

A growing body of literature has documented that unreliable electric vehicle (EV) charging poses a major barrier to public infrastructure for climate mitigation. However, prior methods for detecting reliability have been inadequate for revealing regional and social disparities in EV charging reliability at scale. This study addresses that gap by developing a machine learning pipeline that analyzes 838,785 unstructured consumer reviews to uncover critical disparities in EV charging performance. We demonstrate that zero-shot and few-shot learning with iterative expert prompts substantially reduces training data needs, achieving new performance benchmarks for domain-aware reliability detection (F1 score = 0.97, SD = 0.02). We further show how station reliability detection can be combined with diversity indices (Shannon and Simpson indices) for spatial analysis to inform economic and policy decision-making in infrastructure management. Our framework provides credible, evidence-based, and scalable measurement of infrastructure risks, supporting a more reliable and equitable transition to electric mobility.

Unreliable public chargers have become an epidemic, posing a fundamental barrier to EV adoption. This issue has drawn attention in major news outlets such as Forbes (2023), Bloomberg (2023), and the Wall Street Journal (2023), and in academic studies finding that a significant percentage of public chargers fail across the U.S., Europe, and Asia (Rempel et al., 2024; Liu et al., 2022, 2023; Karanam et al., 2023). Overcoming this barrier is crucial as vehicle electrification is vital for decarbonizing transportation, which is the second-largest source of global emissions and the leading contributor in many developed countries (UNEP, 2024; U.S. Environmental Protection Agency, 2025). Expanding EV charging infrastructure is a cost-effective way to boost adoption through indirect network effects (Li et al., 2017; Springel, 2021; Cole et al., 2023; Asensio et al., 2025), yet the success of this strategy is substantially undermined by poor charger reliability. The urgency is further amplified by the rapid growth of the global EV charging market, which is projected to increase from USD 32.26 billion in 2024 to USD 125.39 billion by 2030, at a compound annual growth rate of 25.5%.

Machine learning has emerged as a key strategy for optimizing EV charging management, including algorithm-based decision-making, load balancing, and demand forecasting (Yaghoubi et al., 2024; Zhang et al., 2024). Scalable measurement remains challenging due to poor data interoperability. Most stations lack sub-metering, limiting access to transaction-level data, while the decentralized growth of infrastructure has produced siloed and fragmented datasets with incompatible standards. Without mandatory reporting, providers have little incentive to share or standardize data, and consumers lack real-time charging information, which discourages adoption. Conventional methods such as dashboard data, simulations, and surveys are limited in scale and fail to capture user experiences, including unsuccessful charging attempts. Some studies integrate consumer perspectives: Asensio et al. (2020) and Yu et al. (2025) analyzed sentiment about charging experiences, and Ha et al. (2021) purely classified review topics. However, detecting reliability issues automatically from consumer data is difficult because failure modes span dozens of imbalanced classes, and problems are not always explicitly or negatively stated. Prior models struggle to distinguish access issues or congestion from operational failures such as hardware, software, or transaction errors. As a result, previous methods require extensive expert training yet achieve only modest F1 scores between 0.45 and 0.88.

In this paper, we examine whether zero-shot and few-shot learning with large language models can reduce costs and improve accuracy in detecting charging reliability. We illustrate how machine learning, combined with geographic performance indices, captures regional and social disparities in consumer charging experiences. Building on earlier classification strategies (Asensio et al., 2020, 2025; Ha et al., 2021), we incorporate iterative expert feedback into prompt design, substantially reducing Type I and II errors and setting new benchmarks in this domain. We further demonstrate how machine learning approaches for station reliability detection can be combined with diversity indices for spatial analysis to inform economic and policy decision-making. For climate mitigation strategies in the transportation sector, we find that EV infrastructure reliability is currently highest in rural areas and less populated communities, whereas reliability issues are widespread in urban centers and metropolitan areas. In recent years, federal policies have prioritized charger installation at 50-mile intervals along designated EV corridors (Hanig et al., 2025). We have uncovered those EV corridors not only have low reliability issues but also have the widest spread of station reliability. Current investment incentives focus on deployment rather than operational reliability, leaving challenges persistent without further incentives or policy interventions. Our contributions include generating behaviorally informed predictions of charging reliability from consumer voices, quantifying disparities overlooked by climate policy, and developing a scalable framework for ensemble learning.