Dissecting Electric Vehicle Conversations on Twitter/X: A Tree-Based Analysis of Discussions

Nina Varchavsky–Bergin[1], Stephany Rajeh[1,2], Guillaume Vincent[3], Faten Chaieb-Chakchouk[1]
[1]Efrei Research Lab, Université Paris-Panthéon-Assas
[2]LIP6, Sorbonne Université
[3]Carism, Université Paris-Panthéon-Assas
nina.varchavsky-bergin@assas-universite.org

Abstract

Electric vehicles (EVs) have become a focal point in public debates on climate change and transportation, particularly in Europe following the European Parliament's decision to ban the sale of new combustion engine vehicles by 2035. This study analyzes French conversations on EVs using a novel open-source dataset of Twitter discussion trees. The dataset includes over 1 million tweets, structured as 22,635 discussion trees, and offers insights into the dynamics of online discourse about EVs. We manually annotated a subset of tweets and fine-tuned a BERTweetFr model for automatic classification and sentiment analysis. Our findings reveal specific structural patterns in online discussions, varied behavior and influence of key actors, and thematic contexts in which EV-related topics are addressed. This research contributes to understanding public opinion on EVs and provides a methodological framework for analyzing online conversations.

Datasets — https://doi.org/10.5281/zenodo.19100383

Introduction

Online Social Networks (OSNs) represent a significant and influential part of public space. Scientists, politicians, and citizens express themselves and interact on a wide range of topics, from everyday concerns to major international issues. Statements and conversations online are such an important part of public debate that journalists and researchers turn to social media to better understand media events and opinion dynamics (Hernández-Fuentes and Monnier 2022).

Social media are often perceived as spaces of polarization and conflict (Conover et al. 2011; Kubin and Von Sikorski 2021). Nevertheless, they also serve as spaces where communities can connect and bond (Jouët 2018) creating welcoming and lighthearted environments in the same way as a third place (McArthur and White 2016). Ultimately, OSNs are spaces where informal discussions occur, where debate (confrontational or not) unfolds, and where controversies are built.

Since the European Parliament (EP) voted to ban the sale of new combustion engine vehicles by 2035, electric vehicles (EVs) have become in Europe a symbol of public authority actions to tackle climate change. Transport accounts for 23% of CO2 emissions in the world, making the decarbonization of this sector a critical issue (IEA 2020). However, EVs remain a controversial subject, generating considerable debate and lacking consensus. They raise many questions in the public debate: Are infrastructures ready? What type of energy powers the vehicle? Under which conditions are EVs environmentally efficient? Are the ores resources sufficient to supply the demand? Can people afford EVs? Do people actually like EVs?

Over the past few years, EVs have attracted substantial attention across various disciplines (Todorovic, Aldakkhelallah, and Simic 2023). In addition to the extensive research on the technical development and optimization of EVs (Sanguesa et al. 2021; Arif et al. 2021), public perception has emerged as a key research area. To guide public policy, enhance adoption, and adapt infrastructure effectively, understanding public opinion is crucial. Social sciences approach this through traditional surveys and interviews to understand consumer practices (Henriksen et al. 2021), assess the acceptability of EVs as an eco-innovation (Coulbaut-Lazzarini and Danteur 2017), and describe the uneven growth of the EV market (Gomes, Pauls, and ten Brink 2023). Recently, researchers have started to investigate the public perception of EVs on OSNs, especially through sentiment analysis approaches (Wang et al. 2022b). From the work of Ruan and Lv, we know that EVs are discussed exponentially on OSNs, including Twitter (Ruan and Lv 2023). The authors showed the diversity of topics addressed in discussions about EVs, such as charging stations, public policies, or gas prices. Interestingly, politicians tend to have more positive views on EVs than the general public. However, in their study, they only focus on English-language tweets from 2011 to 2020, leaving unanswered potential cultural differences over EVs perception on social media.

Given that EVs have been actively promoted by politicians as instruments of sustainable development, we hypothesize that environment and ecology themes will be closely linked to EV discussions and that politicians are more likely to express positive sentiment toward EVs on Twitter/X. However, recognizing that opposition voices have also mobilized strong resistance, we investigate which political side has the most influence in shaping EV discourse on the platform.

In our work, we propose to bridge the gap through the analysis of French conversations about EVs in 2023. To capture a large range of EV discussions and to better understand the context of conversations, we base our work on a novel open-source dataset of Twitter discussion trees and make it publicly accessible. We present the exploration of these discussions through structural and content analysis. With this work, we intend to (1) describe the structural forms of discussions, (2) identify the types of actors involved and their impact on the discussions, and (3) understand the conversational context in which the EV discourse occurs.

This study focuses on French tweets as part of a broader project aimed at understanding online conversations in French. Our dataset is the first public corpus on electric vehicle discussions in French. It enables research on public opinion and contributes to language-specific research works that go beyond the English language.

Related Work

Public perception of EVs

There have been a growing number of studies focused on the public opinion of EVs over the past decade. Understanding public perspectives on EVs is important to implement appropriate public policies, marketing strategies, and more generally observe the evolution of public opinion dynamics. Most of this work is based on traditional social science methods, such as surveys and interviews (Singh, Singh, and Vaibhav 2020; Broadbent, Wiedmann, and Metternicht 2021). Firstly, they show that demographic factors such as age, gender, and income influence EV acceptance (Curtale, Liao, and van der Waerden 2021). Environmental concerns can also positively impact EV adoption (Biresselioglu, Kaplan, and Yilmaz 2018). Krishna highlights limitations for adoptions such as concerns about safety due to lack of noise, the risk of explosion, the non-affordability, and insufficient infrastructure (Krishna 2021). This body of work shows the strong interest from the research community in user adoption (Maybury, Corcoran, and Cipcigan 2022; Wang et al. 2022a; Lashari, Ko, and Jang 2021) and consumer perspectives (Henriksen et al. 2021; Buhmann and Criado 2023). Other researchers investigated public opinion about EVs on social media, using large datasets and computational methods. (Balla et al. 2023; Debnath et al. 2021; Ruan and Lv 2023). The analysis of platforms like Reddit and Twitter has provided insights into public perception at a larger scale, revealing a wide spectrum of topics covered when discussing EVs (Ruan and Lv 2022, 2023; Priyam, Ruan, and Lv 2023). They studied the evolution of discussions around EVs between 2011 and 2020, showing a greater interest among Twitter and Reddit users. Moreover, they showed that those users have predominantly positive views about EVs.

Analysis of online conversations

Conversations on social media are often seen as disjointed messages or threads, restricting the understanding of the global context. Reconstructing these conversations involves identifying and linking related posts, replies, and comments to form conversations as graph structures. This has been explored in the beginning of social network analysis research. For example, Cogan et al. (2012) showed that Twitter conversations can be seen and analyzed as reply graphs, which exhibit a tree-like structure. These trees are usually small and short in time. It was discovered that 85% of tweets with replies only received one reply, and a majority of the replies occur within the hour following the post (Belkaroui, Faiz, and Elkhlifi 2014). More recently, Brambilla et al. analyzed conversations about the COVID vaccine on Reddit (Brambilla et al. 2022). They observed that when discussions start with a negative sentiment (such as "covid side effects"), they tend to contain more side-topics in the following messages of the conversation.

Public debate analysis on Twitter

Twitter has been well-studied as an arena of public debate (Boyadjian 2014). The platform has already been analyzed as a tool for protest movements during the Arab Spring (Tufekci 2017), as a space for sharing experiences during COVID-19 (Zhang et al. 2021), or as a witness of public space reconfiguration during elections (Gaumont, Panahi, and Chavalarias 2018). Various methods are employed to analyze online debate. Some studies use networks to represent interactions, identifying different forces and how they evolve over time (Gaumont, Panahi, and Chavalarias 2018; Gaisbauer et al. 2021). Another approach to public debate is content analysis. This often involves sentiment analysis and topic modeling, which has been used in a variety of contexts such as to analyze the "Nazi in Ukraine" disinformation narrative on Twitter and Telegram (Kloo, Cruickshank, and Carley 2024) or as a tool to characterize the issues discussed by members of the US Congress and the American public on Twitter (Barberá et al. 2019). Additionally, Uthirapathy and Sandanam (2023) used LDA to identify commonly discussed topics in a tweet dataset about climate change, as well as a BERT-based sentiment model to better understand user's position towards climate change.

Methods

Overview of the study

In this study, we started by collecting tweets from discussion trees. Then, a subset of the tweets was manually labeled for conversation subject (EV or not) and sentiment. Models were then trained to annotate the entire dataset. From there, additional analysis (identification of the main users, topic modeling, etc.) was performed to better understand the trends in the conversation about EVs on Twitter (Figure 1).

Dataset construction

In April 2022, Elon Musk bought Twitter and progressively changed the name to X. However, as the URL address of the platform was *twitter.com* at the time of data collection and for clarity purposes, we will refer to the platform as Twitter and to the posts as tweets. Over the last few years, it has become harder to gather data from Twitter. Due to internal pressure to generate profit and to policy change, the Twitter

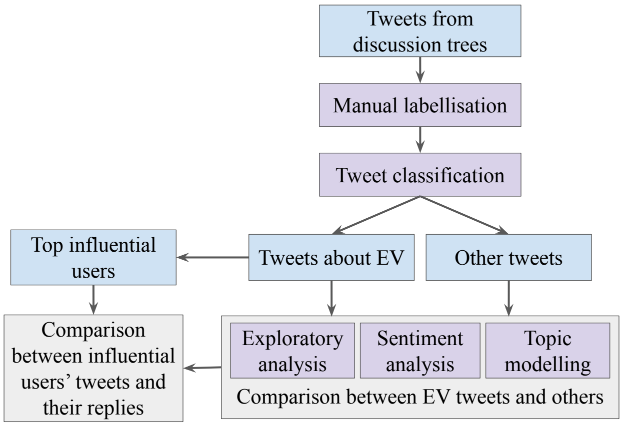

Figure 1: Schema of the study.

API for Academic Research is no longer available. Moreover, since 2023, Twitter has implemented a daily rate limit on the number of tweets an account can access.

To pursue our analysis of conversations on Twitter, tweet collection was performed using the Python open source scrapping library *snscrape*[1], adapted to the new Twitter constraints. Our scrapping started after Musk's policy change, which implied long waiting times. In order to speed up the collection process, we used 2 collection streams to gather tweets by keywords and then 4 collection streams to collect discussion trees. Each stream being associated with one Twitter account.

Collection of tweets by keywords The first part of the dataset was collected from a keyword search. We used two queries based on French keywords representing EVs : "vehicule électrique" (electic vehicle) and "voiture électrique" (electric car), not sensitive to case nor accents. For each public tweet, we collected the following information: tweet ID, tweet date, author username and metadata, tweet content (including text, hashtags, mentions, and URLs), number of retweets, number of likes, number of replies, and conservation ID. If the tweet is a reply to another tweet, we also obtain the ID of the parent tweet. In total, the dataset collected by keywords contains 125,000 tweets corresponding to tweets posted between May 2021 and May 2023 by 65,432 users.

Collection of discussion trees For each keyword tweet already collected, we also gathered its discussion tree. Discussion trees are series of reply cascades initiated by one tweet (root tweet) in a tree-like structure (Fig. 2). They are identified by a unique conservation ID which corresponds to the ID of the root tweet. For each unique conversation ID of our keyword dataset, we collected all its reply branches. The collection includes only publicly available tweets. The collection started in July 2023 and ended in July 2024. The tree dataset contains 1,040,262 tweets corresponding to 22,635 trees, from 241,540 accounts (Table 1). This dataset is publicly available on Zenodo[2]. In this version, we applied anonymization by removing user identifiers such as usernames and account IDs, as well as mentions within the tweet text.

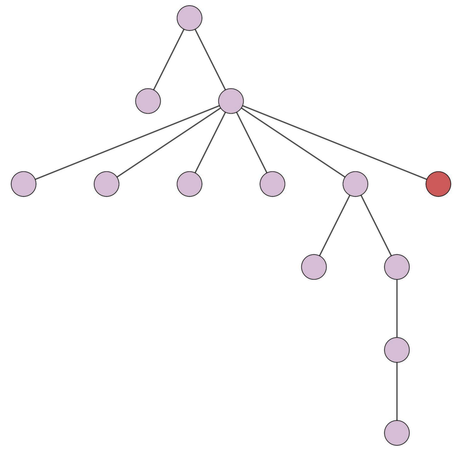

Figure 2: Example of a discussion tree. Nodes correspond to tweets and two nodes are linked if one is a reply to the one above. The red node is the seed tweet containing a collection keyword which led to the collection of the discussion tree.

Dataset	# Tweets	# Users	Unique conv.
Keyword tweets	125,000	65,432	98,178
Discussion trees	1,040,262	241,540	22,635

Table 1: Statistics of the datasets.

Detection of EV tweets

In tree branches, several conversation topics can occur. In order to identify tree segments dedicated to EV topics, we propose a method based on tweet annotation and classification. First, we manually annotate a subset of tweets, then we automatically classify all the other tweets using a fine-tuned BERTweetFr model. The manual annotation part allows one to train and evaluate our classification models.

Manual annotation The annotation was performed by two Ph.D. students. One annotator has a background in information and communication studies, and the other one in computer science. Both annotators labeled the same tweets. The tweets were annotated based on three mutually exclusive labels: "About EV", "Related to EV" or "Other". The label "About EV" includes all tweets talking about EVs (including hybrid and hydrogen vehicles), their use, and their impact, such as car batteries, recharge stations, price or autonomy. The label "Related to EV" corresponds to tweets covering topics close to EVs stakes (Ruan and Lv 2022): combustion-engine cars, mobility, environment, and energy. All the other tweets not covered by the two previous labels are assigned to "Other". The annotation was performed on

[1]https://github.com/JustAnotherArchivist/snscrape

[2]https://doi.org/10.5281/zenodo.19100383

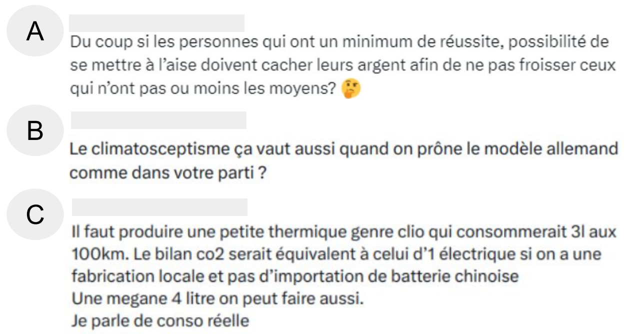

Figure 3: Examples of labeled tweets. The tweets translate to (A) "So, if people who have a minimum of success and are able to make themselves comfortable have to hide their money to avoid offending those who have less or no means?" (Other), (B) "A small combustion engine car Clio-like that consumes 3l per 100km should be produced. The co2 balance would be equivalent to an electric car if it were made locally and no Chinese batteries were imported A Megane 4l we can also do it. I'm talking about real consumption" (About EV), (C) "Does climate scepticism also apply when you advocate the German model, as your party does?" (Related to EV).

the basis of precise written guidelines. Examples of tweets with their label are shown in Figure 3. As reply tweets often refer to previous contexts, we annotated the tweets in their chronological order, per branch. In order to select a representative subset of tweets, we selected random branches of our corpus. First, we selected one random tweet among all the tweets and then we went up from recipes to replies to the root of the discussion tree. The tweets from the root to the randomly picked tweets were manually annotated in chronological order. After a first batch of annotations, it became clear that the different labels are unequally represented in the dataset. To balance our annotated dataset, we finally picked some additional branches, where we knew EVs were mentioned (based on keywords). Even if our seed tweets are in French, a small portion of trees contain tweets in other languages. Only French and English tweets were annotated, being restrained by the languages spoken by the authors. The manually labeled dataset contains 1600 tweets. Cohen's kappa (Cohen 1960) between the two annotators is 0.87, so we consider that the agreement between the two annotators is strong (McHugh 2012). Where the annotations diverged between annotators, they ran through the tweets to discuss and find a consensus.

Automatic annotation To generalize the annotation of our tweets beyond the manual annotation, we opted for a supervised classification approach. All experiments were performed on a standard personal laptop equipped with a Nvidia GeForce RTX 4050 Max-Q/Mobile GPU. We trained several classifier models: SGD, BERT, BERTweetFr. In order to make the classification task simpler, we fused the labels "Related to EV" and "Other" to create a large "Other" category. With the SGD classifier, we pre-processed tweets' content by performing a TF-IDF vectorization. BERTweetFr is a language model based on CamemBERT (a generic French model following RoBERTa architecture) and re-trained on French tweets (Guo et al. 2021). For transformer models, tweets are tokenized. Models were trained or fine-tuned on the manually annotated tweets, we used a 60-20-20 split for training, evaluating, and testing our models. Performance results are detailed in Table 2. Models were evaluated by training on 4 different data partitions (i.e., using different seeds for the 60-20-20 split), which were replicated 5 times to obtain the class-weighted average performances. The results presented (Table 2) are the best models for fine-tuning the hyperparameters (epoch, maximum sequence length, and batch size). BERTweetFr obtained the best performance. This aligns with prior evidence that domain-adaptive pretraining on in-domain data, followed by task-specific fine-tuning, typically outperforms generic language models (Barbieri et al. 2020; Münker, Kugler, and Rettinger 2024). Once the hyperparameters were chosen (here 3 epochs, 8 batch size, 128 maximum sequence length, $5e^{-5}$ learning rate), a new model was trained with these hyperparameters without the test set. This model was used to classify all the other tweets of the dataset.

Model	Precision	Recall	F1-score
SGD	$0.910{*}\pm0.012$	$0.910{*}\pm0.011$	$0.908{*}\pm0.012$
BERT	$0.907{*}\pm0.014$	$0.907{*}\pm0.014$	$0.905{*}\pm0.015$
FlauBERT	$0.778{*}\pm0.220$	$0.845{*}\pm0.118$	$0.805{*}\pm0.179$
BERTweetFr	$0.933{*}\pm0.008$	$0.934{*}\pm0.008$	$0.934{*}\pm0.008$

Table 2: Classification performances. Precision, recall and F1-score are weighted. * indicates the mean performances after 5 independent trainings on 4 different data splits, followed by the standard deviation.

Sentiment analysis

Detection of tweets' sentiment was performed in a similar fashion to the detection of EV tweets. The same tweet subset was labeled for sentiment analysis with positive, negative, neutral, and mixed classes. Using these labeled tweets as an evaluation, we explored two sentiment analysis approaches: lexicon-based models and a task-specific BERT model. For the lexicon-based approach, we used VADER and AFINN as they are known to work well on social media data. For both of them, we used their French version. We compare in Table 3 their results with a fine-tuned BERTweetFr model (6 epochs, 8 batch size, 128 maximum sequence length, $5e^{-5}$ learning rate) using our manually labeled tweets (preprocessed with tokenization).

Due to a small number of tweets labeled and to reduce complexity in training, we trained and tested (for all models) using only positive and negative labeled tweets. We manually checked a substantial portion of the BERTweetFr results to confirm the coherence of the labels on previously unseen data. In the following of the work on sentiment analysis, BERTweetFr was chosen for its better performance.

Model	Precision	Recall	F1-score
VADER-fr	0.761	0.546	0.589
AFINN-fr	0.779	0.447	0.476
BERTweetFr	0.895∗±0.027	0.898∗±0.025	0.896∗±0.026

Table 3: Sentiment analysis performances. Precision, recall and F1-score are weighted. * indicates the mean performances after 5 independent trainings on 4 different data splits, followed by the standard deviation.

Results

In this section, we provide an explanatory analysis of the discussion trees and their tweets, a dive into the different topics related to the EV conversation, and finally, we explore the sentiment of tweets over time and in relation to the status of the author.

Data description and exploratory analysis

A significant portion of trees are trivial trees: 55% of trees (12,515 trees) contain only one tweet (i.e., one node). The mean number of tweets per tree is 46. When taking into account only trees with a root and at least one reply, the mean number of tweets per tree is 99.3. The mean number of replies (i.e., mean out-degree) is 0.9 replies per tweet. If we take into account only tweets with at least one reply, the mean out-degree is 2.9 and with at least two replies, the mean out-degree is 8.9. Except for a small increase at out-degree $k = 200$, we see that the degree distribution of the tree nodes follows a power law distribution (Figure 4). Within trees with at least two nodes (i.e., one reply), 80% of them had all their replies within 7 days after the root tweet was posted.

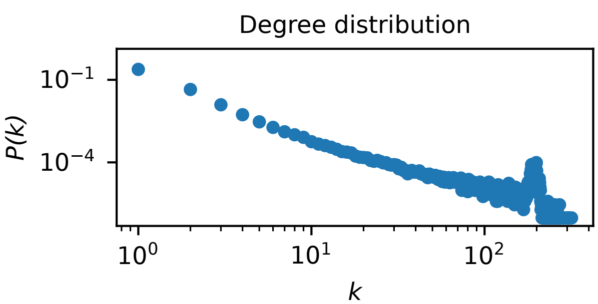

Figure 4: Out-degree (k) distribution of the tweets. The tweet's out-degree corresponds to its number of replies.

After tweet classification (see *Automatic annotation*), we obtained 81,880 tweets about EV and 935,402 tweets related to other subjects. EV tweets that represent only 8% of total conversations show that conversations within our dataset cover a wider range of subjects. Indeed, discussion trees exclusively related to EV are a minority, most trees contain several discussion topics (especially large ones). These conversations can allow us to better understand the context in which EV topics appear.

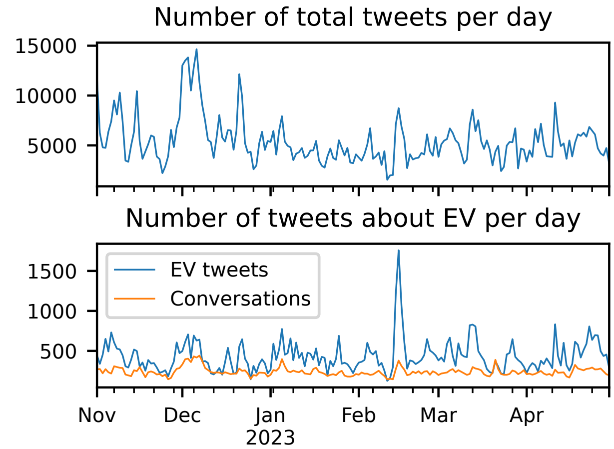

Figure 5: Evolution per day of the number of tweets. Above is the timeline of all tweets and below is the number of tweets about EV (blue) with the number of different conversations active on that day (orange).

Timeline The tweet timeline shows the evolution of the number of tweets posted per day in our dataset (Figure 5). On the timeline of EV tweets, we can see a peak on February 14^{th}. This date corresponds to the day of the adoption by the European Parliament (EP) of a ban on new combustion engine vehicles by 2025, which was vastly discussed by users. This increase in activity is partly due to an increase in the number of discussion trees, but mainly due to larger conversations (more tweets per tree).

Top words To gain a better understanding of the discourse differences held in the EV tweets and the other tweets, we extracted the most specific terms of each tweet set compared to the other. After preprocessing and lemmatization, we computed the log odds ratio of the terms (Monroe, Colaresi, and Quinn 2008). In other words, the ratio of the odds of a term in EV tweets vs. non-EV tweets is computed; this was performed for each term of the dataset. To reduce the importance of small occurrence terms that appear in a single set, we multiplyed the term score by its frequency for each set to select the top 25 (see Table 4 and the Appendix for English translation).

Keywords associated with EV tweets focus mainly on the use and properties of EVs. The largest category of words in Table 4 mentions different car brands (“peugeot”, “bmw”, “volkswagen”, “audi”, “hyundai”, “nissan”, “byd”, “mustang”, “opel”) which is completed by mentions of specific car models “zoé” and “megane” of the French car brand Renault which are both available in electric version. Terms like “recharge”, “recharger”, “rechargeable”, “rechargement”, feature the importance of charging functions in the EV discussions. This discussion is broadened to the type of energy used (“hydrogène”) and the technical characteristics (“batterie”, “lithium”). A special case is highlighted with the “hybride” cars, combining electric and combustion engines which can be rechargeable (”phev”) or nonrechargeable. The last category of keywords is different from the other, "geeko" and "frandroid" are both online media specialized in tech news. They often report information related to new EVs or technological breakthroughs. This selection of keywords shows discussions showcasing and comparing electric car models with their different properties.

On the other hand, keywords describing the other tweets show different conversation patterns (Table 4). The terms refer to different topics in the public sphere, with an emphasis on politics. Several political parties are on the forefront: "rn" (Rassemblement National) is a French far-right party with Marine Le Pen ("pen") as its leader, and lfi (La France Insoumise), a French left-wing party with Louis Boyard ("boyard") being one of its members of parliament. The dates of the tweets overlap with a pension reform in France associated with important protests, some keywords are references of this political event and its repression ("gréviste", "policier", "soignant"). The keywords "hanouna" and "boyard" refer to a mediatic event of that period where both men had a fight on a TV set which had an important mediatic and political impact. The keywords of the non-EV tweets represent the diversity and conflictual conversation happening in public space at the time.

EV terms	Freq.	Non-EV terms	Freq.
électrique	19598	raciste	894
batterie	4503	rn	1212
recharger	1740	soignant	891
recharge	1116	lfi	977
hydrogène	1127	hanouna	538
lithium	1121	gréviste	433
hybride	785	soigner	427
frandroid	217	voix	463
berline	216	mandat	343
peugeot	211	racisme	425
zoé	214	musulman	413
rechargeable	130	présidentiel	401
bmw	207	fasciste	394
volkswagen	153	langue	352
megane	137	(le) pen	385
audi	147	policier	300
rechargement	86	communisme	238
hyundai	102	islam	210
nissan	84	fascisme	249
byd	49	darmanin	202
phev	50	soldat	188
geeko	32	plainte	188
mustang	33	cnews	180
opel	29	boyard	178
taycan	25	terrorisme	169

Table 4: Most specific keywords of EV tweets and non-EV tweets with their frequencies.

Qualitative analysis of long EV segments From the discussion trees, we identified EV segments (i.e., consecutive EV tweets). We read the 47 segments containing at least 10 consecutive EV tweets to better understand how EV conversations develop. These segments represent 561 tweets; as segments can have partial overlaps, they correspond to 391 unique tweets. In these segments, users frequently discuss detailed comparisons of performance, advantages, and disadvantages across different models. Often, they compare EVs, hybrid vehicles, and combustion engine vehicles. These comparisons are both theoretical (manufacturer-claimed performance or the average performance of a vehicle category) and grounded in everyday life, often referencing specific personal needs (e.g., the necessity of three child seats or the feasibility of long-distance travel). A strong focus is placed on usage scenarios and practical considerations. Environmental and economic aspects are frequently addressed through a technical lens, with questions centered on measurable factors such as model type, battery size, and energy costs. Conversations also reflect an interest in the future of technological developments, such as battery improvements or the impact of hydrogen in the automotive market, with participants debating which choices might prove most beneficial and cost-effective in a few years. The segments are often conversational, involving several users and characterized by opposing viewpoints and back-and-forth arguments ("I think X is better" followed by "Yes, but X has this problem"). Six segments are exceptions; they have a thread structure where only one user develops arguments across a series of tweets.

Topic modeling

To deepen our insight into conversations, we used a topic modeling approach, which allows identifying diverse underlying subjects in the tree dataset. We utilized the Latent Dirichlet Allocation (LDA) model on all the tweets from the tree discussions. As tweets are short and heterogeneous in terms of both vocabulary and syntax, all tweets from one tree were concatenated to form one document. Hyperparameters were optimized using the CV coherence score. It ended up with 19 topics, the alpha value as 'symmetric' and eta at 0.1.

Based on the terms representing each topic, provided in Table 5, certain clusters stand out. Firstly, there is a large cluster containing seven topics related to transportation (topics 5, 7, 10, 11, 14, 17, 19). Topic 17 is specific to planes (*jet*, *avion*), but all the other ones are related to cars. This demonstrates the diverse range of car-related discussions, encompassing topics such as engine types (*moteur*, *thermique*), fuel sources (*essence*), vehicle categories (*SUV*), and specifically EVs (*Tesla*, *batterie*, *lithium*). The second large topic cluster is about French national politics and social movements (topics 1, 2, 8, 9, 15). They refer to Macron's pension reform (*reforme*, *retraite*, *travailler*) and people's reactions such as strikes (*grève*). The topics also refer to energy policy including the risk of temporary shutdown (*électricité*, *coupure*, *nucléaire*) in winter 2023. Topic 18 is a special case of political discussions as it is related to local politics and policies in Paris, tackling the mayor (*Anne Hidalgo*, *maire*), cleanness (*rat*, *saleté*), outdoor spaces (*rue*, *arbre*, *place*) and the Olympic games (*JO*). There are mentions of energy aspects within the topics 12 and 4. Topic 12 is centered around energy questions in Quebec. This suggests that there might be some regional stakes which can be found back in the discussions. Topic 4, on the other hand, is tackling the subject of energy in Europe and especially in France and Germany. Topics 3 and 6 relate to science and climate, covering matter from COVID-19 and its vaccines (*covid*, *vaccin*, *scientifique*) to the many challenges of climate crisis (*réchauffement*, *climatique*, *co2*, *planète*, *problème*).

To examine the specific topics related to EVs in greater detail, we apply topic modeling exclusively to the EV-related tweets (Table 6). From the 10 topics extracted from the EV tweets, we identify 4 categories: characteristics and comparison of vehicles, energy production, charging infrastructure, and negative impact of EVs.

- **Characteristics and comparison of vehicles** (topics 1, 2, 7, 8, 9): Topics 1, 2, and 8 highlight the technical and performance characteristics of EVs (*batterie*, *recharge*, *lithium*, *moteur*) and conversations involve discussions and comparisons related to power efficiency, vehicle size, performance, and comfort (*comparer*, *autonomie*, *kWh*, *trajet*, *prix*, *temps*, *poids*, *confort*). Comparison can be with other EV models and brands (*tesla*, *megane*, *hybride*), with combustion engine vehicles (*thermique*, *diesel*, *essence*), more broadly generic car characteristics (*suv*, *gamme*, *break*, *citadine*, *berline*), brands (*Peugeot*, *BMW*, *Renault*), and even alternative vehicles such as bikes and electric-assisted bikes in topic 9 (*vélo* and *VAE*). In topic 7, affordability is emphasized (*prix*, *acheter*, *coûter*, *moyens*), as well as charging infrastructures (*borne*, *électricité*, *batterie*, *recharge*, *station*, *place*). Topic 8 is more focused on market activities (*dévoiler*, *marché*, *constructeur*, *sav*) and shows the central place of *Tesla* (*Musk*) in this market.
- **Energy production** (topics 3, 6): These topics explore the different types of energy sources, either fossil energy (*gaz*, *combustion*, *charbon*) or alternatives (*hydrogène*, *vert*, *nucléaire*) and their *production* (*produire*, *rendement*, *centrale*) including the consequences (*désastre*, *carbon*, *co2*) and advantages (*solution*, *écologique*). These discussions are held in the context of discussion on transportation (*véhicule*, *kilomètre*, *vélo*, *transport*, *automobile*, *avion*). Topic 6 explores the international stakes of energy production and consumption market (*chinois*, *européen*, *france*, *industrie*).
- **Charging infrastructure** (topic 4): This topic focuses on the infrastructure required for EVs, particularly charging stations (*borne*, *recharger*, *ionity*, *kw*) and their practical use during long-distance travel (*autoroute*), reflects user discussions about charging times (*temps*), range limitations (*autonomie*, *kilomètre*), and price (*prix*).
- **Negative impact of EVs** (topics 5, 9): The increasing need for mineral ores (*batterie*, *cobalt*, *lithium*) for batteries, often discussed as a humanitarian risk in the tweets, is covered in topic 5. It also highlights concerns about battery safety and end-of-life management (*fin de vie*), including the fear of fire (*problème*, *incendie*, *feu*). Safety concerns (*accident*, *frein*, *assistance*, *permis*) are also part of the discussions in topic 9.

From the analysis of the EV topics, we see a wide range of subjects covered in the discussions. Many are related to long-term public policies such as infrastructures, energy production, and resources management. However, the keyword extraction suggests that the subjects specific only to EV tweets are mainly about car recharge and model specificities. We can deduce that public policies-related subjects are discussed both in "EV" and "Other" tweets.

Sentiment analysis

Tweets sentiment Tweets sentiment over time is shown in Figure 6 as the proportion of negative tweets. The overall sentiment is displayed in blue, and EV-only tweets sentiment in orange. The sentiment is computed per day then, for readability, the curve is smoothed by a 3-day window.

The overall sentiment is consistently between 77% and 85% of negative tweets. EV tweets sentiment is systematically more positive, with higher variations. Of all data, 73% of EV tweets are negative and 82% of non-EV tweets are negative. Minimum and maximum sentiment overtime are determined by two notable peaks: 61% on January 1st 2023 and 82% on March 26th 2023, respectively. The minimum peak happens to be a day with fewer tweets about EVs and the day after New Year's Eve, when users might be less prone to engage with controversial and heated content.

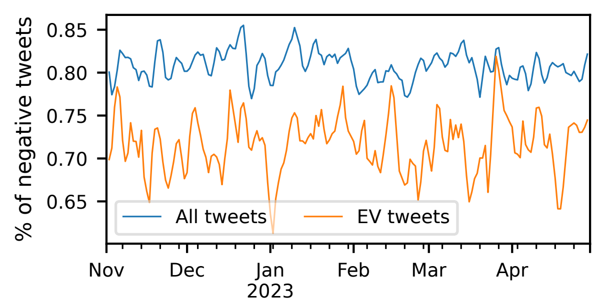

Figure 6: Evolution of the proportion of negative tweets per day. The blue line corresponds to the sentiment of all tweets and the orange one to only EV tweets. The curve is smoothed by a 3-day window.

Another important peak is at mid-February, corresponding to the day when the EP decided to ban selling new combustion-engine cars by 2035. This peak indicates a more negative trend at that date and corresponds to people criticizing members of parliament and the result of the vote. The March peak, corresponding to the period with the highest percentage of negative EV tweets, can be explained by two different events. The first reason is the publication by different accounts, on March 24th and in the following days, of a video described as Congolese miners escaping from a collapsing mine. This video led to conversations about the human cost of the ores needed to produce batteries for EVs. The second reason for the maximal peak is the agreement on March 25th by the European Commission of the regulation on CO2 emissions for new cars. This agreement followed the vote of the EP and led to the adoption by the Council of the European Union on March 28th. After March 25th, many

Topic 1	gauche, macron, france, français, pays, extrême, peuple, rn, guerre, voter, politique, dire, croire, vrai, jamais
Topic 2	an, retraite, france, français, réforme, payer, riche, macron, vie, argent, jamais, dire, travailler, petit, pays, travail, pauvre
Topic 3	scientifique, croire, dire, an, climatique, lire, comprendre, question, vaccin, science, réchauffement, covid, jamais, vrai
Topic 4	nucléaire, énergie, allemagne, charbon, france, allemand, électricité, gaz, centrale, production, produire, europe, pays
Topic 5	vélo, véhicule, kilomètre, route, ville, thermique, moteur, rouler, autoroute, permettre, paris, transport, vitesse, place
Topic 6	écologie, climatique, écolo, co, nucléaire, planète, climat, écologique, politique, problème, dire, eau, réchauffement
Topic 7	prix, payer, taxe, euro, essence, acheter, mois, carburant, borne, coûter, argent, augmenter, français, électricité, an, france
Topic 8	payer, train, travail, grève, travailler, droit, service, public, transport, an, jour, salaire, sncf, paris, temps, retraite, vie
Topic 9	france, macron, français, politique, ministre, pays, gouvernement, arrêter, temps, nucléaire, écologie, an, croire, écolo, fin
Topic 10	tesla, suv, kilomètre, acheter, véhicule, gros, petit, ville, an, vrai, dire, grand, vraiment, enfant, autonomie, problème, vie
Topic 11	co, thermique, batterie, kilomètre, recharge, émission, véhicule, énergie, borne, production, solution, gt, problème
Topic 12	québec, hydro, électricité, montréal, temps, réseau, devoir, jour, année, gaz, auto, gouvernement, pis, char, an, caq
Topic 13	pays, noir, france, afrique, africain, blanc, histoire, raciste, français, venir, dire, religion, algérien, algérie, musulman, croire
Topic 14	batterie, tesla, jour, réseau, recharger, nouveau, courant, temps, lithium, heure, recharge, an, charge, acheter, compteur
Topic 15	macron, france, gouvernement, français, ministre, électricité, coupure, payer, vraiment, con, jour, dire, petit, grand, darmanin
Topic 16	eau, manger, animal, viande, agriculteur, produit, élevage, petit, nourrir, problème, venir, insecte, agriculture, utiliser
Topic 17	eau, jet, co, carbone, priver, avion, france, macron, taxe, pays, train, jour, français, arrêter, petit, payer, pluie
Topic 18	paris, parisien, hidalgo, rat, mairie, ville, français, arbre, maire, rue, jo, anne, france, moron, réticent, place, saleté
Topic 19	abordable, spring, dacia, ligier, séduire, chargemap, simon, permettre, nikola, nft, mythe, petit, ivoirien, commercialisation

Table 5: Description of all tweet topics by their most frequent words.

Topic 1	batterie, thermique, kilomètre, heure, véhicule, recharge, autonomie, pourcent, kwh, charge, borne, prix, recharger, temps
Topic 2	suv, tesla, gros, véhicule, berline, clio, kilomètre, petit, diesel, hybride, citadin, megane, prix, constructeur, comparer, peugeot
Topic 3	hydrogène, moteur, eau, pourcent, énergie, produire, gaz, batterie, vert, électricité, rendement, production, véhicule, thermique
Topic 4	kilomètre, essence, recharge, thermique, borne, recharger, autonomie, autoroute, acheter, batterie, tesla, rouler, prix, temps
Topic 5	batterie, tesla, cobalt, lithium, changer, savoir, recycler, thermique, problème, penser, incendie, insecte, fin, vie, pis, électricité
Topic 6	électricité, acheter, nucléaire, chinois, français, véhicule, avion, passer, industrie, centrale, production, thermique, européen
Topic 7	véhicule, thermique, borne, prix, électricité, essence, hybride, pourcent, diesel, acheter, batterie, mois, recharge, recharger
Topic 8	tesla, prix, véhicule, autonomie, renault, kilomètre, rouler, dévoiler, marché, acheter, batterie, essence, pourcent, monde
Topic 9	vélo, euro, permis, gt, lt, vae, rouler, véhicule, frandroid, thermique, frein, payer, batterie, accident, volkswagen, assistance
Topic 10	énergie, thermique, véhicule, monde, borne, tesla, lithium, payer, batterie, recharger, chine, apparaître, essence, recharge, pays

Table 6: Description of EV tweet topics by their most frequent words.

tweets criticized this decision which opens the way for a replacement of combustion-engine by EVs. Conversely, to a lesser degree, some tweets negatively reacted to the fact that Germany negotiated to modify the measures voted by the EP.

Roots are more positive than replies. They are still mainly negative (59%), but significantly less than replies (82%). A random reply is more probable to be negative than positive (Figure 7). We observed the same trend when considering only EV parents tweets or only EV reply tweets. Yet, positive tweets tend to receive more replies than negative ones (Figure 8). However, this difference is reduced for EV tweets and for EV replies.

Influential users Building upon our analysis of overall sentiment, we now examine the posts of key users within our dataset to assess whether their contributions skew toward positive or negative sentiment.

Influential users are identified by the number of replies to their tweets. The top-150 users with the highest number of EV-related replies and the top-150 users with the highest number of replies to an EV tweet are considered influential. The total number of unique influential users is 202. For each of these users, we manually assign a status defined by the type of their organization if the account is a moral entity or by their job if the account represents a physical person. Most of them are from the political or media sphere, such as the official account of Emmanuel Macron, French ministers (Bruno Le Maire, Agnès Pannier-Runacher), opposition (Sandrine Rousseau, Karima Delli), French public media (France 3 Bretagne, France Info), and French-Canadian media (TVA Nouvelles, La Presse). Many accounts are difficult to identify as they are specific to online behavior: mixing news and humor content, self-proclaiming experts in a topic without the possibility to verify the claim, and/or using pseudonyms. We leave these ambiguous accounts on the side for later work, as they need further investigation and a theoretical framework to be analyzed. A smaller portion of the users are from academia (researchers), from industry (CEOs or brands), or are journalists posting under their own name (and not under their employer's).

Finally, we selected politicians, traditional and online media, industry, and journalists to be analyzed. Academia was excluded for its low number of accounts. The number of accounts per selected category is shown in Table 7.

For all categories of influential users, EV tweets are more positive than the other tweets. Politicians and industry accounts post tweets significantly more positive than the average, both about EVs or other topics. The media tweets are slightly more positive for general subjects, but much more positive about EVs. We notice that online media are even

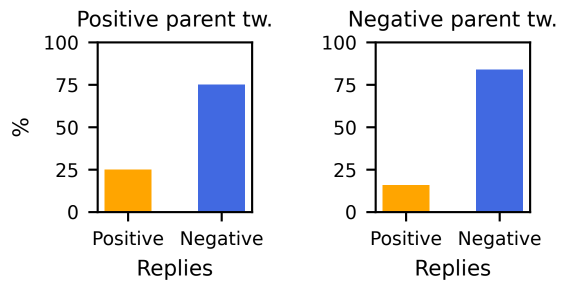

Figure 7: Repartition of reply sentiment based on parent tweet sentiment.

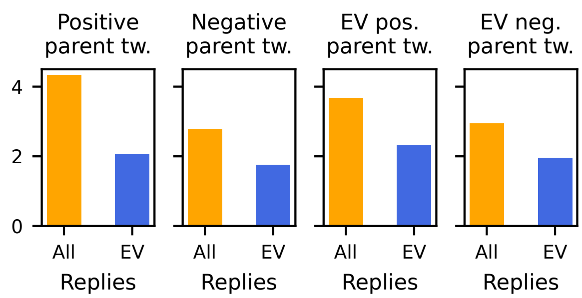

Figure 8: Mean number of total replies and EV replies per tweet (only tweets with at least one reply are included) based on the parent tweet type.

more positive than traditional media. However, for each category (except journalists), the replies are more negative than their parent tweets (posted by the influential users). Politicians, traditional media, and journalists (for EV replies) receive more than 80% of negative replies. Online media and industry accounts receive the most positive replies.

Generally, positive tweets are posted by larger accounts. On average, the number of followers of the accounts which posted negative tweets is 8,679, while accounts which posted positive tweets have on average 32,031 followers.

Conversation context

To better understand the influence of the conversational structure, we constructed a conditional probability tree that shows the impact of tweet ancestors in the class probability of a tweet (Figure 9, full tree in Appendix). In this tree, each node represents a class probability, and edges represent the reply-relationship. Each depth level describes the probability of a tweet's class conditioned on the classes of its ancestor tweets. At depth 1, the tree has two nodes, corresponding to the tweet's probability to be classified as "EV" or "Other" regardless of its ancestors (whether they have ancestors or not). At depth 2, the four nodes contain the class probability of the tweets having at least one parent tweet, depending on the class of their parent tweet. This allows to analyze how the class probability of a tweet differs depending on the conversational depth taken into account. We observe that the more a tweet has EV tweets as ancestors, the more

Category	# Accts	Neg. tweets		Neg. replies	
		EV	Other	EV	Other
All accounts	241,540	0.73	0.82	0.77	0.82
Politicians	30	0.43	0.46	0.87	0.82
Trad. media	24	0.50	0.72	0.87	0.83
Online media	11	0.30	0.62	0.69	0.65
Journalists	7	0.71	0.80	0.85	0.76
Industry	7	0.17	0.20	0.68	0.32

Table 7: Proportion of negative tweets in tweets posted and their replies to EV tweets according to the category of the parent post's author. For comparison, the same proportions are displayed for all the user accounts of the dataset.

it is probable to be about EV. For example, the probability of a tweet to be about EV with four ancestors and only the parent about EV is 0.280 but when the four ancestors are labeled EV the probability is 0.585. Additionally, we note that the place of EV tweets in the ancestors has an impact on the probability of an EV tweet. At depth 4, a tweet with only an EV parent is more probable to be EV than a tweet with only an EV great-grandparent. However, we see that it can be compensated for by a higher number of EV distant ancestors. For example, at depth 5, tweets with their grandparent to their great-great-grandparent about EV (but not their parent) have higher probability to be about EVs than the tweets which only have an EV parent.

Based on the observation that an EV tweet is more probable to appear in a conversation path where ancestors tweets are also labeled EV, we implemented a new tweet classification which takes parent text and parent label as input features. For machine learning models, the training is done in the same conditions as described in the subsection *Automatic annotation*, only adding the parent text (preprocessed with TF-IDF) and the parent label as features. For BERTweetFr, we concatenated the input as a single formatted string: the parent text, label, and tweet text. For both classification approaches, we classified only replies. Performance results (Table 8) show improved classification scores compared to Table 2 for machine learning models. However, BERTweetFr's performance did not improve significantly. This can be explained as classical machine learning models handle categorical features such as class label better than transformers.

Model	Precision	Recall	F1-score
AdaBoost	0.944*±0.007	0.943*±0.006	0.943*±0.006
GradientBoost	0.959*±0.008	0.957*±0.010	0.958*±0.010
BERTweetFr	0.937*±0.007	0.936*±0.007	0.936*±0.007

Table 8: Performance of classification models with parent context. Precision, recall and F1-score are weighted. * indicates the mean performances after 5 independent trainings on 3 different data splits, followed by the standard deviation.

Discussion

Our analysis of discussion trees reveals that EVs are discussed in a variety of contexts, including national politi-

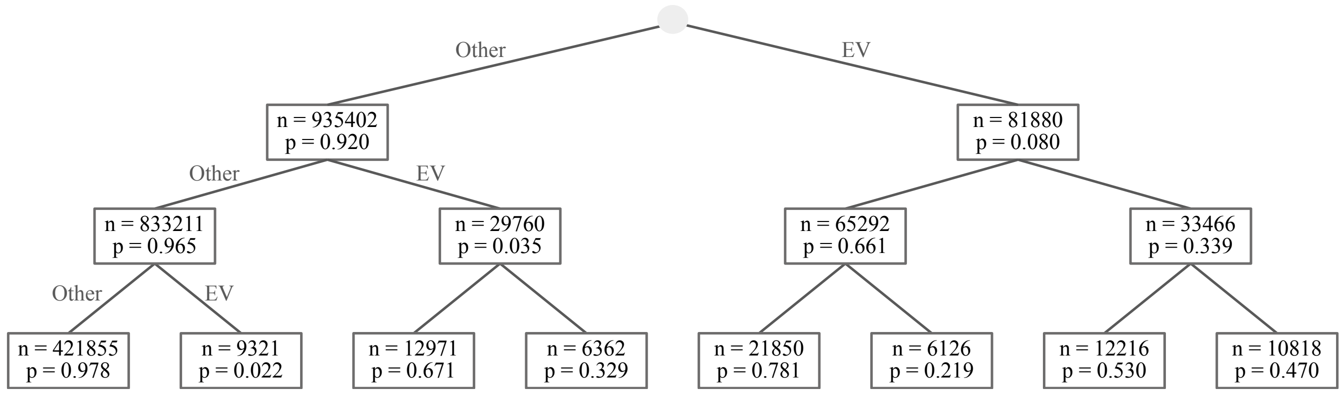

Figure 9: Simplified probability tree of tweet classes ("EV" or "Other") based on their ancestors' classes. Left branches correspond to the class "Other" and right branches to the class "EV". The first level, after the root, includes tweets regardless of their ancestors. Their child includes tweets with at least one parent, and their probability is the conditional probability of the tweet's class depending on its parent class.

cal news, transportation and energy public policies, and resources concerns. When compared to the tweets labeled as "Other", the specific keywords of EV tweets are related to charging and car characteristics. This suggests that political debates are present across all the dataset, but practical and technical aspects are specific to EV tweets. These diverse contexts highlight the multifaceted nature of EV-related discussions on Twitter, ranging from political debates to practical and technical topics. While there are recurring themes related to the environment (energy, ores, pollution), our qualitative analysis suggests that users discussing EVs do not frame particularly these discussions within a context of environmental concerns, but rather consider CO2 consumption as one characteristics of the vehicle among others. Further research is required to determine this.

A notable finding is that tweets directly discussing EVs tend to be less negative than the other parts of the discussion trees. This observation suggests that the portions of the trees where EVs are not explicitly mentioned may serve as a reflection of the broader French Twitter discourse, which often leans towards political, confrontational, or controversial topics. In contrast, EV-related segments encompass not only political aspects, such as environmental consequences and policy decisions, but also more practical and day-to-day implications, which naturally tend to evoke less negativity.

Our analysis also identifies distinct patterns in the sentiment expressed by different types of accounts. Politicians and industry-related accounts (e.g., CEOs and brands) are generally more positive than average, a behavior that could be explained by their supportive attitude towards EVs and the strategic importance of maintaining a positive tone in political and marketing communication. Another group consists of online and traditional media accounts, which exhibit a more positive sentiment towards EVs than other topics. Finally, we observed accounts (journalists) that predominantly post negative tweets, regardless of the topic. Interestingly, journalists exhibit sentiment behavior similar to that of most users in the dataset, reflecting the overall sentiment ratio. This suggests that they may not adopt the same tone or style under their own name as the media organizations they represent.

We observe that replies to tweets from online media and industry accounts (including Elon Musk, who owns Twitter/X) are more positive compared to replies to other accounts. This could indicate higher levels of trust towards these actors. Conversely, the fact that replies to traditional media accounts are less positive may reflect a growing distrust towards established media compared to newer online media.

Our findings also show that politicians maintain a consistently positive tone in their tweets about EVs, comparable to their tone on other topics. This observation expands on the work of Ruan and Lv (2023), who analyzed EV-related tweets in English and found that politicians tend to express more positivity towards EVs than other users. While our results align with this trend, we diverge from their conclusion regarding overall sentiment. Using a lexicon-based method on English tweets, Ruan and Lv reported consistently positive sentiment for EV-related tweets. In contrast, our analysis of French tweets, using a supervised machine learning model, revealed predominantly negative sentiment. This discrepancy raises questions about the influence of methodological choices and potential cultural differences in online discourse across languages.

Positive tweets, especially those from high-profile accounts, tend to attract more replies. However, we observed that replies are, on average, more negative than the original tweets. This finding prompts further investigation: Are replies inherently more critical and negative? Does this depend on the topic or the status of the original author? Is this specific to the platform?

It is important to note several limitations of this study. First, the Twitter data collected may not be exhaustive due to the scraping process, particularly for large tweets. Second, social media users, and especially the ones posting content, do not accurately represent the socio-demographics of the French-speaking community. While Twitter/X users tend to be on average younger, more urban, and more politically engaged than the general French population, social networks remain a critical public space for opinion sharing and discourse (Boyadjian 2014; Mellon and Prosser 2017). In our dataset, we see that French political context dominates over other national context. Third, sentiment analysis has inherent challenges in capturing the nuanced meanings of tweets. For instance, a negative tweet might criticize EV autonomy or target a politician who opposes EVs. Our sentiment annotations contain four classes (positive, negative, neutral, and mixed), which are later reduced to two. This simplifies the modeling task but may miss emotional subtleties like sarcasm or specific affective states. This trade-off was chosen to reduce annotation complexity and improve classification performance, though it inevitably limits the granularity of sentiment insights. Addressing biases and improving the performance of French-language sentiment models will require larger annotated datasets. Future work aims to enrich sentiment analysis with finer-grained emotional categories to better capture the nuances of public opinion. Despite these limitations, our findings underscore the value of studying discussion trees to better understand sentiment dynamics and the broader context of EV-related discourse on Twitter.

Conclusion

This research presents an exploration of public discourse on EVs in French through Twitter conversations with an open source dataset. The analysis reveals that studying discussions instead of only tweets related to EVs provides valuable information on public sentiment and context conducive to EV conversations. The analysis showed that discussion trees contain a wide range of topics beyond EVs, raising questions for future research: How do these conversations evolve over time? Which subjects are most likely to initiate discussions about EVs? And to what other topics do EV-related conversations typically transition? Leveraging the history of discussion trees is an interesting perspective to explore in order to improve tweet classification and to enable a more granular analysis of conversational contexts. Our dataset and findings can support evidence-based decision-making for policymakers, strategic planning for automotive companies, public awareness efforts by NGOs, and further research by social and data scientists into public opinion and sustainability narratives. This work contributes to diversifying language representation in social media research and opens avenues for cross-cultural and regional discourse analysis. Future studies can build on this foundation to compare national or regional narratives around electric vehicles. Furthermore, it would be valuable to investigate how policy changes on Twitter/X and the potential resulting migration of users to alternative platforms such as Mastodon, Bluesky, or Threads might influence EV-related discussions. Comparing the nature and dynamics of these conversations across platforms could offer unique insights into the broader landscape of public discourse on EVs.

Acknowledgments

ChatGPT was occasionally used for translation and rephrasing purposes.

References

Arif, S. M.; Lie, T. T.; Seet, B. C.; Ayyadi, S.; and Jensen, K. 2021. Review of electric vehicle technologies, charging methods, standards and optimization techniques. *Electronics*, 10(16): 1910.

Balla, S. N.; Pani, A.; Sahu, P. K.; and González-Feliu, J. 2023. Examining shifts in public discourse on electric mobility adoption through Twitter data. *Transp. Res. D: Transp. Environ.*, 121: 103843.

Barberá, P.; Casas, A.; Nagler, J.; Egan, P. J.; Bonneau, R.; Jost, J. T.; and Tucker, J. A. 2019. Who leads? Who follows? Measuring issue attention and agenda setting by legislators and the mass public using social media data. *American Political Science Review*, 113(4): 883–901.

Barbieri, F.; Camacho-Collados, J.; Neves, L.; and Espinosa-Anke, L. 2020. Tweeteval: Unified benchmark and comparative evaluation for tweet classification. *arXiv preprint arXiv:2010.12421*.

Belkaroui, R.; Faiz, R.; and Elkhlifi, A. 2014. Conversation analysis on social networking sites. In *2014 tenth international conference on signal-image technology and internet-based systems*, 172–178. IEEE.

Biresselioglu, M. E.; Kaplan, M. D.; and Yilmaz, B. K. 2018. Electric mobility in Europe: A comprehensive review of motivators and barriers in decision making processes. *Transportation Research Part A: Policy and Practice*, 109: 1–13.

Boyadjian, J. 2014. Twitter: A New Opinion Poll? *Participations*, 8(1): 55–74.

Brambilla, M.; Javadian Sabet, A.; Kharmale, K.; and Sulistiawati, A. E. 2022. Graph-based conversation analysis in social media. *Big Data and Cognitive Computing*, 6(4): 113.

Broadbent, G. H.; Wiedmann, T. O.; and Metternicht, G. I. 2021. Electric vehicle uptake: Understanding the print media's role in changing attitudes and perceptions. *World Electric Vehicle Journal*, 12(4): 174.

Buhmann, K. M.; and Criado, J. R. 2023. Consumers' preferences for electric vehicles: The role of status and reputation. *Transp. Res. D: Transp. Environ.*, 114: 103530.

Cogan, P.; Andrews, M.; Bradonjic, M.; Kennedy, W. S.; Sala, A.; and Tucci, G. 2012. Reconstruction and analysis of twitter conversation graphs. In *Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research*, 25–31.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1): 37–46.

Conover, M.; Ratkiewicz, J.; Francisco, M.; Gonçalves, B.; Menczer, F.; and Flammini, A. 2011. Political polarization on twitter. In *Proceedings of the international aaai conference on web and social media*, volume 5, 89–96.

Coulbaut-Lazzarini, A.; and Danteur, T. 2017. Electric mobility analysis: contributions from sociology. *International Journal of Sustainable Development*, 20(1-2): 56–67.

Curtale, R.; Liao, F.; and van der Waerden, P. 2021. User acceptance of electric car-sharing services: The case of the Netherlands. *Transportation research interdisciplinary perspectives*, 149: 266–282.

Debnath, R.; Bardhan, R.; Reiner, D. M.; and Miller, J. 2021. Political, economic, social, technological, legal and environmental dimensions of electric vehicle adoption in the United States: A social-media interaction analysis. *Renew. Sustain. Energy Rev.*, 152: 111707.

Gaisbauer, F.; Pournaki, A.; Banisch, S.; and Olbrich, E. 2021. Ideological differences in engagement in public debate on Twitter. *Plos one*, 16(3): e0249241.

Gaumont, N.; Panahi, M.; and Chavalarias, D. 2018. Reconstruction of the socio-semantic dynamics of political activist Twitter networks—Method and application to the 2017 French presidential election. *PloS one*, 13(9): e0201879.

Gomes, A. D. P.; Pauls, R.; and ten Brink, T. 2023. Industrial policy and the creation of the electric vehicles market in China: demand structure, sectoral complementarities and policy coordination. *Camb. J. Econ.*, 47(1): 45–66.

Guo, Y.; Rennard, V.; Xypolopoulos, C.; and Vazirgiannis, M. 2021. BERTweetFR: Domain adaptation of pre-trained language models for French tweets. *arXiv preprint arXiv:2109.10234*.

Henriksen, I. M.; Throndsen, W.; Ryghaug, M.; and Skjølsvold, T. M. 2021. Electric vehicle charging and end-user motivation for flexibility: a case study from Norway. *Energy, Sustainability and Society*, 11: 1–10.

Hernández-Fuentes, A.; and Monnier, A. 2022. Twitter as a source of information? Practices of journalists working for the French national press. *Journal. Pract.*, 16(5): 920–937.

IEA. 2020. Global energy-related CO2 emissions by sector.

Jouët, J. 2018. Digital feminism: Questioning the renewal of activism. *J. Res. Gend. Stud.*, 8(1): 133–157.

Kloo, I.; Cruickshank, I. J.; and Carley, K. M. 2024. A Cross-Platform Topic Analysis of the Nazi Narrative on Twitter and Telegram During the 2022 Russian Invasion of Ukraine. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, 839–850.

Krishna, G. 2021. Understanding and identifying barriers to electric vehicle adoption through thematic analysis. *Transp. Res. Interdiscip. Persp.*, 10: 100364.

Kubin, E.; and Von Sikorski, C. 2021. The role of (social) media in political polarization: a systematic review. *Ann. Int. Commun. Assoc.*, 45(3): 188–206.

Lashari, Z. A.; Ko, J.; and Jang, J. 2021. Consumers' intention to purchase electric vehicles: Influences of user attitude and perception. *Sustainability*, 13(12): 6778.

Maybury, L.; Corcoran, P.; and Cipcigan, L. 2022. Mathematical modelling of electric vehicle adoption: A systematic literature review. *Transp. Res. D: Transp. Environ.*, 107: 103278.

McArthur, J. A.; and White, A. F. 2016. Twitter chats as third places: Conceptualizing a digital gathering site. *Social media+ society*, 2(3): 2056305116665857.

McHugh, M. L. 2012. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3): 276–282.

Mellon, J.; and Prosser, C. 2017. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, 4(3): 2053168017720008.

Monroe, B. L.; Colaresi, M. P.; and Quinn, K. M. 2008. Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16(4): 372–403.

Münker, S.; Kugler, K.; and Rettinger, A. 2024. Zero-shot prompt-based classification: topic labeling in times of foundation models in German Tweets. *arXiv preprint arXiv:2406.18239*.

Priyam, T.; Ruan, T.; and Lv, Q. 2023. Demographic-Based Public Perception Analysis of Electric Vehicles on Online Social Networks. *Sustainability*, 16(1): 305.

Ruan, T.; and Lv, Q. 2022. Public perception of electric vehicles on reddit over the past decade. *Communications in Transportation Research*, 2: 100070.

Ruan, T.; and Lv, Q. 2023. Public perception of electric vehicles on Reddit and Twitter: A cross-platform analysis. *Transportation research interdisciplinary perspectives*, 21: 100872.

Sanguesa, J. A.; Torres-Sanz, V.; Garrido, P.; Martinez, F. J.; and Marquez-Barja, J. M. 2021. A review on electric vehicles: Technologies and challenges. *Smart Cities*, 4(1): 372–404.

Singh, V.; Singh, V.; and Vaibhav, S. 2020. A review and simple meta-analysis of factors influencing adoption of electric vehicles. *Transp. Res. D: Transp. Environ.*, 86: 102436.

Todorovic, M.; Aldakkhelallah, A.; and Simic, M. 2023. Managing Transitions to Autonomous and Electric Vehicles: Scientometric and Bibliometric Review. *World Electric Vehicle Journal*, 14(11): 314.

Tufekci, Z. 2017. *Twitter and tear gas: The power and fragility of networked protest*. Yale University Press.

Uthirapathy, S. E.; and Sandanam, D. 2023. Topic Modelling and Opinion Analysis On Climate Change Twitter Data Using LDA And BERT Model. *Procedia Computer Science*, 218: 908–917.

Wang, N.; Tian, H.; Zhu, S.; and Li, Y. 2022a. Analysis of public acceptance of electric vehicle charging scheduling based on the technology acceptance model. *Energy*, 258: 124804.

Wang, Y.; Chi, Y.; Xu, J.-H.; and Yuan, Y. 2022b. Consumers' attitudes and their effects on electric vehicle sales and charging infrastructure construction: An empirical study in China. *Energy Policy*, 165: 112983.

Zhang, X.; Yang, Q.; Albaradei, S.; Lyu, X.; Alamro, H.; Salhi, A.; Ma, C.; Alshehri, M.; Jaber, I. I.; Tifratene, F.; et al. 2021. Rise and fall of the global conversation and shifting sentiments during the COVID-19 pandemic. *Humanities and social sciences communications*, 8(1): 1–10.

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, in Discussion section.
   (e) Did you describe the limitations of your work? Yes, in Discussion section.
   (f) Did you discuss any potential negative societal impacts of your work? No because we did not identify any potential negative societal impacts.
   (g) Did you discuss any potential misuse of your work? No because we did not identify any potential misuse.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, in Methods section.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes.
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? NA.
   (b) Have you provided justifications for all theoretical results? NA.
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA.
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA.
   (e) Did you address potential biases or limitations in your theoretical framework? NA.
   (f) Have you related your theoretical results to the existing literature in social science? NA.
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA.
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA.
   (b) Did you include complete proofs of all theoretical results? NA.
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes, in Methods section.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes, in Methods section.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? Yes, in Table 2.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? Yes, in Methods section.
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes in Discussion section.
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? No, because cost is low.
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes, in Methods section.
   (b) Did you mention the license of the assets? No because as far as we know there is no license.
   (c) Did you include any new assets in the supplemental material or as a URL? No because it was not needed.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? No, we collected public data.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? No.
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? No, it will be discussed on the Zenodo dataset page.
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? No.
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? NA.
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA.
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA.
   (d) Did you discuss how data is stored, shared, and deidentified? Yes, anonymization is discussed in Methods section.

Appendix

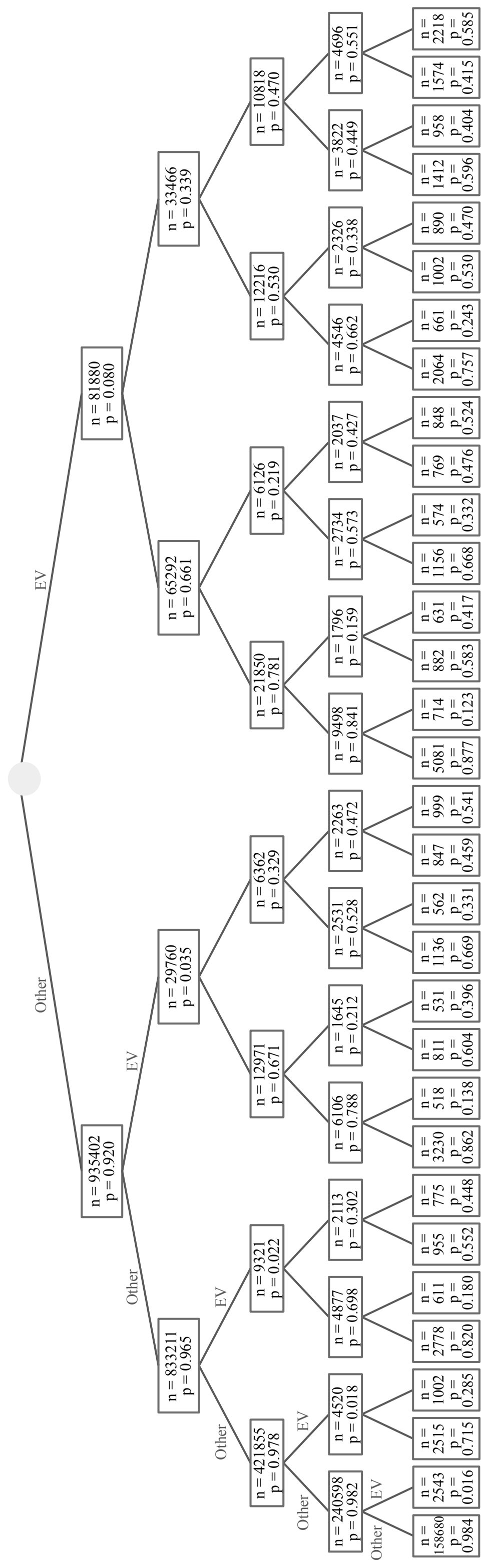

Figure 10: Probability tree of tweet classes ("EV" or "Other") based on their ancestors' classes. Left branches correspond to the class "Other" and right branches to the class "EV". The first level, after the root, includes tweets regardless of their ancestors. Their child includes tweets with at least one parent, and their probability is the conditional probability of the tweet's class depending on its parent class.

EV French terms	English translation	Non-EV French terms	English translation
électrique	electric	raciste	racist
batterie	battery	rn	rn
recharger	to recharge	soignant	caregiver
recharge	recharge	lfi	lfi
hydrogène	hydrogen	hanouna	hanouna
lithium	lithium	gréviste	striker
hybride	hybrid	soigner	treat/care
frandroid	frandroid	voix	voice
berline	sedan	mandat	mandate
peugeot	peugeot	racisme	racism
zoé	zoé	musulman	muslim
rechargeable	rechargeable	présidentiel	presidential
bmw	bmw	fasciste	fascist
volkswagen	volkswagen	langue	language
megane	megane	(le) pen	le pen
audi	audi	policier	policeman
rechargement	recharging	communisme	communism
hyundai	hyundai	islam	islam
nissan	nissan	fascisme	fascism
byd	byd	darmanin	darmanin
phev	phev	soldat	soldier
geeko	geeko	plainte	complaint
mustang	mustang	cnews	cnews
opel	opel	boyard	boyard
taycan	taycan	terrorisme	terrorism

Table 9: Most specific keywords of the EV tweet set and of the non-EV tweet set with their respective translation.