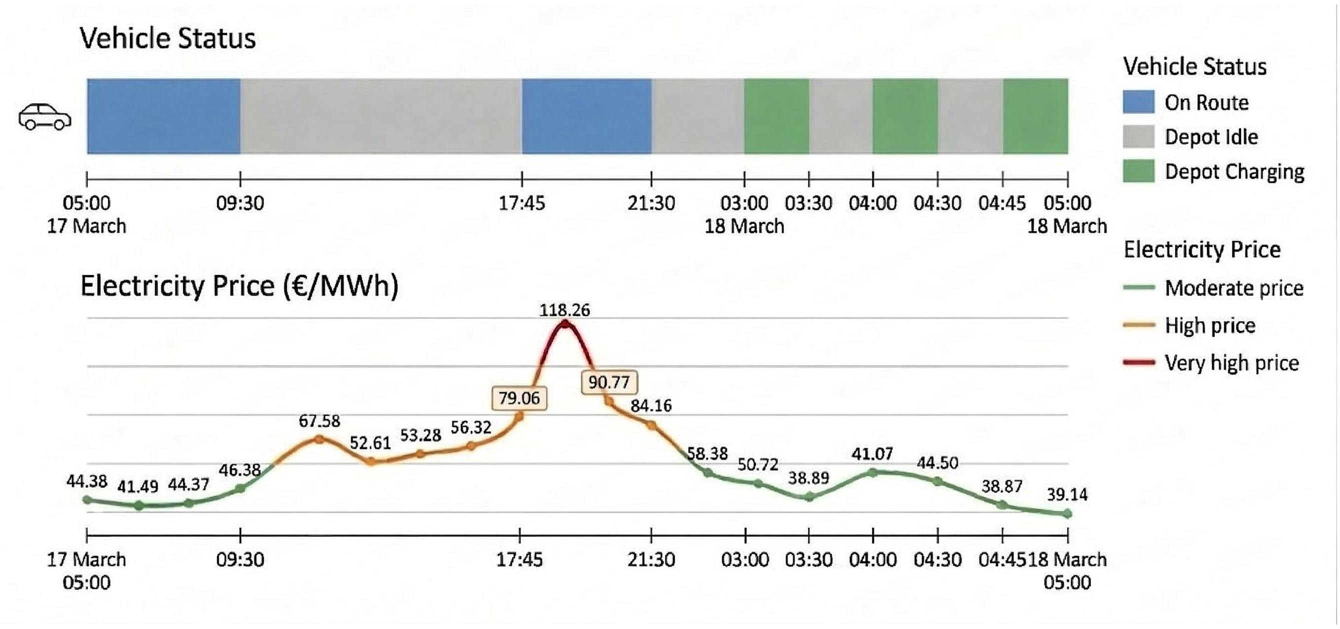

Figure 3: Operational context and electricity price trajectory for EV 2 (AI-generated)

It should be noted that Figure 3 represents only the operational behavior of the selected vehicle and the overall electricity price trajectory. While this visualization focuses on EV 2, the explanation pipeline incorporates the schedules and charging behavior of all vehicles in the fleet, as well as shared system constraints such as transformer capacity. These interactions influence charging priorities and therefore play a critical role in the resulting explanation.

In practice, explanations are generated for a specific time step of interest. For example, a fleet operator may want to understand why EV 2 is not charging at a particular time, such as around 17:00. To provide a meaningful explanation, the system constructs a temporal context window of 24 hours centered around the selected time. This window includes 12 hours before and 12 hours after the chosen time step, ensuring that both past conditions and future opportunities are taken into account.

Within this window, the pipeline retrieves all relevant information, including electricity price trajectories, the operational schedule of EV 2, the behavior of other vehicles in the fleet, and system-level constraints. This comprehensive context is necessary because charging decisions are not based solely on the current state but are influenced by anticipated future conditions and interactions within the system.

Based on this structured context, the explanation model generates the following interpretation:
"The optimization model schedules EV 2's charging sessions between 03:00 and 05:00 on March 18th because these hours present the lowest electricity prices within the planning horizon, making it the most cost-effective window for energy acquisition. Charging is explicitly avoided during the vehicle's earlier depot periods on March 17th due to substantially higher electricity costs.

The observed fragmented charging pattern, with multiple short charging intervals interspersed with pauses, is a direct consequence of a fleet-wide strategy to distribute limited transformer capacity among multiple vehicles. This coordinated approach ensures that all vehicles meet their required state of charge for upcoming routes by optimally allocating the cheapest available power.

The final charging session concludes at 05:00, indicating that either the required state of charge has been reached or that no further cost-optimal charging is feasible within the planning horizon given ongoing fleet demand and capacity limitations."

This example demonstrates how the proposed pipeline reconstructs and explains optimization decisions using only input and output data. In particular, it highlights the importance of incorporating temporal context and system-wide interactions to generate explanations that are both meaningful and causally grounded.

5 Evaluation and Results

5.1 Evaluation Methodology

This section presents the methodology used to evaluate the proposed AI-based explanation approach. The primary objective of the evaluation is to assess whether the generated explanations (i) accurately reflect the underlying optimization decisions and (ii) provide meaningful and understandable insights for human users.

Evaluating such a system poses a fundamental challenge, as it involves both *technical correctness* and *human interpretability*. While correctness can be assessed through comparison with optimization outputs, interpretability depends on subjective user perception. To address this dual nature, the evaluation follows a two-part approach combining objective validation with human-centered assessment.

First, a scenario-based analysis is conducted to verify the correctness and completeness of the extracted and reconstructed data used for explanation generation. This step ensures that the explanation pipeline operates on a faithful representation of the optimization results.

Second, a user study is performed to evaluate the interpretability, clarity, and perceived usefulness of the generated explanations from a human perspective. This allows assessing whether the system effectively bridges the gap between complex optimization outputs and user understanding.

Together, these two components provide a comprehensive evaluation of the proposed approach, covering both system-level correctness and practical usability.

Scenario-Based Evaluation The scenario-based evaluation focuses on analyzing a set of representative charging situations extracted from the optimization results. The selection of scenarios is designed to cover a diverse range of behaviors that are particularly relevant for explanation generation.

These include:

- delayed charging despite vehicle availability,
- fragmented charging across multiple time intervals,
- charging during negative electricity price periods,
- and idle periods without charging despite favorable conditions.

Such scenarios are particularly suitable for evaluation, as they reflect non-trivial decision patterns that require explanation. In contrast to straightforward charging behavior, these cases involve trade-offs between economic objectives and operational constraints, making them ideal for assessing the explanatory capabilities of the system.

For each scenario, the extracted and reconstructed data is compared directly with the original optimization outputs. The goal is to ensure that the explanation pipeline preserves all relevant information required for meaningful interpretation.

The evaluation focuses on the following aspects:

- **Data correctness:** Verification that vehicle availability, charging activity, and electricity prices are accurately reconstructed from the original data,
- **Temporal alignment:** Preservation of the chronological structure of events, including transitions between driving, idle, and charging states,
- **Consistency:** Absence of discrepancies between reconstructed data and original optimization outputs.

This step is critical, as any inconsistency in the reconstructed data would directly affect the validity of the generated explanations. Therefore, the scenario-based evaluation serves as a prerequisite for all subsequent analysis.

User Study Evaluation To complement the technical validation, a user study is conducted to evaluate the explanations from a human-centered perspective. While scenario-based analysis ensures correctness, it does not provide insight into how explanations are perceived by users.

Participants are presented with representative charging scenarios together with the corresponding generated explanations. They are then asked to evaluate these explanations based on multiple criteria, including clarity, readability, usefulness, and the ability to answer the underlying “why” question.

Special attention is given to whether the explanations successfully communicate key decision drivers, such as electricity prices and operational constraints. This is particularly important, as these factors form the core of the optimization logic.

The study includes participants with varying levels of technical knowledge, allowing the evaluation to capture both non-expert and semi-expert perspectives. This diversity is essential to assess the practical applicability of the system in real-world settings, where users may have different levels of familiarity with optimization models.

By combining structured questionnaires with aggregated evaluation metrics, the user study provides a systematic assessment of explanation quality while accounting for variability in individual perception.

5.2 Scenario-Based Analysis

The scenario-based analysis evaluates the correctness and reliability of the data retrieval and reconstruction process used in the explanation pipeline. Since the explanation generation relies entirely on reconstructed context, this step is essential to ensure that the system operates on accurate and complete information.

Beyond verifying correctness, the analysis also provides insight into how different charging behaviors emerge from the optimization process and how these behaviors can be interpreted.

Example Scenario

Figure 4 illustrates a representative scenario for a selected vehicle over a 24-hour period. The figure combines the reconstructed vehicle status with the corresponding electricity price trajectory.

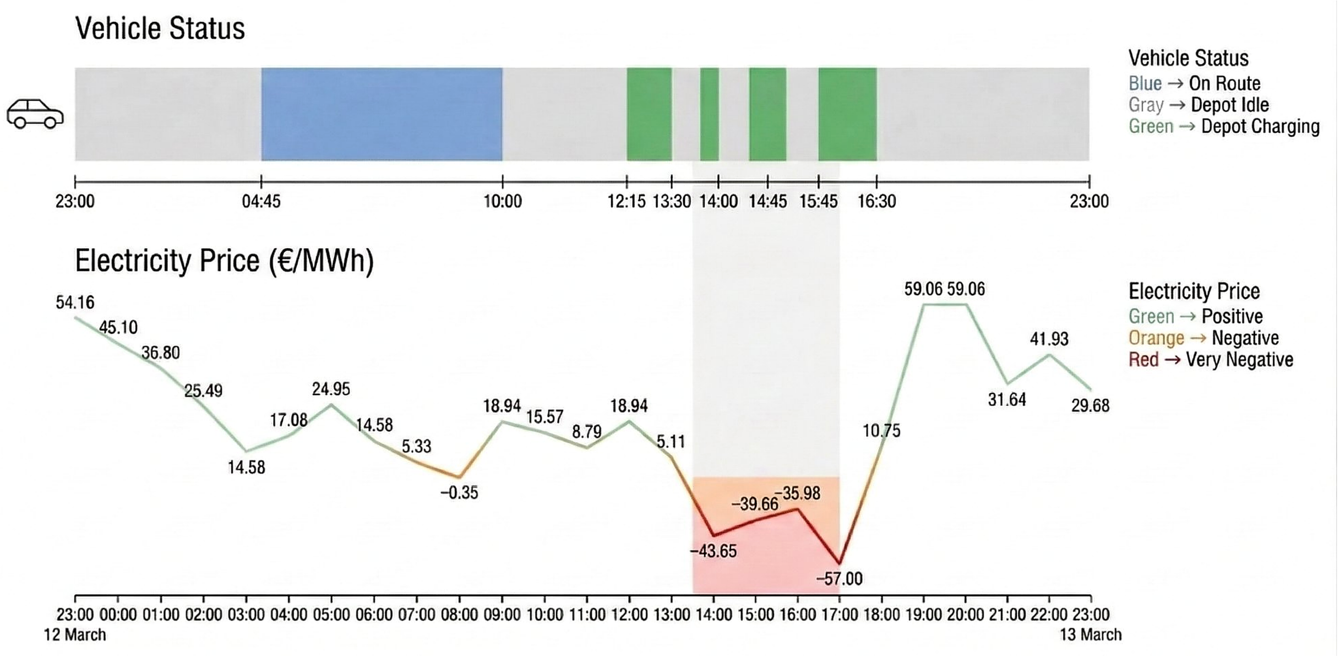

Figure 4: Charging behavior and electricity prices for EV 8 over a 24-hour period (AI-generated)

The reconstructed data shows that the vehicle is initially on route and becomes available at the depot later in the day. Following its return, charging activity occurs in several distinct intervals, interspersed with idle periods despite vehicle availability.

From an optimization perspective, this behavior reflects a non-trivial decision pattern. Instead of charging continuously upon arrival, the system distributes charging over multiple intervals. This suggests that the optimization model balances multiple factors, such as electricity price fluctuations and system constraints.

A closer inspection of the price trajectory reveals that charging activity is aligned with periods of lower or negative electricity prices. This indicates that economic signals play a significant role in shaping the charging schedule. At the same time, the presence of idle periods despite favorable prices suggests that additional constraints, such as capacity limits or interdependencies with other vehicles, may influence the decision.

To validate the correctness of this reconstruction, the extracted data was compared with the original optimization outputs. The comparison confirmed that:

- vehicle availability periods match the original schedule,
- charging and idle intervals are correctly identified,
- and the electricity price trajectory corresponds to the input data.

No discrepancies were observed. In particular, the temporal alignment between vehicle availability, charging behavior, and electricity prices was preserved, ensuring that the reconstructed scenario accurately reflects the original system behavior.

Validation of Retrieved Data

To further evaluate the robustness of the pipeline, a manual validation was conducted on five representative temporal windows selected from different parts of the dataset. These windows were chosen to cover a range of charging behaviors, including delayed charging, fragmented charging, and charging during negative price intervals.

For each window, the reconstructed data was compared directly with the original solver outputs. The validation focused on:

- correctness of vehicle operational states (driving, idle, charging),
- accuracy of charging intervals and power allocation,
- consistency of state-of-charge evolution,
- and alignment of electricity price data.

The validation was performed through manual inspection of raw data and corresponding time-series representations. Across all evaluated windows, no inconsistencies were detected.

This result demonstrates that the data processing pipeline reliably preserves both the structural and temporal properties of the optimization output.

Summary

The scenario-based analysis confirms that the data extraction and reconstruction process is both accurate and reliable. All relevant system variables are correctly represented, and the temporal structure of the optimization output is preserved.

This is a crucial result, as it ensures that the generated explanations are grounded in a faithful representation of the underlying system. Consequently, any insights derived from the explanations can be attributed to the optimization behavior itself rather than artifacts of data processing.

5.3 User Study and Questionnaire Results

The user study was conducted with a group of 13 participants, including both students and individuals with some domain-related knowledge. The primary objective of the study was to evaluate the generated explanations from a human-centered perspective, focusing on their clarity, interpretability, and practical usefulness.

In contrast to the technical validation presented in the scenario-based analysis, this evaluation aims to assess whether the explanations effectively support users in understanding the underlying optimization decisions. Since explainability is inherently subjective, the study is designed to capture user perception while maintaining a structured and comparable evaluation framework.

Study Design and Model Usage

To evaluate the quality and interpretability of the generated explanations, a questionnaire-based user study was conducted. The questionnaire consisted of two complementary parts, allowing both a comparative and an absolute evaluation of the proposed approach.

In the first part, participants were presented with three different explanations for the same charging scenario and were asked to select the one they found easiest to understand. The purpose of this comparison was to assess how the explanation style of the proposed approach is perceived relative to alternative language models.

To ensure a fair and controlled comparison, all models were provided with identical input data and the same prompt structure. This isolates the effect of the language model itself, allowing differences in explanation quality to be attributed primarily to the model's generation behavior rather than variations in input or prompt design.

Specifically:

- Explanation A was generated using Gemini 2.5 Flash, which is also the model used in the proposed explanation pipeline,
- Explanation B was generated using ChatGPT 5.2,
- Explanation C was generated using ChatGPT 5.1.

In the second part of the questionnaire, participants evaluated individual explanations generated by the proposed system across multiple scenarios. For this purpose, only the Gemini 2.5 Flash model was used, ensuring consistency with the implemented pipeline.

The evaluated scenarios were selected to represent a range of charging behaviors, including delayed charging, fragmented charging, and charging under varying electricity price conditions. This diversity ensures that the evaluation captures different types of decision patterns rather than focusing on a single case.

Participants rated the explanations using Likert-scale questions addressing:

- clarity and readability,
- interpretability and understanding,
- perceived usefulness for decision support,

- and the ability to explain the influence of electricity prices.

This design enables both a relative comparison between explanation styles and an absolute assessment of the proposed approach across different scenarios.

Results

The results of the questionnaire are presented using aggregated visualizations to provide a clear and structured overview of participant responses. Beyond descriptive statistics, the results are analyzed to derive insights into the strengths and limitations of the proposed explanation approach.

Comparison of Explanation Styles Figure 5 shows the distribution of participant preferences for the three explanations generated by different language models.

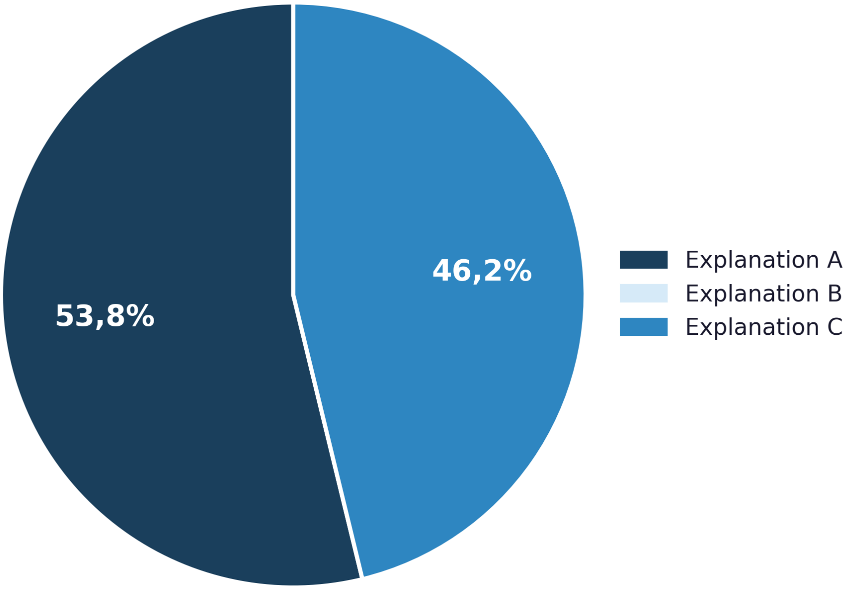

Figure 5: Participant preference for explanations generated by different language models

The results indicate that Explanation A was preferred by the majority of participants, receiving 53.8% of the selections. Explanation C was chosen by 46.2% of participants, while Explanation B was not selected.

This result suggests that the explanation generated by the proposed approach is generally perceived as more understandable compared to alternative formulations. At the same time, the relatively close distribution between Explanation A and Explanation C indicates that multiple explanation styles can be effective, although differences in clarity and structure influence user preference.

Qualitative Feedback from Participants In addition to quantitative ratings, participants provided qualitative feedback explaining their preferences and highlighting strengths and weaknesses of the different explanation styles. This feedback offers deeper insight into how explanations are perceived and which aspects contribute to their effectiveness.

Clarity and Simplicity as Key Factors. A recurring observation across participants is the importance of clarity and simplicity. Several participants emphasized that explanations which directly link charging decisions to electricity prices and vehicle requirements were easier to understand. For example, one participant stated:

> “Explanation A works best because it’s simple and easy to understand. It clearly connects the charging decisions to prices and battery needs without adding unnecessary complexity.”

Similarly, another participant noted that explanations should focus on conveying the overall logic rather than excessive detail:

> “The explanation provides a clear understanding of how the charts are structured. The exact figures can be seen in the charts if needed, but this level of detail is not necessary in the text.”

These responses indicate that users prefer explanations that highlight the main decision drivers while avoiding unnecessary technical complexity.

Importance of Coherence and Data Alignment. Participants also valued explanations that were consistent with the underlying data and provided a coherent narrative. One participant commented:

> “It provides the most coherent and technically accurate explanation.”

Another participant highlighted that linking decisions to specific time periods and constraints improves perceived realism:

> “It clearly connects the charging decisions to specific time periods and electricity prices while also considering constraints like state of charge and transformer limits.”

This suggests that explanations are more effective when they explicitly connect observed behavior to concrete system variables.

Preference for Appropriate Level of Detail. The feedback reveals that different users prefer different levels of detail. While some participants favored more comprehensive explanations, others preferred concise and direct formulations. For example:

> “Even though Explanation A is more detailed, I found Explanation C easier to follow, because it is more to the point.”

This indicates that explanation effectiveness is influenced not only by content but also by presentation style. A balance between completeness and readability is therefore essential.

Impact of Language and Terminology. Several participants commented on the role of language and technical terminology. Explanations containing too many technical terms were perceived as harder to understand:

> “Explanation B is written in an unnecessary technical style and had too many technical terms, which makes it harder to understand.”

In contrast, simpler and more natural language was preferred, particularly for users with less technical background. This suggests that explanation systems should adapt language complexity to the target audience.

Stylistic and Structural Observations. Participants also provided feedback on writing style and phrasing. For example, the use of first-person formulations (e.g., “I charge”) was perceived as less appropriate in a technical context. Additionally, expressions indicating uncertainty (e.g., “indicating X or Y”) were criticized for reducing confidence in the explanation.

Furthermore, some participants identified redundant or obvious statements, such as explicitly mentioning that the objective is cost minimization. These observations indicate that concise and confident formulations improve perceived quality.

Summary of Qualitative Insights. Overall, the qualitative feedback highlights several key factors influencing explanation quality:

- clear and simple language improves understandability,
- explicit links between decisions and observable data increase trust,
- avoidance of unverifiable assumptions is crucial,
- and the level of detail should be carefully balanced.

These insights complement the quantitative results and provide valuable guidance for improving the design of AI-generated explanations.

Aggregated Evaluation Across Scenarios To evaluate the overall performance of the proposed approach, participant ratings from all evaluated scenarios were aggregated. The aggregated results are obtained by first computing the average rating for each evaluation dimension within each scenario, and subsequently averaging these values across all considered scenarios. This two-step aggregation ensures that the results reflect both participant responses and variations between different charging situations, providing a robust and balanced overall assessment.

Figure 6 summarizes the average ratings for key evaluation criteria, including clarity, usefulness, trustworthiness, and the ability to explain electricity price effects.

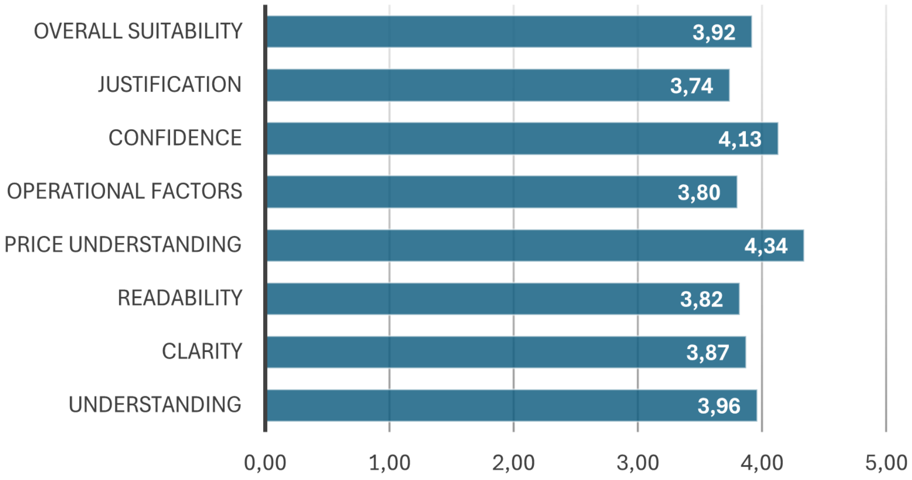

Figure 6: Average participant ratings across evaluation dimensions

The aggregated results show consistently high scores across all dimensions, indicating that the generated explanations are generally perceived as clear, informative, and reliable. In particular, the role of electricity prices is effectively communicated, supporting the interpretability of the charging decisions.

Additional Observations and Feedback Beyond individual preferences, participants provided additional feedback highlighting recurring strengths and limitations of the generated explanations. These observations offer important insights into how the explanations are interpreted in practice and where improvements are necessary.

Uncertainty Due to Missing or Implicit Information. Several participants pointed out that certain explanations relied on assumptions that were not explicitly supported by the available data. In particular, reasons such as transformer capacity limits or state-of-charge (SOC) constraints were sometimes inferred without clear evidence:

> "Some reasons, like transformer limits or SOC being reached, were assumed and not clearly shown in the data."

Similarly, participants questioned whether charging termination was always sufficiently justified:

> "In some examples it was not sure if the battery is full or why it stops charging."

This indicates that explanations should more explicitly distinguish between observable facts and inferred reasoning. Providing clearer justification or explicitly marking assumptions could improve transparency and user trust.

Need for More Precise Data Linking. While explanations were generally perceived as clear, some participants noted that they remain relatively high-level and do not always establish precise links between decisions and specific data points:

> "The explanation is generally clear, but it remains somewhat high-level and does not always precisely link each charging decision to specific timestamps or price values."

This suggests that improving the granularity of explanations—by referencing concrete timestamps, price levels, or state changes—could enhance interpretability and allow users to verify reasoning more easily.

Positive Feedback on Decision Logic Explanation. Participants consistently highlighted the effective communication of core decision drivers, particularly the role of electricity prices and operational constraints:

> "The explanation clearly highlights the role of electricity prices and operational constraints in driving the charging strategy."

This confirms that the system successfully captures and conveys the fundamental logic of the optimization model, which is a key objective of this work.

Preference for Simpler Language. The feedback also emphasizes the importance of language style. Several participants expressed a preference for explanations written in a simpler and more accessible way:

> "Personally, I found explanations in a more casual language style more intuitive and quicker to understand."

This suggests that explanation systems should consider adapting language complexity to the target audience. While technical language may improve precision, simpler formulations are often more effective for understanding.

Ambiguity in Priority and Decision Justification. In some cases, participants questioned the reasoning behind prioritization decisions:

> "It is not clear why other vehicles have higher priority."

This indicates that explanations involving implicit prioritization or resource allocation should either provide justification or avoid introducing concepts that cannot be supported by the available data.

Summary of Additional Insights. The additional feedback reveals several cross-cutting themes:

- explanations should clearly distinguish between observed data and inferred reasoning,
- verifiability is critical, especially for system-level or fleet-level statements,
- more precise references to data (e.g., timestamps, prices) improve transparency,
- simpler language enhances usability for non-expert users,

These findings provide valuable guidance for refining the explanation approach and further improving its practical applicability.

Conclusion Across all evaluated scenarios, the explanations demonstrate a stable level of quality, as reflected by the relatively consistent scores across different evaluation dimensions. The aggregation of results highlights that the observed performance is not limited to individual cases but generalizes across multiple charging situations.

Participants consistently rated the explanations highly in terms of understanding, confidence, and the ability to convey the influence of electricity prices. This indicates that the proposed approach successfully captures and communicates the core decision logic of the optimization model, particularly with respect to price-driven charging behavior and operational constraints.

At the same time, the qualitative feedback reveals important limitations that are not fully reflected in the aggregated scores. In particular, participants identified that some explanations rely on inferred reasoning—such as assumptions about transformer capacity limits or state-of-charge conditions—that are not explicitly supported by the available data. This affects the verifiability of the explanations and may reduce user trust in certain cases.

Furthermore, several participants noted that the explanations remain relatively high-level and do not always establish precise links between decisions and specific data points, such as timestamps or exact price values. This suggests that while the overall reasoning is understandable, the level of detail is sometimes insufficient for step-by-step verification.

Another key observation is the importance of language and presentation. Simpler and more direct explanations were generally preferred, particularly by non-expert users, while overly technical formulations were perceived as less accessible. This highlights the need for explanation strategies that balance precision with readability.

Overall, the results demonstrate that the proposed system is effective in providing intuitive and meaningful explanations of complex optimization decisions. However, they also indicate that further improvements are necessary to enhance transparency, strengthen data grounding, and adapt explanations to different user needs.

5.4 Discussion, Limitations and Future Work

5.4.1 Discussion

The results presented in this chapter demonstrate that the proposed AI-based explanation approach is capable of providing meaningful and interpretable insights into optimization decisions in smart charging systems. Both the scenario-based analysis and the user study indicate that the generated explanations successfully capture key decision drivers, particularly the influence of electricity prices, vehicle availability, and system-level interactions.

The aggregated evaluation results show that participants generally perceived the explanations as clear, informative, and reliable. In particular, high ratings in dimensions such as price understanding and confidence suggest that the approach effectively communicates the economic reasoning underlying charging decisions. At the same time, slightly lower scores in readability and justification indicate that further improvements in explanation structure and clarity could enhance user understanding.

Qualitative Comparison of Explanation Styles To complement the quantitative evaluation, a qualitative comparison of explanation styles across different language models was conducted. The objective of this analysis is to identify structural and stylistic differences that influence user perception and interpretability. An overview of the key characteristics and differences between the evaluated explanation styles is summarized in Table 2.

Table 2: Qualitative comparison of explanation styles across language models

Criterion	Gemini 2.5 Flash	ChatGPT 5.2	ChatGPT 5.1
Optimization framing	Cost and price oriented; limited marginal trade-off language	Explicit global objective and allocation framing	Explicit constraint and marginal reasoning
Shared resource clarity	Mentions transformer limits, sometimes implicitly	Clearly frames shared transformer constraint	Explicit time-slicing under capacity constraints
Handling of negative prices	Clearly described as revenue opportunity with concrete values	Revenue framed within objective comparison	Strong economic and constraint-based framing
Fragmentation explanation	Attributed to dynamic allocation effects	Attributed to binding shared resource trade-offs	Explicit sharing under capacity constraints
Temporal narration	Often includes timestamps and price values	Abstract over time; avoids detailed timestamps	Structured reasoning chain over time
Abstraction level	Medium	High	High
Hedging language	Occasional	Minimal	Minimal
Conciseness	Moderate	High	Moderate

The comparison highlights that differences between models arise primarily from the level of abstraction and explanatory framing rather than factual correctness. In particular, explanations that combine concrete system details, such as timestamps and price values, with clear causal reasoning tend to be perceived as more understandable. This observation is consistent with the results of the user study.

5.4.2 Limitations

Despite the promising results, several limitations of this work must be acknowledged. These limitations arise from restricted access to the underlying optimization model, methodological design choices, and inherent constraints of language-model-based explanation systems.

Limited Access to the Optimization Process A fundamental limitation of this work is the restricted practical access to the underlying optimization model. While the source code of the optimization model was available and could be analyzed, the model could not be executed within the scope of this thesis due to missing dependencies and the complexity of the required runtime environment.

As a result, it was not possible to perform controlled experiments, such as modifying input parameters or using synthetic (dummy) data to observe how the optimization behavior changes under specific conditions. This limited the ability to systematically investigate cause-effect relationships within the model.

Consequently, the analysis of the optimization process is based primarily on static inspection of the code and the available input-output data, rather than on dynamic experimentation. In particular, the internal execution behavior and intermediate decision states of the solver were not directly observable.

This has important implications for explainability. The proposed system reconstructs explanations in a post hoc manner, based on observable patterns in the input and output data, complemented by insights gained from code analysis. However, optimization solvers follow complex internal decision paths that cannot be fully inferred from final results alone.

Therefore, the system cannot always explain why a specific decision was chosen over an equally optimal alternative, nor can it fully capture the influence of intermediate solver states. This limits the ability to provide causal explanations and instead results in plausible but indirect interpretations of decision-making behavior.

Constraint Interpretability Limitations The optimization model includes a large number of constraints, such as energy balance equations, state-of-charge (SOC) updates, transformer limits, depot capacity restrictions, and inter-temporal consistency constraints. While these constraints are essential for ensuring physical feasibility, they do not necessarily contribute directly to human-understandable reasoning.

Many constraints are structural in nature, meaning they enforce system validity rather than represent active decision drivers. As a result, even if constraint-level analysis were incorporated into the explanation process, a significant portion of these constraints would provide limited explanatory value to end users. This creates a gap between mathematical correctness and human interpretability.

However, this limitation also highlights an important opportunity for more advanced analysis. Constraints become particularly relevant when they are binding or close to their operational limits. In such cases, they directly influence the optimization outcome by restricting feasible decisions. For example, transformer capacity limits or depot constraints may force the system to delay or redistribute charging processes, even if electricity prices would otherwise favor immediate charging.

The current approach does not explicitly identify or highlight such situations. As a result, it may fail to capture when and how constraints actively shape decisions, focusing instead on more observable factors such as price signals or schedules.

A more advanced explanation system could address this limitation by detecting active or near-binding constraints, quantifying their impact on decision variables, and explicitly linking constraints to specific deviations from economically optimal behavior. This would enable a clearer distinction between decisions driven by economic objectives and those enforced by technical or physical limitations. However, such an approach would require deeper integration with the optimization model and access to additional internal solver information, which was not available within the scope of this work.

Data Complexity and Preprocessing Overhead Another limitation arises from the complexity and format of the available data. A significant portion of the input and output data is stored in serialized `.pkl` (pickle) files, which are not directly human-readable and require explicit deserialization and preprocessing.

This introduces additional processing steps before interpretation can take place. Transforming raw data into structured formats suitable for analysis increases implementation complexity and may introduce potential sources of error or information loss. Although the proposed pipeline addresses this challenge through structured data processing and context construction, handling high-dimensional and time-dependent data remains inherently demanding.

Temporal Scope Limitation The explanation system operates on a fixed 24-hour time window around a selected timestamp. While this simplifies the explanation process and ensures consistency, it introduces a limitation in capturing long-term dependencies.

Optimization decisions may be influenced by events outside the selected time window, such as future electricity prices, multi-day scheduling constraints, or boundary conditions of the optimization horizon. By restricting the explanation to a fixed temporal scope, relevant influencing factors may be omitted, leading to incomplete representations of decision logic.

Static Prompt Design and Limited Adaptability From a technical perspective, the system relies on a predefined and static prompt structure for generating explanations. While this ensures consistency across different scenarios, it limits adaptability.

The current approach does not dynamically adjust to different scenario types, optimization objectives, or stakeholder perspectives. As a result, explanations may not always emphasize the most relevant aspects for a given context, which can reduce their effectiveness in more complex or diverse scenarios.

Lack of Interactivity The system generates a single, static explanation for each scenario and does not support interactive exploration. Users cannot ask follow-up questions, focus on specific vehicles or time intervals, or explore alternative scenarios.

This limits the system's role to that of a reporting tool rather than an interactive decision-support system. In real-world applications, explainability is often an iterative process, where users refine their understanding through interaction. The absence of such capabilities restricts the depth of user engagement and insight.

Evaluation Constraints and Subjectivity The evaluation of the proposed approach is based on a relatively small number of participants. While the results provide useful initial insights, a larger and more diverse sample would be required to obtain more generalizable conclusions.

Furthermore, the evaluation of explanations inherently involves a qualitative component. Although structured questionnaires and aggregated metrics were used, individual interpretations, prior knowledge, and personal preferences may influence the results. This introduces a degree of subjectivity that cannot be entirely eliminated.

Limitations of Language Model-Based Explanations The explanation generation relies on a large language model, which introduces additional limitations. Language models are optimized for generating coherent and contextually plausible text, but they do not provide formal guarantees of correctness.

As a result, explanations may appear convincing even if they are incomplete, and the reasoning is not mathematically verifiable. The model may omit relevant factors or overemphasize others. Therefore, the generated explanations should be understood as supporting tools for interpretation rather than exact representations of the underlying optimization logic.

Dependence on Model Capabilities Finally, the quality of the generated explanations is directly dependent on the capabilities of the underlying language model. Improvements or limitations in the model's reasoning and language generation abilities will directly affect system performance. This creates an external dependency that is not fully controllable within the scope of this work.

5.4.3 Future Work

The results of this work demonstrate that AI-based explanations can effectively improve the interpretability of complex optimization decisions. However, several limitations identified in the evaluation highlight important directions for future research and development.

Extended and More Diverse Evaluation One important direction for future work is the expansion of the evaluation methodology. The current user study is based on a relatively small sample size and primarily includes students and participants with limited domain expertise. Future studies should involve a larger and more diverse group of participants, including domain experts such as fleet operators and energy system specialists.

A more extensive evaluation would enable a deeper analysis of how different user groups perceive explanations and whether explanation requirements vary depending on technical background and application context. In addition, longitudinal studies could be conducted to assess how explanations influence user understanding and decision-making over time.

Adaptive and User-Centered Explanation Strategies The results indicate that different users prefer different levels of detail and language complexity. While some participants favored concise explanations, others preferred more detailed and technically precise descriptions. This suggests that a one-size-fits-all explanation strategy is not sufficient.

Future work could therefore focus on developing adaptive explanation systems that dynamically adjust:

- the level of detail,
- the language complexity,
- and the focus of the explanation

based on user preferences, roles, or expertise.

Such personalization could significantly improve usability and ensure that explanations remain both accessible for non-experts and informative for advanced users.

Integration of Interactive Exploration Capabilities Another promising direction is the integration of interactive features. The current system provides static explanations, which limits the ability of users to explore the reasoning process in more depth.

Future systems could enable:

- interactive querying (e.g., follow-up questions),
- scenario exploration ("what-if" analyses),

- and filtering of explanations for specific vehicles or time intervals.

Combining natural language explanations with interactive visualization tools could further enhance understanding by allowing users to directly relate textual explanations to underlying data and system behavior.

Improved Data Linking and Verifiability User feedback revealed that explanations should more explicitly reference observable data, such as specific timestamps, electricity prices, or state-of-charge values. Future work could focus on strengthening the connection between explanations and data by incorporating explicit references and quantitative evidence.

In addition, mechanisms for distinguishing between observed facts and inferred reasoning could improve transparency. For example, explanations could explicitly indicate when certain statements are assumptions rather than directly verifiable facts.

Constraint-Aware and Model-Informed Explanations As discussed in the limitations, the current approach does not explicitly identify when constraints are binding or actively influencing decisions. Future work could extend the explanation system to detect and highlight such situations.

In particular, identifying near-binding constraints (e.g., transformer capacity limits or SOC constraints) and quantifying their impact on decisions would provide deeper insight into the optimization process. This would enable explanations to distinguish more clearly between economically driven decisions and those enforced by technical limitations.

Such improvements would likely require tighter integration with the optimization model and access to additional solver information.

Trade-off Between Model Independence and Explainability Depth From a methodological perspective, the proposed approach operates as a post-processing layer based solely on input and output data. This design ensures independence from the underlying optimization model, which is advantageous in scenarios where direct model access is not possible.

However, this independence introduces a fundamental trade-off. The lack of access to the internal structure and execution of the optimization model limits the depth of explanation that can be achieved. Certain decision mechanisms and intermediate reasoning steps cannot be directly observed and must be inferred.

Future work could investigate approaches that balance this trade-off, for example by incorporating partial model information or intermediate solver outputs while maintaining a modular and flexible system design.

Conclusion This thesis addressed the challenge of making optimization-based smart charging decisions interpretable for human operators. The central motivation was the observation that while modern energy management systems for electric commercial vehicle fleets are capable of producing globally optimal charging schedules, the reasoning behind individual decisions remains opaque. Operators confronted with counterintuitive behavior — such as vehicles remaining idle during periods of available capacity, or charging occurring at seemingly unfavorable price points — lack the tools to understand, verify, and trust the system's outputs. This gap between mathematical optimality and human interpretability represents a practical barrier to the adoption of automated charging management in real-world depot operations.

To address this challenge, the thesis developed and evaluated an AI-based explanation pipeline that transforms raw optimization outputs into human-readable natural language descriptions of individual charging decisions. The pipeline operates as a post-processing layer and consists of several sequential components: variable extraction and reconstruction, scenario context construction, system behavior analysis including constraint activity detection, temporal context retrieval using a 24-hour window, and finally natural language generation via a large language model. A central design principle throughout was the use of retrieval-augmented generation, which grounds the language model's output in verified, structured data derived directly from the optimization results, thereby reducing the risk of hallucination and improving the factual reliability of the generated explanations.

The evaluation of the proposed approach was structured around two complementary components. The scenario-based analysis confirmed that the data extraction and reconstruction pipeline accurately preserves all relevant system variables, temporal relationships, and constraint interactions from the original optimization output. No discrepancies were detected across the evaluated scenarios, establishing the correctness of the information on which explanations are based. The subsequent user study, conducted with 13 participants of varying technical backgrounds, demonstrated that the generated explanations are perceived as clear, informative, and useful for understanding the underlying decision logic. Aggregated ratings across evaluation dimensions including clarity, understanding, confidence, and price comprehension consistently exceeded 3.7 on a five-point scale, with price understanding receiving the highest average rating of 4.34. In a direct comparison between explanation styles generated by different language models, the system's primary model was preferred by 53.8 percent of participants.

Beyond the quantitative results, the qualitative feedback collected during the user study provided important insights into the practical requirements for explanation systems in operational contexts. Participants valued explanations that establish explicit, verifiable links between decisions and observable data, and expressed a preference for clear and accessible language over technically precise but complex formulations. At the same time, participants identified limitations: some explanations relied on inferred reasoning such as assumptions about binding transformer constraints or state-of-charge conditions that could not be directly verified from the available visualizations. This finding points to a fundamental tension in post-hoc explanation systems: the explanations must be informative enough to be useful, yet they cannot claim more certainty than the data supports.

Taken together, the results demonstrate that the proposed pipeline successfully closes the interpretability gap for the most common and practically relevant class of charging behaviors, particularly those driven by electricity price optimization and fleet-wide capacity coordination. The approach is effective in communicating the economic logic of the optimization model and the role of shared infrastructure constraints, which are precisely the aspects that operators find most difficult to interpret without assistance.

Several limitations of the current work were identified and should be considered when interpreting the results. The explanation system operates exclusively on the input and output data of the optimization model, without access to intermediate solver states or dual variable information. This means that explanations are necessarily reconstructed from observable patterns rather than derived from the internal reasoning of the solver, which limits their causal precision. The fixed 24-hour temporal window may omit relevant factors for decisions influenced by multi-day planning dynamics. The user study sample, while sufficient for exploratory evaluation, was limited in size and did not include professional fleet operators or energy system specialists, which restricts the generalizability of the findings. Finally, the static and non-interactive nature of the generated explanations limits their usefulness in operational workflows where operators may need to ask follow-up questions or explore specific aspects of a decision in more depth.

These limitations point toward clear directions for future work. Tighter integration with the optimization model — for example through access to constraint dual values or sensitivity information — would enable constraint-aware explanations that explicitly quantify how binding a given limit is and how much the charging schedule would change if that limit were relaxed. Adaptive explanation strategies that adjust the level of detail and language complexity to the expertise of the user would improve practical applicability across different operator profiles. The addition of interactive exploration capabilities, such as natural language follow-up queries or what-if scenario analysis, would transform the system from a reporting tool into a genuine decision-support interface. Finally, a larger and more diverse evaluation study involving domain experts would strengthen the empirical basis for the approach and provide more targeted guidance for further development.

The broader contribution of this thesis lies in demonstrating that the boundary between optimization and explainability need not be as sharp as it currently appears in practice. Optimization models encode rich reasoning about constraints, trade-offs, and economic signals that is fully accessible in the model's data structures, even if it is not directly visible in the final schedule. By systematically extracting, structuring, and communicating this reasoning through a language model grounded in verified data, it is possible to provide operators with explanations that are not only technically accurate but also genuinely useful for building understanding and trust in automated energy management systems. As electric commercial fleets grow in scale and as the complexity of energy markets continues to increase, the ability to explain optimization decisions will become an increasingly important component of practical, human-centered energy management.

Declaration of AI Assistance

This thesis was written independently. Language refinement and minor stylistic improvements were supported using tools such as ChatGPT. All content, analysis, and conclusions were developed by the author.

Three figures were generated using AI-based tools solely for design purposes; the underlying data, content, and interpretations presented in these figures are entirely the author's own work.

References

[1] Saleh Afroogh et al. “Trust in AI: Progress, Challenges, and Future Directions”. In: *arXiv e-prints*, arXiv:2403.14680 (Mar. 2024), arXiv:2403.14680. DOI: `10.48550/arXiv.2403.14680`. arXiv: `2403.14680 [cs.CY]`.

[2] Mada Altiary and Farookh Khadeer Hussain. “Explainability in carbon price forecasting systems: A systematic literature review”. In: *Knowledge-Based Systems* 330 (2025), p. 114612. ISSN: 0950-7051. DOI: `https://doi.org/10.1016/j.knosys.2025.114612`. URL: `https://www.sciencedirect.com/science/article/pii/S095070512501651X`.

[3] Alejandro Barredo Arrieta et al. “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI”. In: *Information Fusion* 58 (2020), pp. 82–115. ISSN: 1566-2535. DOI: `https://doi.org/10.1016/j.inffus.2019.12.012`. URL: `https://www.sciencedirect.com/science/article/pii/S1566253519308103`.

[4] Erik Cambria et al. “A survey on XAI and natural language explanations”. In: *Information Processing Management* 60.1 (2023), p. 103111. ISSN: 0306-4573. DOI: `https://doi.org/10.1016/j.ipm.2022.103111`. URL: `https://www.sciencedirect.com/science/article/pii/S0306457322002126`.

[5] Wenlin Chen et al. “Research on Explainability Methods for Unmanned Combat Decision-Making Models”. In: *IEEE Access* 12 (2024), pp. 83502–83512. DOI: `10.1109/ACCESS.2024.3409616`.

[6] Kristien Clement-Nyns, Edwin Haesen, and Johan Driesen. “The Impact of Charging Plug-In Hybrid Electric Vehicles on a Residential Distribution Grid”. In: *IEEE Transactions on Power Systems* 25.1 (2010), pp. 371–380. DOI: `10.1109/TPWRS.2009.2036481`.

[7] Sara Deilami et al. “Real-Time Coordination of Plug-In Electric Vehicle Charging in Smart Grids to Minimize Power Losses and Improve Voltage Profile”. In: *IEEE Transactions on Smart Grid* 2.3 (2011), pp. 456–467. DOI: `10.1109/TSG.2011.2159816`.

[8] Yunfan Gao et al. “Retrieval-augmented generation for large language models: A survey”. In: *arXiv preprint arXiv:2312.10997* 2.1 (2023), p. 32.

[9] A. Gatt and E.J. Krahmer. “Survey of the State of the Art in Natural language Generation: Core tasks, applications and evaluation”. English. In: *Journal of Artificial Intelligence Research* 61.1 (2018), pp. 65–170. DOI: `10.1613/jair.5477`.

[10] Zorik Gekhman et al. “Does Fine-Tuning LLMs on New Knowledge Encourage Hallucinations?” In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. Ed. by Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 7765–7784. DOI: `10.18653/v1/2024.emnlp-main.444`. URL: `https://aclanthology.org/2024.emnlp-main.444/`.

[11] Julian de Hoog et al. “Optimal charging of electric vehicles taking distribution network constraints into account”. In: *2017 IEEE Power Energy Society General Meeting*. 2017, pp. 1–1. DOI: `10.1109/PESGM.2017.8273990`.

[12] Lei Huang et al. “A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions”. In: *ACM Trans. Inf. Syst.* 43.2 (Jan. 2025). ISSN: 1046-8188. DOI: `10.1145/3703155`. URL: `https://doi.org/10.1145/3703155`.

[13] Ziwei Ji et al. “Survey of Hallucination in Natural Language Generation”. In: *ACM Comput. Surv.* 55.12 (Mar. 2023). ISSN: 0360-0300. DOI: `10.1145/3571730`. URL: `https://doi.org/10.1145/3571730`.

[14] Jenia Kim, Henry Maathuis, and Danielle Sent. “Human-centered evaluation of explainable AI applications: a systematic review”. In: *Frontiers in Artificial Intelligence* Volume 7 - 2024 (2024). ISSN: 2624-8212. DOI: `10.3389/frai.2024.1456486`. URL: `https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2024.1456486`.

[15] Jannis Kurtz, Ş. İlker Birbil, and Dick den Hertog. “Counterfactual explanations for linear optimization”. In: *European Journal of Operational Research* 329.1 (2026), pp. 24–41. ISSN: 0377-2217. DOI: `https://doi.org/10.1016/j.ejor.2025.06.016`. URL: `https://www.sciencedirect.com/science/article/pii/S0377221725004886`.

[16] Patrick Lewis et al. “Retrieval-augmented generation for knowledge-intensive NLP tasks”. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. NIPS ’20. Vancouver, BC, Canada: Curran Associates Inc., 2020. ISBN: 9781713829546.

[17] Lei Li, Yongfeng Zhang, and Li Chen. “Generate Neural Template Explanations for Recommendation”. In: CIKM ’20. Virtual Event, Ireland: Association for Computing Machinery, 2020, pp. 755–764. ISBN: 9781450368599. DOI: `10.1145/3340531.3411992`. URL: `https://doi.org/10.1145/3340531.3411992`.

[18] Scott M Lundberg and Su-In Lee. “A Unified Approach to Interpreting Model Predictions”. In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017. URL: `https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf`.

[19] Tim Miller. “Explanation in artificial intelligence: Insights from the social sciences”. In: *Artificial Intelligence* 267 (2019), pp. 1–38. ISSN: 0004-3702. DOI: `https://doi.org/10.1016/j.artint.2018.07.007`. URL: `https://www.sciencedirect.com/science/article/pii/S0004370218305988`.

[20] Zeinab Moghaddam et al. “Smart Charging Strategy for Electric Vehicle Charging Stations”. In: *IEEE Transactions on Transportation Electrification* 4.1 (2018), pp. 76–88. DOI: `10.1109/TTE.2017.2753403`.

[21] G.R. Chandra Mouli, P. Bauer, and M. Zeman. “System design for a solar powered electric vehicle charging station for workplaces”. In: *Applied Energy* 168 (2016), pp. 434–443. ISSN: 0306-2619. DOI: `https://doi.org/10.1016/j.apenergy.2016.01.110`. URL: `https://www.sciencedirect.com/science/article/pii/S0306261916300988`.

[22] Ehud Reiter and Robert Dale. *Building Natural Language Generation Systems*. Studies in Natural Language Processing. Cambridge University Press, 2000.

[23] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. “"Why Should I Trust You?": Explaining the Predictions of Any Classifier”. In: KDD ’16. San Francisco, California, USA: Association for Computing Machinery, 2016, pp. 1135–1144. ISBN: 9781450342322. DOI: `10.1145/2939672.2939778`. URL: `https://doi.org/10.1145/2939672.2939778`.

[24] Omid Sadeghian et al. “A comprehensive review on electric vehicles smart charging: Solutions, strategies, technologies, and challenges”. In: *Journal of Energy Storage* 54 (2022), p. 105241. ISSN: 2352-152X. DOI: `https://doi.org/10.1016/j.est.2022.105241`. URL: `https://www.sciencedirect.com/science/article/pii/S2352152X22012403`.

[25] Ahmed M. Salih et al. "A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME". In: *Advanced Intelligent Systems* 7.1 (2025), p. 2400304. DOI: `https://doi.org/10.1002/aisy.202400304`. eprint: `https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/aisy.202400304`. URL: `https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/aisy.202400304`.

[26] Pedram Salimi et al. "Towards Feasible Counterfactual Explanations: A Taxonomy Guided Template-Based NLG Method". In: *ECAI 2023*. IOS Press, Sept. 2023. ISBN: 9781643684376. DOI: `10.3233/faia230499`. URL: `http://dx.doi.org/10.3233/FAIA230499`.

[27] Nicolas Scharowski et al. "Exploring the effects of human-centered AI explanations on trust and reliance". In: *Frontiers in Computer Science* Volume 5 - 2023 (2023). ISSN: 2624-9898. DOI: `10.3389/fcomp.2023.1151150`. URL: `https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2023.1151150`.

[28] Eric Sortomme and Mohamed A. El-Sharkawi. "Optimal Scheduling of Vehicle-to-Grid Energy and Ancillary Services". In: *IEEE Transactions on Smart Grid* 3.1 (2012), pp. 351–359. DOI: `10.1109/TSG.2011.2164099`.

[29] Brígida Teixeira et al. "Explainable AI framework for reliable and transparent automated energy management in buildings". In: *Energy and Buildings* 347 (2025), p. 116246. ISSN: 0378-7788. DOI: `https://doi.org/10.1016/j.enbuild.2025.116246`. URL: `https://www.sciencedirect.com/science/article/pii/S0378778825009764`.

[30] Sahil Verma et al. "Counterfactual Explanations and Algorithmic Recourses for Machine Learning: A Review". In: *ACM Comput. Surv.* 56.12 (Oct. 2024). ISSN: 0360-0300. DOI: `10.1145/3677119`. URL: `https://doi.org/10.1145/3677119`.

[31] Sandra Wachter, Brent Mittelstadt, and Chris Russell. "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR". In: *arXiv e-prints*, arXiv:1711.00399 (Nov. 2017), arXiv:1711.00399. DOI: `10.48550/arXiv.1711.00399`. arXiv: `1711.00399 [cs.AI]`.

[32] Qinglong Wang et al. "Smart Charging for Electric Vehicles: A Survey From the Algorithmic Perspective". In: *IEEE Communications Surveys Tutorials* 18.2 (2016), pp. 1500–1517. DOI: `10.1109/COMST.2016.2518628`.

[33] Zhiwei Xu et al. "A Hierarchical Framework for Coordinated Charging of Plug-In Electric Vehicles in China". In: *IEEE Transactions on Smart Grid* 7.1 (2016), pp. 428–438. DOI: `10.1109/TSG.2014.2387436`.