# A Generative AI Approach to Pricing Mechanisms and Consumer Behavior in the Electric Vehicle Charging Market

**Sarthak Chaturvedi[1], Edward W. Chen[1], Ila P. Sharma[1], Omar Isaac Asensio[1, 2*]**

[1]Georgia Institute of Technology
[2]Harvard Business School

sarthak@gatech.edu, echen307@gatech.edu, isharma37@gatech.edu, asensio@gatech.edu

**Abstract**

The electrification of transportation is a growing strategy to reduce mobile source emissions and air pollution globally. To encourage adoption of electric vehicles, there is a need for reliable evidence about pricing in public charging stations that can serve a greater number of communities. However, user-entered pricing information by thousands of charge point operators (CPOs) has created ambiguity for large-scale aggregation, increasing both the cost of analysis for researchers and search costs for consumers. In this paper, we use large language models to address standing challenges with price discovery in distributed digital data. We show that generative AI models can effectively extract pricing mechanisms from unstructured text with high accuracy, and at substantially lower cost of three to four orders of magnitude lower than human curation (USD 0.006 pennies per observation). We exploit the few-shot learning capabilities of GPT-4 with human-in-the-loop feedback—beating prior classification performance benchmarks with fewer training data. The most common pricing models include free, energy-based (per kWh), and time-based (per unit time), with tiered pricing (variable pricing based on usage) being the most prevalent among paid stations. Behavioral insights from a US nationally representative sample of 13,008 stations suggest that EV users are commonly frustrated with the slower than expected charging rates and the total cost of charging. This study uncovers additional consumer barriers to charging services concerning the need for better price standardization.

## Introduction

The transportation sector is one of the largest contributors to greenhouse gas and air pollutant emissions in the United States and globally (IEA, 2023; EPA, 2021). The adoption of electric vehicles (EVs) reduces tailpipe emissions and provides air quality co-benefits to communities (Asensio et al., 2020; Carley et al., 2019; Requia et al., 2018; Sheldon, 2022). A variety of pricing strategies can serve as innovation policy levers to increase EV diffusion by supporting new charging business models and making life cycle fueling costs more salient or economically accessible to a broader set of consumers (Asensio et al. 2021). Under current decentralized models of EV infrastructure growth, thousands of individual CPOs and managers set their own pricing and access policies. For climate research communities, this creates long standing data challenges for archiving, sharing, and extracting insights from massively distributed digital data (Knusel et al. 2020; Faghmous and Kumar, 2014; Overpeck et al., 2011). Inadequate information about alternative fueling costs is known to lead to sub-optimal consumer decisions (Allcott, 2013; Larrick and Sol, 2008). Further, recent theory and evidence from the US and Europe indicates that it is more cost effective to promote incentives for infrastructure provision, rather than car sales, due to well-known charging network externalities (Li et al., 2017; Springel, 2021; Cole et al., 2023; Liang et al., 2023).

In recent years, the emergence of open-source EV charging platforms offers new opportunities to investigate consumer behavior and price phenomena in near real-time, using unstructured natural language data (Asensio et al., 2020; Ha et al., 2021; Liu et al., 2023). These data from mobile platforms overcome limitations of government surveys and self-reported data by providing direct market evidence. However, there remain two fundamental challenges for discovery. First, pricing information on these platforms is user-entered free-form text by station operators and managers, leading to inconsistencies and ambiguity. Further, interoperability issues exist between charging networks, making it difficult to extract price data at a large scale. Second, high search costs also make it difficult to observe consumer responses to various price schemes.

In this paper, we investigate the plausibility of using large language models to solve these data challenges regarding scale and cost. Developing automated approaches using generative AI is important because understanding large-scale coordination of price behavior is essential to managing the smart adoption of EVs and directed investments in public infrastructure. Here we demonstrate the capabilities of the zero-shot learning of GPT-4 (OpenAI, 2023) with expert prompting to extract price scheme information at a national level. Prior research has demonstrated that transformers and other neural networks (i.e., CNNs, RNNs, and BERT) can

do well in sentiment classification tasks within this domain but require expensive, high quality training data for policy analysis. We leverage the few-show learning capabilities of GPT-3 (Brown et al., 2020) to classify consumer review sentiments with significantly fewer training examples compared to prior benchmarks, while still understanding domain specific vocabulary and achieving comparable performance with reduced cost. We characterize the relative importance of common pricing strategies to reveal the distribution of pricing decisions in the US. Additionally, we generate a measure of consumer responses within different pricing schemes resolved at the station level. Such capabilities transform the cost and scale of system-level research in engineering and climate-oriented research in transportation and mobility. We evaluate performance and comment on the relative cost of human versus AI assisted price discovery.

# Methods

## Data Collection

We collected data on EV stations from open-source web applications, such as PlugShare and ChargePoint, for the years 2020 and 2022. The data included unique station ID, geographic location, pricing scheme, parking information, consumer reviews, and date of the reviews. The raw data contained 25,607 and 22,208 stations from 2020 and 2022, respectively. We observed that 3,339 stations that existed in 2020 were no longer available in 2022. To examine the possible attrition bias, we performed a two-sample Kolmogorov–Smirnov test to compare the distributions of station counts per state between 2020 and 2022 datasets. We found no statistically significant difference between the distributions ($P > 0.05$). Thus, we ruled out the attrition bias as a factor in our analysis. In 2020, we collected over 38,540 reviews from 10,686 stations. In 2022, the number of reviews increased significantly to 71,330 from 13,008 stations. Given generative AI models have been known to have memorization tendencies of private sensitive information (Biderman et al., 2023) and socially harmful behavior from real world observations (Ganguli et al., 2022), we omitted personally identifiable information (PII) from the consumer reviews and station operators.

## Pricing Taxonomy

We defined five broad pricing categories based on the literature, domain knowledge, and hundreds of stations cost descriptions. Table 1 presents the price scheme, definition, and brief ex-ample. We found that the vast majority of station price schemes (e.g. 99.9%) fell into one of the mutually exclusive models: energy-based, time-based, tiered, or free. Other price schemes such as subscription or other form of payment were rare.

| Price Scheme | Definition | Example |
|---|---|---|
| Energy-based | Cost to charge at either per kWh basis | "$0.35/kWh" |
| Time-based | Cost to charge at per unit of time basis, such as dollars per minute, hr, etc. | "$2.00/Hr" |
| Tiered | Cost varies depending on extend of charging session duration either in terms of time or energy | "Charging is $0.80/hour for first 4 hours. $5/hour after that. Paid parking" |
| Free | No cost for charging | "9/ day parking plus tax. Free to charge" |
| Other | Subscription or other form of payment | "$4/month unlimited wind powered charging, pluginaustin.com" |

Table 1. Taxonomy of Pricing Schemes in EV Stations

## Pricing Scheme Extraction

We experimented with rule-based strategies to extract price schemes using regular expressions as a direct way to programmatically assign stations to a category in our price scheme taxonomy. We used 30 regular expressions to extract valuable price-specific data and categorize it into the associated pricing scheme category. However, we found that regular expressions were not sufficient to capture all the possible variations and nuances of user-entered pricing schemes. For example, some stations had different prices for different plug types or different times of the day. Some stations had conditional prices based on membership or subscription status. Some stations had no explicit price but required a donation or a purchase at a nearby store. Moreover, some pricing schemes were too complex or ambiguous to be parsed by simple rules.

To overcome these limitations, we used GPT-4 (OpenAI, 2023) as an alternative approach to test whether it could handle the widely varying user-entered text. Prior research has demonstrated the use of automated or human led methods for prompt generation that can serve as an effective alternative for model fine tuning (Shin et al., 2020; Meyer et al., 2022; Møller et al., 2023). We used chain-of-thought reasoning (Wei et al., 2022; Chen et al., 2023) to create expert prompts that included labeled definitions and examples of pricing schemes with more complex reasoning that distinguishes between charging and parking costs. We used these

prompts to guide GPT-4 to capture the nuances and variations in the user-entered pricing schemes and reviews. The prompt that yielded the best performance was:

"Write a word that describes the charging rate of an electric vehicle (EV) station. The word should be based on the price and whether it is energy based or time based. If there are multiple charging rates, the word should indicate that it is tiered. Do not include parking fees or any other fees except the charging rate. If there is no charging rate, the word should be free. For example:

- $0.25/kWh -> energy
- $1/hour -> time
- $0.50/kWh for first 10 kWh, then $0.25/kWh -> tiered
- No charge -> free"

We applied this prompt to each pricing record in our data set and obtained a standardized category label from GPT-4. To evaluate performance, we compared the results of GPT-4 with those of regular expressions and found that GPT-4 was able to handle more complex and diverse pricing schemes with higher accuracy and consistency. In the example provided in Table 2, we show examples where GPT-4 successfully identifies the true price scheme even when there are additional parking costs, time limits or seasonality, non-linear pricing, or other comparative price schemes.

### Sentiment Prediction with Fine-Tuned Models

Following price scheme extraction, we used GPT-3 available via the Azure OpenAI service to classify the sentiments of the user reviews for each charging station, which allowed for modified tuning. We compared two GPT-3 models, DaVinci and Curie, that can perform sentiment classification using few-shot learning techniques (Brown et al., 2020) and we found that the DaVinci model to be the most capable with our dataset. We chose Curie as our preferred model because it was more cost-effective and efficient, and it achieved similar results to DaVinci. We used a standard sentiment lexicon (Liu et al., 2005) to assign positive or negative labels to the user reviews. We then fine-tuned Curie on a small subset of expert labeled reviews (about 10% of the total reviews) to learn the domain-specific language and expressions used by the EV users. We devised prompts using labeled definitions from consumer reviews in the curated dataset in Ha et al., 2021 (Ha et al., 2021). We applied these prompts to each user review in our sample and obtained a sentiment prediction with fine-tuned, expert-trained GPT-3. We evaluated our methods using standard metrics such as accuracy, precision, recall, and F1-score for sentiment classification tasks. We also compared our extracted pricing scheme results with several baselines such as rule-based methods like RegEx (Friedl, 2006) and sentiment prediction

| Charging Operator Price Example | True Class | GPT-4 Class | Regex Class |
|---|---|---|---|
| "Parking: First 4 hrs free. Thereafter $2.00 per hour. Charging: $0.25/kWh" | Energy | Energy | Tiered |
| "Charging always free. General metered parking 8am to 9pm, $2.00/hour, 2 hour time limit. Free parking 9pm to 8am and in winter months" | Free | Free | Time |
| "$1 first 4 hours $3/hr afterward" | Tiered | Tiered | Time |
| "$1.50 minimum, $0.35/kWh same as other Disney ChargePoint stations" | Energy | Energy | Tiered |
| "Parking costs $5/hr. $1.25 per hour charging" | Time | Time | Tiered |
| "Energy: All Days: $0.13/kWh Station Parking: First 1 hr(s) Free, thereafter $0.10/hr" | Energy | Energy | Tiered |

Table 2. Complex Price Examples

results with well-known deep learning classifiers such as CNN (Kim, 2014).

To enable station level analysis over time from 2020 to 2022, we computed the negativity score for each station as a measure of customer dissatisfaction, using protocols in Asensio et al., 2020 (Asensio et al., 2020). The negativity score, bounded between 0 and 1, for any station *i* in a given year *t* is defined as:

$$\text{Negativity Score}_{i,t} = \frac{\text{Count of negative reviews}_{i,t}}{\text{Total count of reviews}_{i,t}} \quad (1)$$

To understand possible variations in sentiment across different pricing schemes, we grouped the stations based on the pricing categories determined from the previous step. Furthermore, we excluded the stations with less than two reviews to avoid any potential bias in the sentiment scores. We then used the negativity scores to understand station level phenomena across different price schemes between 2020 and 2022.

## Results and Discussion

### Distribution of Pricing Schemes

We evaluated the performance of GPT-4 with prompt engineering for extracting standardized pricing schemes from user-entered data and compared it with a baseline Regex approach. We randomly sampled 4 sets of 50 stations from our dataset, which includes 100 reviews per set, and manually annotated their price scheme class. GPT-4 achieved a 2 to 26 percentage point improvement in accuracies for extracting pricing schemes (Table 3). This demonstrates that GPT-4 easily outperforms Regex in replication tests with fresh data, resulting in 2.4 to 38.2% improvement from the base-

| Sample | Regex | GPT-4 | No. Stations | Sampled Reviews |
|---|---|---|---|---|
| 1 | 67% | 90% | 50 | 100 |
| 2 | 64% | 86% | 50 | 100 |
| 3 | 85% | 87% | 50 | 100 |
| 4 | 68% | 94% | 50 | 100 |

Table 3. Performance of Regex vs. GPT-4 and Replication Accuracy

line using regular expressions.

Although GPT-4 consistently outperforms Regex with lower misclassification errors, i.e., 6 to 14% error rate as shown in Table 3, we acknowledge that there could exist cases where Regex identifies the correct price scheme while GPT-4 does not. We performed additional experiments to quantify such cases and found that these were rare, with only 6 out of 400 (1.5%) that met this criterion. Examples of such cases are shown in Table 4. Thus, we conclude that GPT-4 is a preferred technical solution to price discovery from unstructured user text in digital platforms.

Given our large-scale data and highly accurate classification with GPT-4, we are able to generate the distribution of pricing schemes across all major charging networks in the US. We used the taxonomy of price schemes defined in Table 1 to classify the user-entered data. Figure 1 shows the distribution of pricing scheme in the year 2022. We find that the most common pricing scheme is free (72.6%), followed by tiered (11.8%), time (8.2%), and energy-based (7.4%) respectively. Although a majority of stations do not include charging fees, many free stations could have restricted access, i.e. such as to employees or to customers, and/or significant parking fees. For example, in Table 2, the price description: "Charging always free. General metered parking 8am to 9pm, $2.00/hour, 2 hour time limit. Free parking 9pm to 8am and in winter months", was successfully identified by GPT-4 as a free station despite also having detailed parking rates. We note that other price schemes not in one of the four major price schemes only make up 0.1% of parking rates

| Example | True Class | GPT-4 Class | Regex Class |
|---|---|---|---|
| "Fast Charger: $0.25 Min / Level2 Charger: $1 per hour max $10 Session fee" | Time | Tiered | Time |
| "DC Fast Charge $21hr, 35¢/minute prorated- L2 $1.50/hour, 2.5¢/m" | Time | Tiered | Time |
| "Free 2 hours limit" | Free | Time | Free |
| "$1/hr, $2 min., $4 max." | Time | Tiered | Tiered |

Table 4. Examples of GPT-4 Misclassification

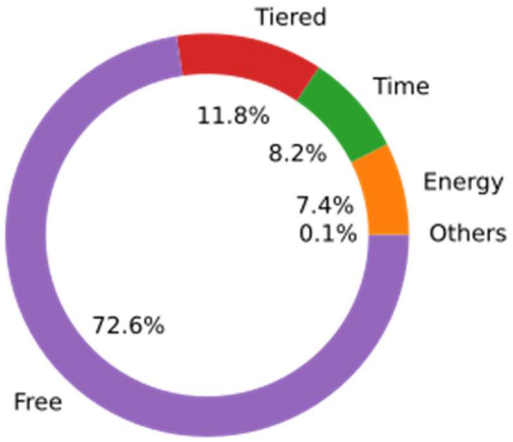


Figure 1. Distribution of pricing schemes for a national sample of U.S. charging stations in 2022 (N=10,527)

## Classifier Performance

We evaluate the performance of GPT-3/4 versus several benchmarks for sentiment classification on a national sample of EV user reviews. The results of our experiments show that the GPT-3 (Curie) model that has been fine-tuned with expert annotated labels outperformed the benchmark convolutional neural network (CNN) model (Asensio et al., 2020) by 5 percentage points in accuracy and 0.04 points in F1 score (Table 5). This is notable because GPT-3 achieves this performance with only 13% of high-quality training examples (i.e., F1 Score 0.894 (s.d. 0.0048) for GPT-3 versus 0.86 (s.d. 0.0037) for CNN). Given the impressive performance with few-shot learning we also conducted additional experiments to investigate whether GPT-4 could further reduce dependence on high quality expert annotated data. In a series of prompting experiments where we provided GPT-4 with contextual definitions and relevant examples (Ha et. al., 2021), we show that the GPT-4 can achieve F1 scores in the similar range of approximately 0.88 with zero-shot learning (i.e., F1 Score 0.877 (s.d. 0.0018) for GPT-4 versus 0.894 (s.d. 0.0048) for GPT-3). These findings suggest that fine tuning the GPT-3 model results in slightly better performance versus other existing zero-shot learning strategies. In future work, we suggest evaluation of other zero-shot transformer models to evaluate its use for other multi-labelled topic classification. These results indicate model fine-tuning and expert augmentation are still the dominant pathways to achieving the most effective classification in this domain. These results add to a growing body of literature of using context specific and general-purpose models in climate change and other social science domains (Veghefi et al., 2023; Savelka et al., 2023; Abdelghani et al., 2023).

| Model | Annotation | No. Training | Acc % (s.d.) | F1 (s.d.) |
|---|---|---|---|---|
| GPT-4 † | None | 0 | 87.70 (0.17) | 0.877 (0.0018) |
| GPT-3 (Curie) | Expert Annotated | 1212 | 89.59 (0.54) | 0.894 (0.0048) |
| CNN * | Human Annotated | 8953 | 84.70 (0.80) | 0.860 (0.0037) |

Table 5. Classifier Metrics for Sentiment Analysis
* (Asensio et al., 2020). Benchmark Model
† Neutral classifications were not considered for consistency with other benchmark models

To generate uncertainty estimates for GPT model performance, we conducted 10 experimental replications creating 84/13/3 random data splits into testing, training, and validation, respectively. The sentiment classification task was to predict a "Positive" or "Negative" label for each review. We assessed model performance using the macro-averaged F1 score. This evaluates the F1 score independently for each class, taking the average to account for class imbalance. Our fine-tuned GPT-3 model achieved a macro F1 score of 0.894 (s.d. 0.0048), while the GPT-4 zero-shot model achieved a macro F1 score of 0.877 (s.d. 0.0018), indicating comparable predictive performance for each model. This is notable since GPT-4 uses zero training examples and outperforms the prior benchmark with human annotated CNNs (Table 5).

## Comparing AI vs Human Costs

Given the strong performance, we evaluated the potential reductions in research evaluation costs between AI-driven classification and human expert annotation. To calculate the human costs, we assume that each annotator is paid minimum wage (currently Massachusetts minimum wage of $15/hr) and that they can annotate between 2-3.3 labels per minute, based on our human-labelling experiments (Asensio et al., 2020, Ha et al., 2021, Liu et al., 2023). Based on this, we estimate research evaluation costs to range from 75,000 to 125,000 USD. For the expert trained AI costs, we assume 9,000 prompt tokens for training, and 1,000k sampled tokens for predictions. For GPT-4 models with 8,000 contexts lengths (e.g. GPT-4 or GPT-4-0314), the current price is $0.03/1k prompt tokens and $0.06/1k sampled tokens. The AI costs are a modest 60.27 USD, which are three to four order of magnitude lower than human annotation.

## Sentiment Analysis by Pricing Models

Given that we have pricing information resolved at the station level, we analyzed the large-scale consumer sentiment for the years 2020 and 2022 by pricing scheme. To conduct this analysis, we computed the average sentiment score for each pricing model by computing the negativity score at each station ID, defined in Equation 1. The negativity score ranges from 0 to 1, where 0 means that all listed reviews per station per year are classified as positive and 1 means all listed reviews per station per year are classified as negative. We summarized the negativity scores at each station location for the most prominent pricing categories for 2020 and 2022 (Figure 2). With the exception of stations with time-based pricing, we find all other pricing models are associated with an increase in negative consumer sentiment at EV chargers.

Because correlated external factors or latent variables can influence consumer sentiment and pricing perceptions, we conducted post-ML regressions by spatially merging geolocated stations with observable social and economic variables from the U.S. Census and other publicly available sources to adjust the negativity scores. We used a generalized linear model to regress the station negativity scores on input factors such as regional policies (e.g. county-level indicator for metro area), economic conditions (e.g. annual county unemployment rate and household median income as a proportion of a median state income), technological advancements (e.g. networked v.s. non-networked and number of types of connector techonologies), location amenties such as point of interests

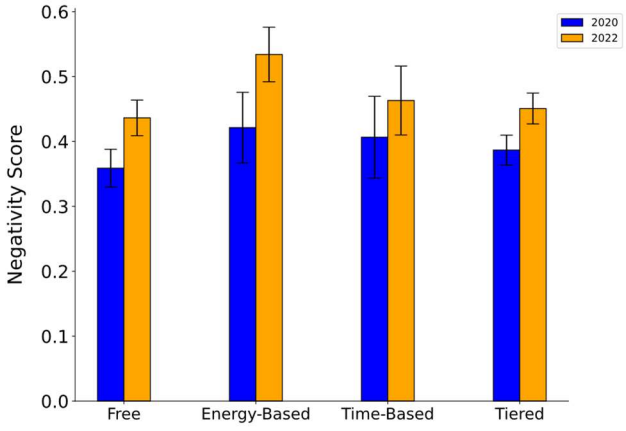


Figure 2. Consumer sentiment by pricing scheme
We find evidence of increasing negative consumer sentiment for free, energy-based, and tiered pricing schemes

(e.g. restaurant, hotels, stores), and a proxy for other potential unobserved factors (e.g. station rating).

Table 6 summarizes the changes in negativity scores from 2020 to 2022 by pricing schemes, while statistically adjusting for external factors. We find that stations are generally associated with statistically significant increases in negative consumer sentiment, particularly for energy-based and tiered pricing schemes (Table 6). We clustered the standard errors at different spatial scales including state, county, and location ID to confirm the robustness of estimates at different scales (Table 7). To give further details on the potential biasing effect of latent variables, Table 8 includes detailed point estimates for the observable station characteristics. For example, we find that the number of connector technologies (e.g. Tesla plug, J-1772, and CHAdeMO) and stations rating are generally associated with a decrease in negative consumer sentiment.

In order to reveal possible explanations or mechanisms, we analyzed qualitative evidence from station reviews by price scheme, which suggests that consumers are dissatisfied with payments and the delivered charging rates. For example, a user at a station with energy-based pricing writes: "*Slow and $. The total duration of this session was 01:00:01 and the energy charged was 17.649 kWh. Your total cost for the session was $19.49*." Similarly, consumers also expressed frustration regarding tiered pricing becoming prohibitively expensive over time, for example a user writes: "*$21.65 to charge!!!!!!! Holy moly!!!! Dont come here unless you are desperate!! And its not fast. 32A is what it pulls.*" We also found similar dissatisfaction at free charging stations, possibly due to the quality or availability of complimentary charging facilities, including being out of service, lack of adequate parking, low charging power, and long wait times (Asensio et al. 2020). For example a user at a free station writes: "*Tried 3 times. Faulted each time. Customer service was very helpful but could not get it going.*", implying frustration with the malfunctioning of the station despite the free rate. Although our statistical estimates for the rise in negative sentiment adjust for a comprehensive set of station characteristics and other other factors, there could be additional factors that also influence the negativity score not directly related to pricing. In addition, although price setting is done independently by station operators and in some cases can be considered quasi-random, the price models are not randomly assigned, and therefore our estimates should be interpreted as conditional correlations. In future work, we suggest merging AI-based predictions with randomized or quasi-randomized price assignment, which is the focus of forthcoming research.

| Price Scheme | Estimate |
|---|---|
| Free | 6.79** |
| | (2.31) |
| Energy | 13.60*** |
| | (4.07) |
| Time | 8.32 |
| | (4.42) |
| Tiered | 8.54*** |
| | (1.79) |

Table 6. Post-ML Regression Estimates
Significant at the level p < *0.05, **0.01, ***0.001
Note: Price schemes are set by independent charge point operators and not randomly assigned.

| Clustering Level | Free | Energy | Time | Tiered |
|---|---|---|---|---|
| None | 6.79** | 13.60*** | 8.32 | 8.54*** |
| | (2.31) | (4.07) | (4.42) | (1.79) |
| Location ID | 6.79* | 13.60** | 8.32 | 8.54*** |
| | (2.95) | (4.90) | (5.46) | (1.86) |
| County | 6.79* | 13.60** | 8.32 | 8.54*** |
| | (3.39) | (5.23) | (6.54) | (1.93) |
| State | 6.79** | 13.60* | 8.32 | 8.54*** |
| | (2.43) | (5.72) | (5.82) | (1.92) |

Table 7. Robustness of Regression to Various Clustering Options
Significant at the level p < *0.05, **0.01, ***0.001

## Closing

This study demonstrates how large language models like GPT-4 can overcome limitations of rule-based approaches for complex price scheme extraction, despite challenges regarding inconsistencies and ambiguities with unstructured data. Additionally, fine-tuning GPT-3 with expert-augmentation precisely classifies consumer sentiment and related classification tasks, outperforming benchmarks with minimal labeled data through transfer learning. These methods are highly accurate, scalable, and substantially reduce evaluation costs by three or four orders of magnitude at an estimated cost of USD 0.006 pennies per observation.

Our nationwide analysis reveals heterogeneity in consumer responses to different pricing schemes over time. The results have important policy implications, highlighting opportunities to address pain points through infrastructure improvements and pricing innovation. Currently, free or subsidized charging is the most popular strategy that has been used by employers, managers, and other station hosts as a complimentary benefit to encourage EV adoption and to increase EV miles traveled. The results suggest significant opportunities to implement pricing mechanisms in both free and restricted access stations that can support business model development and efficiently price energy use externalities.

| | Free | Energy | Time | Tiered |
|---|---|---|---|---|
| **Performance period** | | | | |
| Post | 6.79** | 13.60*** | 8.32 | 8.54*** |
| | (2.31) | (4.07) | (4.42) | (1.79) |
| **Station technologies and characteristics** | | | | |
| Networked station | 2.75* | -14.89** | 3.58 | 3.75 |
| | (1.26) | (4.90) | (4.35) | (3.63) |
| No. of user check-ins | 0.00 | 0.02* | 0.04*** | 0.01* |
| | (0.00) | (0.01) | (0.01) | (0.00) |
| No. of connector technologies | -3.91*** | -2.85* | -6.23** | -3.37*** |
| | (0.78) | (1.31) | (1.99) | (0.96) |
| Station rating | -4.38*** | -3.59*** | -5.54*** | -3.86*** |
| | (0.20) | (0.31) | (0.46) | (0.26) |
| **Geographic/economic conditions** | | | | |
| County in metro area | 8.15*** | 0.71 | -1.48 | 3.54* |
| | (1.92) | (2.93) | (3.27) | (1.51) |
| County unemployment rate | -0.18 | -0.51 | 0.54 | 0.57 |
| | (0.43) | (0.71) | (0.82) | (0.32) |
| Median household income as a percent | 0.01 | 0.04 | 0.19 | 0.04 |
| of state total | (0.03) | (0.05) | (0.08) | (0.03) |
| **Location amenities** | | | | |
| Point of interest indicators | Yes | Yes | Yes | Yes |
| No. of stations | 999 | 324 | 236 | 814 |
| No. of reviews | 9,072 | 3,554 | 2,385 | 14,765 |

Table 8. Post ML Regression of Change in Negativity Score 2020-2022
Significant at the level p < *0.05, **0.01, ***0.001


## Acknowledgements

We gratefully acknowledge funding by Microsoft and National Science Foundation Awards #1945532, #1931980, and #2125390. For valuable assistance in public data collection supporting this work, we thank M. Cade Lawson. We also thank discussants at the Text as Data (TADA) Conference at Cornell Tech and the AAAI Fall Symposium on Artificial Intelligence and Climate. For institutional support, we thank Deborah Spar, Drew Keller, Amram Migdal, Barbara DeLollis, Sarah Gazzaniga, and Olivia Barba at the Institute for the Study of Business in Global Society. For high performance computing access, we thank Aaron J Drysdale and Srinivas Aluru at the Institute for Data Engineering and Science. For cloud service engagement on Azure, we thank Bob Breynaert and Elizabeth Bruce at Microsoft.


## Author Contributions

**Conceptualization**: O.I.A.; **Data curation**: S.C., E.W.C., I.P.S.; **Formal analysis**: S.C., E.W.C., I.P.S.; **Funding acquisition**: O.I.A; **Investigation**: S.C., E.W.C., I.P.S., O.I.A.; **Methodology**: S.C., E.W.C., I.P.S., O.I.A.; **Project administration**: O.I.A.; **Resources**: O.I.A.; **Software**: S.C., E.W.C., I.P.S.; **Supervision**: O.I.A.; **Validation**: S.C., E.W.C., I.P.S.; **Visualization**: S.C., E.W.C.; **Writing–original draft**: S.C., E.W.C., I.P.S., O.I.A.; **Writing–review and editing**: S.C., E.W.C., O.I.A..

## References


Abdelghani, R.; Wang, Y.-H.; Yuan, X.; Wang, T.; Lucas, P.; Sauzéon, H.; and Oudeyer, P.-Y. 2023. GPT-3 Driven Pedagogical Agents to Train Children's Curious Question-Asking Skills *International Journal of Artificial Intelligence in Education*.

Allcott, Hunt. 2013. The Welfare Effects of Misperceived Product Costs: Data and Calibrations from the Automobile Market. *American Economic Journal: Economic Policy*, 5(3): 30-66.

Asensio, O. I.; Alvarez, K.; Dror, A.; Wenzel, E.; Hollauer, C.; and Ha, S. 2020. Real-Time Data from Mobile Platforms to Evaluate Sustainable Transportation Infrastructure. *Nature Sustainability*, 3(6), 463-471.

Asensio, O.I.; Apablaza, C.Z.; Lawson, M.C.; and Walsh S.E. 2022. A Field Experiment on Workplace Norms and Electric Vehicle Charging Etiquette. *Journal of Industrial Ecology*, 26(1), 183-196.

Biderman, S.; Prashanth, U. S.; Sutawika, L.; Schoelkopf, H.; Anthony, Q.; Purohit, S.; and Raff, E. Emergent and predictable memorization in large language models. arXiv:2304.11158.

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.;

Askell, A.; Agarwal, S.; Herbert-Voss, A.; Krueger, G.; Henighan, T.; Child, R.; Ramesh, A.; Ziegler, D.; Wu, J.; Winter, C.; Hesse, C.; Chen, M.; Sigler, E.; Litwin, M.; Gray, S.; Chess, B.; Clark, J.; Berner, C.; McCandlish, S.; Radford, A.; Sutskever, I.; and Amodei, D. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, 1877-1901.

Carley, S.; Zirogiannis, N.; Siddiki, S.; Duncan, D.; and Graham, J. D. 2019. Overcoming the Shortcomings of U.S. Plug-In Electric Vehicle Policies. *Renewable and Sustainable Energy Reviews*, 113.

Chen, J.; Chen, L.; Huang, H.; and Zhou T. When do you need Chain-of-Thought Prompting for ChatGPT? arXiv:2304.03262.

Cole, C.; Droste, M.; Knittel, C.; Li, S.; and Stock, J.H. 2023. Policies for Electrifying the Light-duty Vehicle Fleet in the United States. *AEA Papers and Proceedings*, 113, 316-322.

Ganguli, D.; Hernandez, D.; Lovitt, L.; Askell, A.; Bai, Y.; Chen, A.; Conerly, T.; Dassarma, N.; Drain, D.; Elhage, N.; El Showk, S.; Fort, S.; Hatfield-Dodds, Z.; Henighan, T.; Johnston, S.; Jone., A.; Joseph, N.; Kernian, J.; Kravec, S.; Mann, B.; Nanda, N.; Ndousse, K.; Olsson, C.; Amodei, D.; Brown, T.; Kaplan, J.; McCandlish, S.; Olah, C.; Amodei, D.; and Clark, J. 2022. Predictability and Surprise in Large Generative Models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 1747-1764.

EPA. 2021. Inventory of U.S. Greenhouse Gas Emissions and Sinks. https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks. Accessed: 2023-07-27.

Faghmous, J.H. and Kumar, V. 2014. A Big Data Guide to Understanding Climate Change: The Case for Theory-Guided Data Science. *Big Data*, 155-163.

Friedl, J. E. 2006. *Mastering Regular Expressions*. O'Reilly Media, Inc.

Ha, S.; Marchetto, D.J.; Dharur, S.; and Asensio, O.I. 2021. Topic Classification of Electric Vehicle Consumer Experiences with Transformer-Based Deep Learning. Patterns, 2(2), 100195.

IEA. 2023. Trends in electric light-duty vehicles. https://www.iea.org/reports/global-ev-outlook-2023/trends-in-electric-light-duty-vehicles. Accessed 2023-07-27.

Kim, Y. 2014. Convolutional neural networks for sentence classification. arXiv:1408.5882.

Knüsel, B.; Zumwald, M.; Baumberger, C.; Hirsch Hadorn, G.; Fisher, E.M.; Bresch, D.N.; Knutti, R. 2019. Applying Big Data Beyond Small Problems in Climate Research. *Nature Climate Change*, 9, 196–202.

Larrick, R.P., and Soll, J.B. 2008. The MPG Illusion. *Science*, 320, 1593-1594.

Li, S.; Tong, L.; and Zhou, Y. 2017. The Market for Electric Vehicles: Indirect Network Effects and Policy Design. *Journal of Association of Environmental and Resources Economists*, 4(1), 89-133.

Liang, J.; Qiu, Y.; Liu, P.; He., P.; and Mauzerall, D.L. 2023. Effects of Expanding Electric Vehicle Charging Stations in California on the Housing Market. *Nature Sustainability* 6, 549–558.

Liu, B.; Hu, M.; and Cheng, J. 2005. Opinion Observer: Analyzing and Comparing Opinions on the Web. *In Proceedings of the 14th international conference on World Wide Web*, 342-351.

Liu, Y.; Francis, A.; Hollauer, C.; Lawson, M. C.; Shaikh, O.; Cotsman, A.; Bhardwaj, K.; Banboukian, A.; Li, M.; Webb, A.; and Asensio, O. I. 2023. Reliability of Electric Vehicle Charging Infrastructure: A Cross-Lingual Deep Learning Approach. *Communications in Transportation Research*, 3, 100095.

Møller, A. G.; Dalsgaard, J. A.; Pera, A.; and Aiello, L. M. 2023. Is a Prompt and a Few Samples All You Need? Using GPT-4 for Data Augmentation in Low-Resource Classification Tasks. arXiv:2304.13861.

OpenAI. 2023. GPT-4. https://openai.com/research/gpt-4. Accessed 2023-07-28.

Overpeck, J.T.; Neehl, G.A.; Bony, S.; and Easterling, D.R. 2011. Climate Data Challenges in the 21st Century. *Science*, 331, 700-702.

Santoyo, C.; Nilsson, G.; and Coogan, S. 2021. Sensitivity of Electric Vehicle Charging Facility Occupancy to Users' Impatience. 2021. In *60th IEEE Conference on Decision and Control*, 2862-2867.

Requia, W.J., Mohamed, M., Higgins, C.D., Arain, A., and Ferguson, M. 2018. How clean are electric vehicles? Evidence-based review of the effects of electric mobility on air pollutants, greenhouse gas emissions and human health. *Atmospheric Environment*, 185, 64-77.

Savelka, J.; Ashley, K. D.; Gray, M. A.; Westermann, H.; and Xu, H. 2023. Explaining legal concepts with augmented large language models (GPT-4) arXiv:2306.09525.

Sheldon T.L. 2022. Evaluating Electric Vehicle Policy Effectiveness and Equity. *Annual Review of Resource Economics*, 14, 669-688.

Shin, T.; Razeghi, Y.; Logan IV, R. L.; Wallace, E.; and Singh, S. 2020. Autoprompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 4222-4235.

Si, C.; Gan, Z.; Yang, Z.; Wang, S.; Wang, J.; Boyd-Graber, J.; and Wang, L. 2023. Prompting GPT-3 To Be Reliable. In *International Conference on Learning Representations*.

Springel, K. 2021. It's Not Easy Being "Green": Lessons from Norway's Experience with Incentives for Electric Vehicle Infrastructure. *Review of Environmental Economics and Policy*, 15, 352-359.Vaghefi, S.A.; Wang, Q.; Muccione, V.; Ni, J.; Kraus, M.; Bingler, J.; Schimanski, T.; Colesanti-Senni, C.; Webersinke, N.; Huggel, C.; and Leippold, M. 2023. chatClimate: Grounding conversational AI in climate science. arXiv:2304.05510

Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Ichter, B.; Xia, F.; Chi, E.; Le, Q. V.; and Zhou, D. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In *Advances in Neural Information Processing System*