Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Trustworthy and Explainable AI for Electric Vehicle Reliability: A Four-Domain Analysis of Powertrain Systems and Intelligent Driving Experiences

Li Xinyu
Shinhan University
Qijia Liu
liuqijia679@gmail.com

Shinhan University

Research Article

Keywords: new energy vehicles, electrified powertrains, advanced driver-assistance systems, online consumer reviews, BERTopic, RoBERTa, negative binomial regression

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11722342/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Trustworthy and Explainable AI for Electric Vehicle Reliability: A Four-Domain Analysis of Powertrain Systems and Intelligent Driving Experiences

Li Xinyu[1, 2], **Liu Qijia**[1*]

[1] College of Engineering, Shinhan University, Uijeongbu 11644, Gyeonggi-do, Republic of Korea

[2] Xinjiang University of Foreign Languages, Kashgar, Xinjiang,844000, China

* Corresponding author: Liu Qijia. Email:liuqijia679@gmail.com

Abstract

As electrified powertrains and advanced driver-assistance systems (ADAS) develop rapidly, consumer assessments of new energy vehicles (NEVs) increasingly encompass powertrain performance, intelligent driving functionality, and operational safety rather than conventional driving attributes alone. Overall star ratings cannot adequately distinguish users' experience of battery range and charging reliability from their assessments of environmental perception, driving decisions, and human takeover. This study proposes a four-dimensional framework for analysing NEV owner reviews and integrates topic modelling, sentiment identification, structured rating cues, and community endorsement into an analytical design. The four domains comprise battery range and energy efficiency (T0), charging replenishment and electric-drive reliability (T1), driving-environment perception and scenario adaptation (T2), and driving decision-making and safety takeover (T3). The supplied analytical dataset comprises 16,500 records associated with the BYD Tang New Energy and AITO M9 vehicle labels. The reported results show that T0 accounts for 29.10% of dominant-topic records, whereas T3 displays the highest mean endorsement count (8.753). Negative binomial regression suggests heterogeneous conditional associations between technical themes and observable endorsement, with a positive T3-by-rating-extremity interaction (β = 0.408). These descriptive and model-based patterns illustrate the importance of analytically separating topical prevalence, affective evaluations, and platform interactions. The study contributes a vehicle-specific construct system, an integrated natural language processing and count-regression workflow, and a foundation for interpreting technology-specific user feedback. These findings must be interpreted in light of the underlying data provenance and the distinction between perceived experience and objectively verified technical safety.

Keywords: new energy vehicles; electrified powertrains; advanced driver-assistance systems; online consumer reviews; BERTopic; RoBERTa; negative binomial regression

1. Introduction

The global transition towards low-carbon mobility and increasing automotive intelligence is changing the relationship between consumers and vehicle technologies. Early development of new energy vehicles (NEVs) primarily concentrated on batteries, electric drives, range, and energy cost. As artificial intelligence, environmental sensing, onboard computing, and intelligent control have become integrated with vehicle architecture, NEVs have increasingly evolved into connected mobility systems that combine energy management, environmental perception, assisted decision-making, and human–machine interaction [1,2,19–21]. The broadening of advanced driver-assistance systems (ADAS) means that vehicle value is no longer determined solely by mechanical performance or energy efficiency. It also involves the ability to recognise road environments, perform reliably across scenarios, support driving decisions, and communicate operational limits. Consumer evaluation has consequently shifted from conventional judgments concerning acceleration, comfort, and operating cost to multidimensional assessments of electrification, driver assistance, system reliability, and perceived safety. Technology-acceptance research identifies perceived usefulness, ease of use, and trust as important beliefs in adoption [7,22,41]. Yet intentions measured before purchase do not necessarily capture the judgments users form after repeated driving experiences, charging encounters, and software updates.

This limitation is particularly salient for the interaction between electrified powertrains and ADAS. Real-world driving range depends on ambient temperature, traffic conditions, load, driving style, and battery-management

strategies, while the convenience of charging depends on charger availability, compatibility, waiting time, and power delivery [2,19,20,29,30]. A consumer may appreciate urban energy efficiency and acceleration but remain dissatisfied with winter range loss or charging delays. ADAS creates a different class of experience. Lane keeping, adaptive cruise control, assisted lane changing, and parking assistance may reduce perceived workload under ordinary conditions, but complex intersections, construction zones, poor weather, unexpected braking, or poorly timed takeover requests may undermine confidence [3,15–17,22,26,27]. It is therefore essential to distinguish the perceived quality of recognising an environment from the perceived appropriateness of subsequent control actions. Moreover, subjective descriptions of assisted driving must not be treated as proof of full automation or verified system safety: operational responsibility depends on the relevant automation level and design domain [4]. These developments motivate separate yet integrated constructs for range and efficiency, charging and electric-drive reliability, environmental perception, and decision-making and takeover.

Online automotive communities provide a setting in which these post-adoption judgments are expressed in natural language. Review platforms, including Autohome, present vehicle attributes, owner narratives, overall ratings, posting dates, and various forms of public interaction. Compared with standardised survey scales, reviews often describe concrete conditions of use and the specific technical performance that motivates a favourable or unfavourable assessment [9,10,31–34]. They also allow researchers to distinguish what consumers discuss, how they express their evaluations, and whether other visitors endorse those accounts. However, prior NEV research has largely examined adoption intention, range anxiety, infrastructure constraints, and overall satisfaction, whereas review mining frequently focuses on keywords, aggregate sentiment, or overall rating levels [1,2,21,39–42]. These approaches provide valuable evidence but may obscure differences between powertrain-related and ADAS-related assessments. A review can express satisfaction with charging convenience while criticising system warnings; conversely, a relatively rare description of unsafe-feeling takeover behaviour may attract attention beyond its frequency. Accordingly, topic prevalence should not be equated with either satisfaction or information usefulness, and textual sentiment should not be treated as interchangeable with a structured star rating.

Against this background, the study develops an analytical framework integrating topic identification, affective classification, and statistical modelling of observable community endorsement. The four theoretical domains are battery range and energy efficiency (T0), charging replenishment and electric-drive system reliability (T1), driving-environment perception and scenario adaptation (T2), and driving decision-making and safety takeover (T3). Topic modelling is used to organise technology-related review content; Chinese-language RoBERTa is considered for identifying affective orientation; and negative binomial regression examines the associations between topic membership, sentiment, rating extremity, and endorsement counts, with attention to review length, exposure, and vehicle group. The framework distinguishes topic content from evaluative cues and models the associations between those cues and observable platform behaviour rather than treating all indicators as a single satisfaction construct. In line with this objective, the study addresses three research questions:

RQ1. What technical evaluation themes concerning electrified powertrains and advanced driver assistance are expressed in NEV owner reviews, and can they be organised into four conceptually distinguishable domains?

RQ2. How are the technical domains and textual sentiment associated with observable review endorsement, and do the patterns differ between powertrain and driver-assistance content?

RQ3. Does the association between each technical domain and endorsement vary according to star-rating extremity, and are these conditional associations heterogeneous across domains?

The study makes three contributions. Conceptually, it shifts attention from adoption intention and overall satisfaction towards technology-specific judgments formed after use. By differentiating powertrain characteristics from environmental perception and decision-making, the four-dimensional framework accommodates the heterogeneous benefits and risks of intelligent electrified vehicles. Methodologically, it links unstructured review text to structured rating indicators and count outcomes through an interpretable analysis workflow, while explicitly separating substantive content from affective cues and platform interactions. Managerially, its taxonomy can help manufacturers organise feedback on range, charging, ADAS perception, and takeover performance into engineering and communication priorities. Nevertheless, observational correlations and review endorsement should not be interpreted as objective fault rates, verified safety outcomes, or causal effects. These contributions

are best understood alongside the limitations of the data source, sampling frame, and variable construction [9,10,16,17].

2. Literature Review and Theoretical Framework

2.1. Consumer Evaluation of Electrified Powertrains and Intelligent Driving Technologies

As NEV technologies have progressed from electrification towards integrated electrification and intelligence, consumer evaluations have become more multidimensional, context-sensitive, and technically differentiated. Conventional vehicle assessment often emphasises engine performance, fuel economy, ride comfort, and reliability. For NEVs, consumers also assess battery utilisation, charging access, electric-drive stability, the usability of assistance functions, and perceived operational safety [1,2,19,20,39,40]. The value of an electrified powertrain lies not only in nominal range and efficiency but also in whether drivers can complete intended journeys reliably under actual conditions. Driving range is sensitive to temperatures, route conditions, vehicle load, driving habits, and battery-management strategy. In particular, winter driving, highway travel, and lengthy trips can expose differences between advertised and realised performance [20,29,30]. Similarly, charger coverage, replenishment speed, compatibility, and the stability of power delivery contribute to consumers' assessments of convenience and long-term utility. Satisfaction with driving range can therefore coexist with dissatisfaction over charging queues or unstable replenishment, underscoring the need to distinguish energy consumption from replenishment and power-delivery experience.

Two dimensions characterise user perceptions of the powertrain. T0, battery range and energy efficiency, captures assessments of realised mileage, range-estimation accuracy, seasonal changes in consumption, energy use, and regenerative braking. It represents perceived operating efficiency and trip reliability. T1, charging replenishment and electric-drive system reliability, captures the accessibility and stability of charging services, charging speed, equipment compatibility, electric-motor response, power delivery, and system dependability. This distinction limits the tendency to attribute heterogeneous issues to 'battery performance' as a broad label. Differences in powertrain configuration also matter: battery-electric, plug-in hybrid, and range-extended systems differ in energy pathways, replenishment modes, and the meaning of pure-electric versus combined range. Comparisons should therefore account for actual vehicle specifications and operating mode rather than equating technically different range indicators [2,21,30].

ADAS evaluations depend more directly on dynamic interactions among vehicles, road geometry, traffic participants, weather, and human drivers. Environmental perception and sensor fusion allow systems to assist with lane following, adaptive cruise control, lane changes, and parking; nonetheless, performance and perceived comfort can change across urban roads, congested intersections, construction areas, and low-visibility conditions [3,15,22,35,36]. Human-factors research further emphasises that appropriately calibrated trust requires drivers to understand system limits and maintain the ability to monitor or regain control [16,17,26,27]. Drivers may assess obstacle detection favourably while expressing concern about abrupt control actions, false braking, or insufficiently clear takeover alerts. Consequently, T2, driving-environment perception and scenario adaptation, comprises perceived lane and obstacle recognition, road understanding, highway and urban assistance, parking, and operation in adverse conditions. T3, driving decision-making and safety takeover, focuses on braking and acceleration choices, following distances, alerts, unexpected disengagement, driver monitoring, and manual interventions. T2 relates primarily to the perception of environmental information, whereas T3 reflects the reasonableness and predictability of control and transfer-of-control behaviour. Both are user perceptions rather than validated engineering measurements; positive reviews do not demonstrate compliance with safety standards, and negative reviews alone do not establish defect incidence [4,16,17].

Recent applications published in Discover Applied Sciences address battery thermal management, charging infrastructure, autonomous-driving anomaly detection, charging-demand prediction, and state-of-charge estimation [56–60]. Related work in Discover Internet of Things explores battery-state estimation, coordinated fast-charging reservations, connected thermal management, intelligent vehicle path tracking, and uncertainty in connected devices [61–65]. These papers broaden the engineering and digital-infrastructure context for the framework; their relevance does not imply that review perceptions are objective evaluations of technical safety.

The four domains share a common foundation in post-adoption experience but differ in technological mechanism, operational context, and risk sensitivity. Powertrain evaluations relate primarily to energy use and mobility convenience; ADAS evaluations involve scene understanding, driver–system coordination, and safety expectations. A single overall star rating may mask these distinctions, including the possibility that one reviewer praises a feature and criticises another. The proposed framework thus provides a domain-specific classification for subsequent topic modelling, sentiment analysis, and endorsement modelling, with the understanding that real review narratives can contain multiple themes and require transparent dominant-topic or multi-label annotation rules.

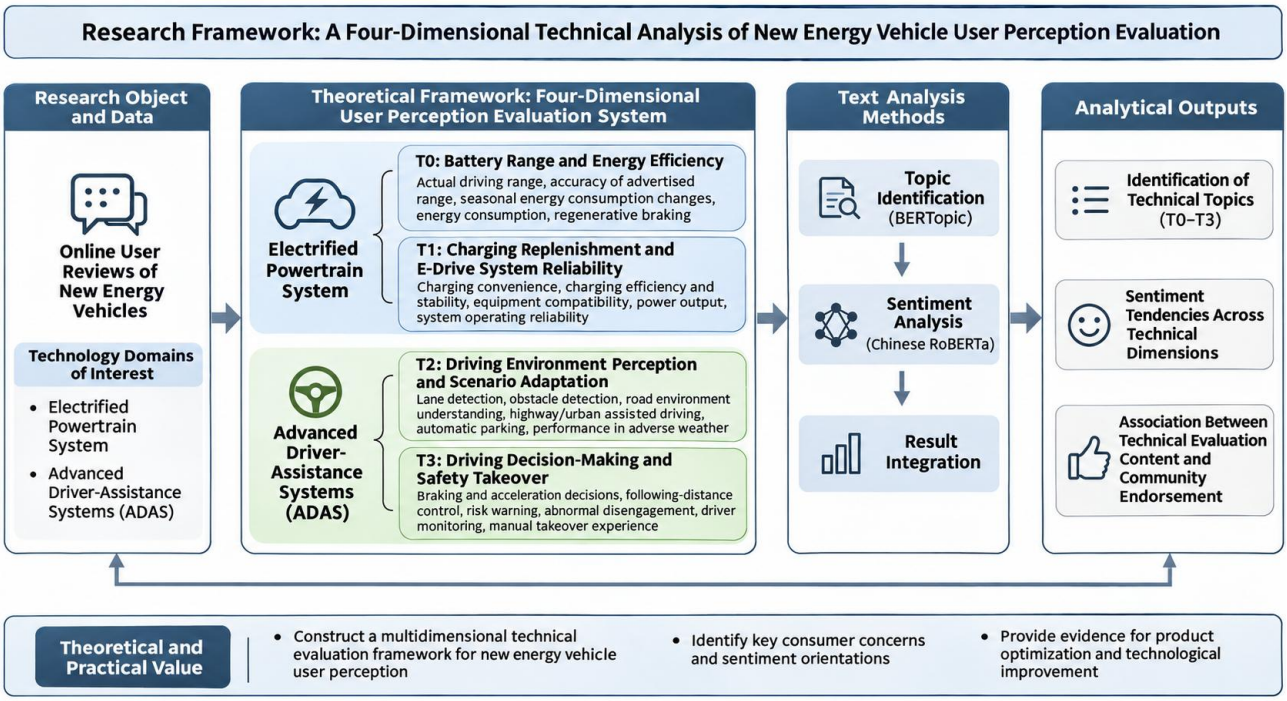


Figure 1. Four-domain research framework for NEV technical user perceptions.

2.2. Theoretical Foundations: Consumption Values, Technology Acceptance, and Review Diagnosticity

The Theory of Consumption Values (TCV) holds that consumer choice and product evaluation involve multiple sources of value rather than a single functional utility [5]. The PERVAL framework similarly distinguishes quality or performance, emotional, social, and price-related dimensions [6]. In the context of NEVs, users evaluate not only whether a car fulfils travel needs but also its energy efficiency, charging convenience, assistance functionality, and reliability. A driver may value low energy consumption but feel frustrated by charging availability, or appreciate automated assistance while expressing unease about sudden braking. TCV therefore provides an overarching perspective for interpreting the simultaneous benefits, sacrifices, and perceived risks expressed in vehicle reviews. The four domains T0–T3 are not claimed to constitute a validated TCV psychometric scale; they are vehicle-specific content constructs interpreted with the aid of relevant theories.

The Technology Acceptance Model (TAM) highlights perceived usefulness and perceived ease of use as central beliefs in technology adoption [7]. These concepts illuminate why consumers value convenient charging, stable ADAS operation, and comprehensible takeover instructions. Importantly, continued product use extends evaluation beyond pre-adoption beliefs to realised performance. For a powertrain, usefulness can relate to adequate range and reliable replenishment; for ADAS, it can involve reduced driving burden and stable assistance under supported conditions. Ease of use can relate to charging procedures, system interfaces, and recognition of operating limitations. Nevertheless, text-mined themes are not automatically equivalent to TAM latent constructs, and review language cannot be regarded as a direct validated measure of behavioural intention. TAM is used as an interpretive lens, not as an assumed measurement identity [7,22,41].

The Elaboration Likelihood Model (ELM) distinguishes relatively detailed processing of substantive arguments from judgments based more heavily on accessible peripheral cues [8]. In automotive review settings, specific descriptions of winter range loss, charging failures, false braking, or takeover problems can provide content-rich information; sentiment and star ratings may function as more immediate evaluative cues. Previous research has associated review depth, rating extremity, and perceived diagnosticity with review usefulness and information adoption [9,10,32–34]. It is therefore analytically appropriate to include substantive topic content alongside rating direction, extremity, text length, and other characteristics. The count of thumbs-up votes or likes constitutes observable platform endorsement, however, and should not be interpreted as a direct measure of trust, diagnosticity, or a reader's cognitive route. Together, TCV explains the multidimensional content of value judgments; TAM helps explain technology-related utility and convenience; and ELM frames how substantive and heuristic signals may coexist in the processing of online reviews.

Table 1. Selected Relevant Studies and Their Implications for the Analytical Framework

Study	Context and method	Contribution to the framework	Scope of application
Egbue and Long [1]	EV adoption; consumer perceptions	Identifies range and technology-related barriers	Adoption barriers are not direct measurements of post-use performance
Hardman et al. [2]	Review of EV charging preferences	Separates charging access from driving experience	Infrastructure findings depend on the context of use
Kyriakidis et al. [3]	International automated-driving survey	Connects driving automation with public concerns	Not a model-specific safety assessment
Mudambi and Schuff [9]	Online consumer reviews	Relates review depth and extremity to helpfulness	Likes are not identical to perceived helpfulness
Filieri [10]	Review diagnosticity	Explains information adoption through review quality	Endorsement alone does not measure diagnosticity
Grootendorst [11]; Liu et al. [12]	Topic modelling and language models	Methodological basis for BERTopic and RoBERTa	Predictive performance requires independent evaluation

Table 2. Theoretical Perspectives and Correspondence With Analytical Components

Theoretical perspective	Principal proposition	Correspondence to four-domain framework	Interpretive limitation
Theory of Consumption Values [5]	Value is multidimensional	Functional, experiential and risk-related content	Technical domains are not validated TCV factors
Technology Acceptance Model [7]	Usefulness and ease of use influence acceptance	Charging convenience and ADAS usability	Review content does not directly measure intention
Elaboration Likelihood Model [8]	Central and peripheral information processing	Topic content versus sentiment and rating cues	Model coefficients do not establish cognitive routes
Review diagnosticity [9,10]	Informative content may shape adoption of reviews	Endorsement count and review depth	Likes represent observable interactions, not diagnosticity
Human factors of automation [16,17]	Appropriate reliance and safe transitions matter	T2 perception and T3 decision/takeover	Subjective sentiment does not prove objective safety

2.3. Research Propositions and Domain Classification

The proposed technical taxonomy is operationalised through representative review expressions and their semantic boundaries (Table 3). The terminology serves as an initial coding vocabulary rather than proof that statistically coherent topics have already been discovered through an unsupervised model. In authentic reviews, conceptual overlap is expected: a sentence such as 'charging is so slow that I have lost confidence in the vehicle' may involve T1, negative sentiment, and a general trust judgment simultaneously. A rigorous analysis must specify whether the unit of classification is the full review or a sentence, and whether exclusive dominant-topic labels or multiple labels are used. Coding rules should be established before comparative analysis rather than forcing ambiguous content into mutually exclusive categories without justification.

Table 3. Four-Domain Evaluation Vocabulary and Semantic Boundaries

Domain	Representative Chinese review terms (English equivalents)	Conceptual content	Ambiguity requiring control
T0: Range and efficiency	Actual range; highway	Expected versus realised range	Pure-electric versus combined

Domain	Representative Chinese review terms (English equivalents)	Conceptual content	Ambiguity requiring control
	consumption; winter range; energy use; range attainment; regenerative braking	and energy consumption	driving range
T1: Charging and drive reliability	Charging speed; fast-charge compatibility; charging convenience; powertrain response; e-drive smoothness; battery management	Replenishment and consistent propulsion	Charger network limitations versus vehicle faults
T2: Perception and adaptation	Lane recognition; obstacle detection; urban navigation assistance; highway assistance; parking; night-time detection	Detection and scenario-dependent assistance	Advertised function versus actual availability
T3: Decision and takeover	Takeover reminder; false braking; following distance; braking decisions; disengagement; driver monitoring	Intervention, handover, decision-making and perceived risk	Subjective worry versus verified hazardous event

On this basis, three propositions are advanced. **P1:** The relative prevalence of powertrain and ADAS themes may vary across vehicle configurations and usage contexts. **P2:** After accounting for exposure time, review depth, and evaluative cues, the relationship between safety-takeover content and endorsement may differ from that of routine range content. **P3:** Star-rating extremity may condition the association between technical domain and endorsement, with heterogeneous patterns across domains. These propositions clarify the analytical questions addressed by the framework; whether they are supported requires appropriately obtained and validated empirical evidence.

3. Research Methodology

3.1. Research Design, Data Sources, and Preprocessing

The analytical procedure comprises five linked stages (Figure 2): review-data acquisition and preprocessing, technical-topic identification, construction of the four evaluation domains, sentiment classification and measurement of evaluative cues, and statistical estimation with robustness checks. At the first stage, vehicle characteristics, review text, star ratings, dates, and available interaction counts are organised within a consistent data structure. Duplicate and low-information entries should be screened according to documented criteria; the language and technical terminology of Chinese automotive comments require particular attention. At the second stage, unsupervised topic models identify semantically related expressions. At the third stage, the resulting topics are interpreted against the four vehicle-specific domains, T0–T3. The fourth stage relates textual sentiment to structured star ratings, rating extremity, and endorsement counts. Finally, count regression and sensitivity analysis assess conditional associations across technological domains and evaluative contexts. This workflow integrates natural-language processing with statistical modelling without assuming that textual content, expressed satisfaction, and observable community response are identical constructs [11–14,32–34].

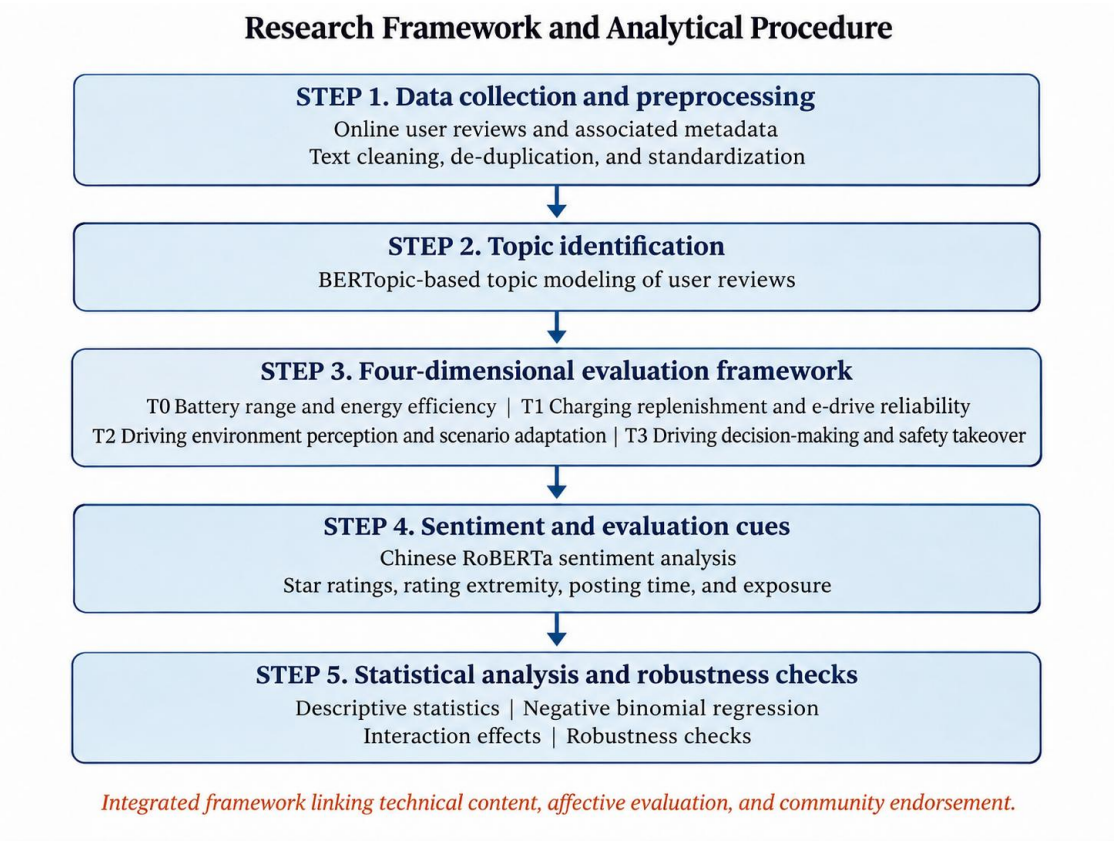


Figure 2. Research framework and analytical procedure.

Autohome's publicly accessible vehicle-review interfaces provide a reference for the fields potentially available in an automotive review study. The vehicle pages for BYD Tang New Energy (https://k.autohome.com.cn/3430/21522) and AITO M9 (https://k.autohome.com.cn/7207/ge11/) illustrate the presentation of model names, configurations, overall ratings, written comments, posting dates, and selected interaction measures. The two vehicle groups represent technologically different configurations and product positioning and thus provide a meaningful setting for defining a shared technical taxonomy. The platform interfaces themselves, however, do not establish a particular extraction procedure or independently authenticate a research database. A fully reproducible study should disclose its collection permissions and dates, accessible fields, sampling criteria, preprocessing script, and matching rules across textual and structured records. Figure 3 illustrates the relationship between the review interface and the analytical dataset.



Figure 3. Automotive review interface and analytical variable structure.

The document describes an analytical dataset comprising 16,500 Chinese-language review records, with 8,250 observations assigned to each of the two vehicle groups and a specified review period from 1 January 2024 to 31 October 2026. A review-level analytical record includes an identifier, vehicle name or configuration, comment text, rating, posting date, interaction count, and constructed fields for domain membership, sentiment, rating extremity, review length, and exposure duration. The main thematic variables differentiate T0 (range and efficiency), T1 (charging and electric-drive reliability), T2 (environmental perception and adaptation), and T3 (decision-making and takeover). Textual sentiment records affective direction, while the rating represents a voluntarily selected overall evaluation. Endorsement count is the non-negative response variable in the count models. Exposure-related measures, comment length, rating direction, and vehicle-group indicators are treated as covariates because interaction opportunities and review characteristics may differ across observations. Substantive interpretation requires observing actual vehicle type, operating mode, ADAS availability, and time-varying software characteristics where possible, particularly before attributing differences between the two vehicle labels to technical quality [2,4,19,30].

3.2. Determining the Number of Topics: LDA Model Comparison

To evaluate an appropriate number of latent themes, the study compares Latent Dirichlet Allocation (LDA) specifications with K = 3, K = 4, and K = 5 under a consistent Chinese-text preprocessing strategy [18]. Model selection considers validation-set perplexity, semantic coherence, and the substantive interpretability of domain boundaries. Perplexity quantifies the predictive fit of a topic model for held-out tokens: smaller values indicate more favourable out-of-sample likelihood, all else equal. Let D_test contain M validation documents, N_d denote the number of tokens in document d, and p(w_d | K) be the predictive likelihood of the document given K themes. The measure is defined as follows:

$$Perplexity(K) = exp[-(\Sigma_{d \in Dtest} \ln p(w_d \mid K)) / (\Sigma_{d \in Dtest} N_d)]$$

Because increased topic complexity can improve likelihood without improving conceptual clarity, the analysis also considers relative reductions in perplexity when the topic number rises from K to K + 1:

$$\Delta Perplexity(K \rightarrow K+1) = [Perplexity(K) - Perplexity(K+1)] \,/\, Perplexity(K) \times 100\%$$

The three-theme solution risks combining distinct aspects of range, charging, and ADAS performance, whereas a five-theme solution may fragment a technical domain into overlapping semantic subtopics. The four-theme interpretation is substantively aligned with the domain-specific taxonomy developed in Section 2.1. Table 4 presents the comparative values reported in the supplied manuscript. The reported perplexities are 865.42, 742.68, and 728.35 for K = 3, 4, and 5, respectively, while the corresponding coherence scores are 0.512, 0.638, and 0.601. Relative to K = 3, the four-topic solution reduces perplexity by approximately 14.18% and increases topic coherence; increasing K from four to five reduces perplexity by approximately 1.93% but is accompanied by lower coherence and potential semantic redundancy. On these reported criteria, K = 4 offers the most defensible balance of parsimony, interpretability, and fit. These values are reproduced from the supplied comparison table and have not been independently recalculated from original review records in preparation of this English manuscript. The selection remains contingent on consistent train–validation splits, vocabulary construction, random seeds, and independently inspected topic words.

Table 4. Comparison of LDA Specifications With Three, Four, and Five Topics

Candidate topic number	Validation perplexity ↓	Topic coherence C_v ↑	Main interpretive check	Overall assessment
K = 3	865.42	0.512	Range and charging content may merge; limited technical differentiation	Under-partitioned
K = 4	742.68	0.638	Four technical themes retain interpretable boundaries	Preferred trade-off
K = 5	728.35	0.601	ADAS content may become redundant or excessively fragmented	Potential over-partitioning

Note: Lower perplexity (↓) and higher coherence (↑) are preferred; values are reproduced from the supplied comparison table and have not been independently recomputed.

The four interpretive domains are T0, battery range and energy efficiency; T1, charging replenishment and electric-drive system reliability; T2, driving-environment perception and scenario adaptation; and T3, driving decision-making and safety takeover. For each candidate specification, the most probable topic words, representative texts, and overlap across clusters should be inspected. A four-topic specification is preferable only if these distinctions are supported by the fitted model; a prespecified taxonomy is not in itself empirical proof of four independently discovered topics. This distinction is particularly relevant when comparing LDA and BERTopic because their word distributions and embedding-based clusters are generated by different statistical procedures [11,18].

3.3. BERTopic-Based Identification of Technical Themes and Semantic Analysis

Online NEV reviews frequently contain short and heterogeneous descriptions of technical performance. Expressions such as 'battery consumption rises in winter' and 'cold-weather range is much shorter' may convey similar meanings despite limited overlap in exact word forms. BERTopic combines contextual document embeddings, dimensionality reduction, density-based clustering, and class-based TF–IDF (c-TF–IDF) to capture semantic proximity and produce interpretable representations [11,13]. Within the proposed workflow, it serves to assess whether the textual discourse can be linked to the four theoretical domains while allowing potentially distinct subtopics to emerge from the data. A document-level semantic vector is generated by a Chinese-language sentence encoder. Let d_i denote review i and f_θ the embedding function; then:

$$h_i = f_\theta(d_i), \quad h_i \in \mathbb{R}^{D} \qquad (9)$$

Here h_i is a D-dimensional representation of the contextual meaning of review i. Such representations help capture semantically related statements involving range reduction, charging incompatibility, false braking, or driver takeover even when consumers use different expressions. The choice of pretrained checkpoint, tokeniser,

domain adaptation, and maximum sequence length affects reproducibility and should be documented rather than inferred from a generic model name [11–13].

UMAP reduces high-dimensional embeddings while aiming to preserve relevant local geometry; HDBSCAN subsequently identifies dense semantic clusters and may designate outliers without a reliable assignment. The process can be summarised as follows:

$$u_i = UMAP_\psi(h_i); \quad z_i = HDBSCAN_\varphi(u_i) \quad (10)$$

In this notation, u_i is the reduced representation, z_i is the preliminary cluster label, and ψ and φ are the respective dimensionality-reduction and clustering hyperparameters. Crucially, BERTopic need not produce exactly four initial clusters. The selected four-domain taxonomy is an interpretive framework for organising relevant subtopics, not a requirement that all texts be forcibly assigned to four density clusters. Ambiguous, cross-domain, and outlying reviews should be retained for adjudication or analysed under an explicit alternative coding strategy.

For topic interpretation, c-TF–IDF aggregates texts associated with a cluster into a class document and identifies words that are representative of that cluster relative to the collection of topic documents. A normalised form of the keyword-weighting procedure is represented below:

$$c - TF - IDF_\{r,w\} = [f_\{r,w\} / \Sigma_\{v \in V\} f_\{r,v\}] \cdot log(1 + A/f_w) \quad (11)$$

In the expression, f_{r,w} is the frequency of word w in the merged text of cluster r, V is the vocabulary, f_w denotes the total frequency of word w across the class documents, and A is the average class-document length. The resulting weights guide, but do not determine, substantive topic labels. Range, cold-weather energy use, and regenerative braking are candidate semantic cues for T0; fast charging, charger compatibility, and electric-motor stability relate to T1; lane recognition, obstacle detection, and parking assistance relate to T2; and braking decisions, risk alerts, disengagement, and manual intervention relate to T3. The final high-weight words must be obtained from fitted model output rather than substituted from the conceptual vocabulary.

Table 5. Four Technical Evaluation Domains and Their Semantic Boundaries

Topic	Technical evaluation domain	Representative semantic cues	Core review content
T0	Battery range and energy efficiency	Range; cold-weather electricity use; regenerative braking	Realised driving range, energy consumption and management
T1	Charging replenishment and electric-drive reliability	Fast charging; compatibility; charging efficiency; motor output	Charging convenience and stable e-drive operation
T2	Driving-environment perception and scenario adaptation	Lane recognition; obstacle detection; parking	Environmental detection and scenario-dependent assistance
T3	Driving decision-making and safety takeover	Braking decisions; alerts; manual takeover	Control response, driver coordination and operational limits

To relate initial BERTopic topics to the four theoretical domains, the analytical procedure combines cluster keywords, prototypical reviews, and semantic similarity with domain descriptions. Let v_r be the embedding representing initial topic r and g_k the embedding of theoretical domain k. Their cosine similarity is computed as:

$$Sim(r,k) = (v_r^{T} g_k) / (\|v_r\|_2 \|g_k\|_2) \quad (12)$$

The similarity score captures semantic proximity, but automated matching alone is insufficient when reviews discuss overlapping components. For example, an ADAS comment may describe both pedestrian detection and the vehicle's subsequent braking decision. The mapping should therefore be supported by a codebook, exclusion criteria, review-level contextual reading, and independent human checks. Let m(r) denote the validated mapping from an eligible initial cluster to an interpretive domain. Then the domain label associated with review i is given by:

$$k_i = m(z_i), \quad k_i \in \{T0, T1, T2, T3\} \quad (13)$$

The mapping is defined only for comments that can be meaningfully classified; outliers and ambiguous records require separate handling. When mutually exclusive dominant-domain labels are used, the share of valid classified reviews belonging to domain k is:

$$R_k = (N_k / N*) \times 100\%, \quad N_k = \Sigma_\{i = 1\}\^\{N*\} I(\hat{k}_i = k) \quad (14)$$

In this equation, N_k is the number of eligible reviews assigned to domain k and N* is the number of reviews receiving a valid classification. Topic shares describe the frequency of particular subjects within the analysed review set; they do not directly measure the importance of the subject for satisfaction, purchase decisions, or safety. Reliability should be examined using within-topic coherence, between-topic diversity, representative-text inspection, and independently coded validation subsets. Cohen's κ or an appropriate multi-label agreement statistic can quantify agreement between human coders; disagreements should be adjudicated before finalisation. Alternative random seeds, dimensionality-reduction settings, and topic-merging thresholds provide further tests of semantic stability.

After technical-domain mapping, Chinese-language RoBERTa can be used to identify positive, neutral, and negative sentiment from actual review text. Unlike dictionary-based counting, contextual classification can respond to negation, contrast, and conditional praise, although domain adaptation and manual validation are essential [12]. Sentiment and star ratings are conceptually distinct: a user may award a high overall rating while criticising a specific technical feature. They should accordingly be preserved as separate evaluative indicators. Precision, recall, F1 score, confusion matrices, and inter-annotator agreement should be reported from independently labelled evaluation data if used to substantiate a claim about classification performance; such estimates cannot be supplied solely from the definitions of the models. Together, semantic topic analysis, affective classification, and review-level metadata support the subsequent comparison of technical content, expressed evaluations, and community endorsement.

4. Results

4.1. Four-Domain Technical Topic Structure and Sentiment Distribution

The descriptive analysis distinguishes topic prevalence, affective orientation, overall star ratings, and observable community endorsement. Table 6 summarises 16,500 analytical records classified into the four technical domains. Battery range and energy efficiency (T0) account for 4,802 records, or 29.10% of the sample. Charging replenishment and electric-drive reliability (T1) represent 4,271 records (25.88%), driving-environment perception and scenario adaptation (T2) account for 4,190 (25.39%), and driving decision-making and safety takeover (T3) account for 3,237 (19.62%). These proportions show that powertrain-related descriptions constitute a substantial share of the analysed text, while perceptions of ADAS also form a prominent part of technical discourse. The distribution of themes provides a basis for evaluating whether subject prevalence is associated with user attitudes and interaction counts; it does not in itself indicate whether the corresponding vehicle technology is objectively better or worse.

Table 6. Distribution and Affective Characteristics of Four Technical Evaluation Domains

Technical domain	Reviews (n)	Topic share (%)	Positive sentiment (%)	Mean stars	Mean endorsements (SD)
T0 Range and efficiency	4,802	29.10	83.01	4.125	2.083 (2.213)
T1 Charging and e-drive reliability	4,271	25.88	78.01	4.041	3.157 (3.200)
T2 Perception and adaptation	4,190	25.39	77.16	4.000	4.328 (4.189)
T3 Decision and takeover	3,237	19.62	71.21	3.870	8.753 (8.377)
Overall	16,500	100.00	77.92	4.021	4.239 (5.287)

Note: SD = standard deviation. Small deviations in percentage totals may result from rounding.

Positive sentiment varies across the four domains. T0 displays the highest proportion of positive classifications (83.01%), followed by T1 (78.01%) and T2 (77.16%); T3 has the lowest positive proportion (71.21%). This variation is consistent with the proposition that consumer evaluation is technology-specific rather than homogeneous across the vehicle. In particular, comments about decision-making and takeover warrant closer qualitative inspection because these domains involve perceived risk, predictability, and driver expectations. The descriptive difference alone, however, does not provide an estimate of verified failures, objective system safety, or overall technology acceptance. It indicates that different groups of statements convey different affective profiles within the analysed records.

The structured ratings show a broadly similar descriptive ordering. The overall mean rating is 4.021 stars; T0 receives 4.125, T1 receives 4.041, T2 receives 4.000, and T3 receives 3.870. Although these averages align directionally with the positive-sentiment proportions, sentiment reflects what review language conveys about a particular experience whereas stars express the reviewer's general rating signal. Their correlation should not be regarded as identity. An analytical framework combining topic classification and affective signals can recover differences that might be obscured by a single overall rating, but the strength of any association must be established through the corresponding fitted models rather than inferred from group averages alone.

4.2. Comparison of the Two Vehicle Groups

The descriptive comparison considers technical-topic composition, positive sentiment, star ratings, and endorsement counts for the BYD Tang New Energy and AITO M9 groups. According to Table 7, each group comprises 8,250 records. The AITO M9 group has a positive-sentiment proportion of 82.19%, a mean rating of 4.112, and an average endorsement count of 4.842. The BYD Tang group has corresponding values of 73.64%, 3.930, and 3.637. These figures describe the analysed groups and should not, on their own, be used as rankings of technical quality or purchasing recommendations. Differences may reflect vehicle specification, owner characteristics, review contexts, and the composition of the underlying records.

Table 7. Review Characteristics and Topic-Level Comparisons Across Vehicle Groups

Vehicle group and topic	Reviews (n)	Positive sentiment (%)	Mean stars	Mean endorsements
BYD Tang New Energy — All	8,250	73.64	3.930	3.637
BYD Tang — T0	2,766	82.57	4.095	1.997
BYD Tang — T1	2,439	75.03	3.986	2.972
BYD Tang — T2	1,708	65.69	3.759	3.857
BYD Tang — T3	1,337	62.75	3.707	7.960
AITO M9 — All	8,250	82.19	4.112	4.842
AITO M9 — T0	2,036	83.60	4.165	2.199
AITO M9 — T1	1,832	81.99	4.115	3.403
AITO M9 — T2	2,482	85.05	4.166	4.651
AITO M9 — T3	1,900	77.16	3.985	9.310

Topic composition differs visibly between groups. In the BYD Tang group, T0 and T1 comprise 2,766 and 2,439 records, accounting jointly for approximately 63.09% of the group. In the AITO M9 group, T2 and T3 contain 2,482 and 1,900 records, jointly accounting for approximately 53.12%. The Pearson chi-square test for the two-by-four contingency table yields $\chi^2(3) = 438.14$ ($p < 0.001$), indicating a difference in the distribution of assigned topics within the supplied analytical records. This test cannot establish the reasons for the difference or separate vehicle characteristics from selection processes. The two powertrain architectures, operating modes, winter versus summer travel, optional ADAS equipment, software releases, and ownership duration may confound a comparison if not observed and controlled.

Within both groups, T3 displays relatively high mean endorsement counts (7.960 and 9.310), whereas T0 is associated with lower mean counts (1.997 and 2.199). The comparison reinforces the distinction between topical prevalence and platform endorsement: the most frequently discussed subject is not necessarily the most interacted-with subject. Cross-vehicle comparisons therefore require both the proportions of topic categories and their within-category evaluation patterns. In particular, range values from vehicles with different electric and fuel-based

operating modes should not be treated as physically identical measurements, and assisted-driving functions should be compared only with attention to their actual availability and design limits.

4.3. Count Distribution, Correlation Structure, and Model Choice

Endorsement counts are non-negative integers, making their distribution relevant to selecting an appropriate regression family. The reported mean number of likes is 4.239, whereas the sample variance is 27.948—approximately 6.59 times the mean. The standard deviation is 5.287, the maximum is 72, and zero-endorsed records account for 16.8% of the dataset. Such features are consistent with overdispersion relative to the mean–variance equality assumed by the ordinary Poisson model and motivate the use of a negative binomial specification with an estimated dispersion parameter [14]. The average text length is 88.375 Chinese characters, and the mean logarithm of exposure time is 5.946. These covariates are considered because longer reviews may contain more diagnostic information, while reviews posted earlier can accumulate interactions over a longer period.

Table 8. Descriptive Statistics, Correlations, and Multicollinearity Diagnostics

Variable	Mean	SD	Pearson r with endorsements	VIF
Review endorsement count	4.239	5.287	1.000	—
Positive sentiment	0.779	0.415	0.106	3.68
Star rating	4.021	1.066	0.124	—
Rating extremity	1.310	0.681	0.175	1.23
Log exposure duration	5.946	0.980	0.164	1.00
Review length (characters)	88.375	3.747	0.134	1.15
Rating direction	0.656	0.667	0.089	3.37
Vehicle-group indicator	0.500	0.500	0.114	1.04

Note: VIF = variance inflation factor; correlations are Pearson coefficients. Topic-indicator VIFs for T1, T2 and T3 are approximately 1.44, 1.44 and 1.43, respectively.

Bivariate correlations with endorsement are generally small. The Pearson coefficients are 0.175 for rating extremity, 0.164 for log exposure duration, 0.134 for review length, 0.124 for stars, and 0.106 for positive sentiment. These associations do not imply independent conditional effects, as reviewers' sentiment and overall ratings may be related to each other and to omitted characteristics. Variance inflation factors (VIFs) are 3.68 for positive sentiment and 3.37 for rating direction; the listed values for the other covariates are lower, and the T1–T3 indicators have VIF values close to 1.44. Such diagnostics help identify potential estimation instability but cannot demonstrate the absence of omitted-variable bias, exposure selection, or model misspecification. This motivates modelling the substantive and peripheral evaluation cues jointly and retaining non-significant covariates in the reported specification.

4.4. Negative Binomial Regression Estimates and Model Comparison

Table 9 reports three nested negative binomial models fitted to the same set of 16,500 records. Model 1 includes rating extremity, rating direction, log exposure time, review length, and vehicle group. Model 2 adds the three technical-topic indicators T1–T3 with T0 as the reference domain. Model 3 further introduces positive sentiment and topic-by-rating-extremity interaction terms. The Akaike information criterion (AIC) decreases from 82,258.57 in Model 1 to 78,195.69 in Model 2 and then to 77,974.91 in Model 3. Within the specified candidate models and sample, this pattern favours the richer model in terms of relative goodness of fit after accounting for parameter complexity. AIC alone does not provide an external validation test, establish causal relationships, or show that the fitted associations generalise across platforms or vehicle configurations.

Table 9. Negative Binomial Regression Estimates for Review Endorsement Counts

Predictor	Model 1: Controls	Model 2: Topics	Model 3: Full
T1 vs. T0	—	0.423* (0.021)	0.229* (0.048)
T2 vs. T0	—	0.725* (0.021)	0.489* (0.047)

Predictor	Model 1: Controls	Model 2: Topics	Model 3: Full
T3 vs. T0	—	1.412* (0.021)	0.870* (0.047)
Rating extremity	0.318* (0.013)	0.292* (0.012)	0.087* (0.023)
T1 × extremity	—	—	0.143* (0.031)
T2 × extremity	—	—	0.175* (0.031)
T3 × extremity	—	—	0.408* (0.031)
Positive sentiment	—	—	0.223* (0.035)
Rating direction	0.030* (0.013)	0.119* (0.012)	0.012 (0.020)
Log exposure	0.261* (0.009)	0.258* (0.008)	0.256* (0.008)
Review length	0.0367* (0.0022)	0.0038 (0.0020)	0.0028 (0.0020)
AITO M9 indicator	0.242* (0.016)	0.106* (0.015)	0.100* (0.015)
Dispersion α	0.842	0.564	0.551
AIC	82,258.57	78,195.69	77,974.91
Observations	16,500	16,500	16,500

Note: Coefficients with model standard errors in parentheses. T0 is the reference topic. Asterisks reproduce the source-table indicator for $p < 0.05$; no further significance tiers are inferred. AIC = Akaike information criterion.

In Model 3, where rating extremity is zero (a three-star rating), the estimated coefficients for T1, T2, and T3 relative to T0 are 0.229, 0.489, and 0.870, respectively. Their incidence-rate ratios (IRRs) are approximately 1.257, 1.631, and 2.387. The positive-sentiment coefficient is 0.223 (IRR ≈ 1.250). The interactions of rating extremity with T1, T2, and T3 are positive, at 0.143, 0.175, and 0.408, with the largest point estimate for decision-making and takeover content. These estimates indicate that the difference in expected endorsement between a technical domain and the reference domain may vary by rating extremity. They should not be interpreted as evidence that the technical theme causes a review to be useful or that readers assign causal importance to a system's perceived safety. When interactions are included, the topic main effects apply specifically at the reference level of the interacting variable.

Not all included predictors exhibit statistically distinguishable associations. In Model 3, the rating-direction coefficient is 0.012 ($p \approx 0.547$), and the coefficient for review length is 0.0028 ($p \approx 0.169$), both failing to reach a conventional 5% significance threshold. Reporting these estimates alongside the larger topic coefficients avoids selective interpretation and better represents uncertainty. Overall, the sequential model comparison reinforces the analytical relevance of separating review content, affective language, and structured rating cues while limiting substantive interpretation to conditional statistical relationships.

4.5. Conditional Associations Across Levels of Star-Rating Extremity

For a negative binomial model containing an interaction between a topic indicator and rating extremity (EXT), the relative expected endorsement rate of domain k compared with T0 depends on the level of EXT. Holding other covariates fixed, the topic-specific conditional incidence-rate ratio can be written as:

$$IRR_k(EXT) = exp(\beta_k + \delta_k \times EXT), \quad k \in \{1,2,3\}$$

For T3, as EXT increases from 0 to 1 and 2, the model-implied ratio relative to T0 rises from approximately 2.39 to 3.59 and 5.40. Ratios for T1 and T2 also increase but less steeply. Figure 6 visualises the conditional ratios computed from the coefficients of Model 3. The key implication is interpretive: a main-effect coefficient in an interaction model does not represent a constant difference at all rating levels. The positive T3 interaction reflects heterogeneity in the modelled conditional association, not proof that consumers cognitively prioritise takeover concerns or that rating extremity causally alters information processing.

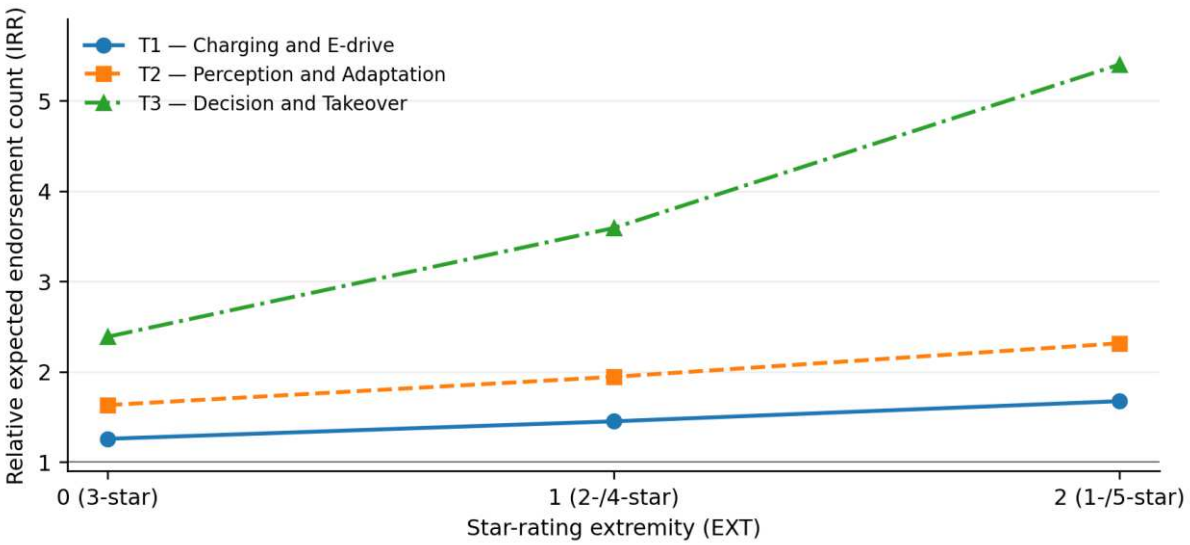


Figure 6. Conditional endorsement ratios across star-rating extremity levels.

The conditional association has theoretical relevance because it is consistent with the possibility that topic-related substantive information and more accessible rating cues coexist in online-review evaluation [8–10,32,33]. However, the observed pattern does not identify whether readers followed a central or peripheral processing route. Establishing that mechanism would require independent designs measuring information exposure, interpretation, and reader response rather than relying solely on aggregated endorsement counts.

4.6. Sensitivity to Extreme Endorsement Observations

A sensitivity analysis considers the influence of unusually high counts on negative binomial parameter estimates. Using the reported 99th-percentile cut-off of 25 endorsements, the restricted sample retains 16,345 records, compared with 16,500 in the complete analytical set. The main coefficient for T3 increases from 0.870 to 0.903, while the T3-by-extremity interaction decreases from 0.408 to 0.303. The corresponding T1 and T2 coefficients remain positive and are comparatively stable. Such results can indicate whether the signs and magnitudes of estimates are particularly sensitive to upper-tail observations; they are not an adequate substitute for robustness to alternative samples, model specifications, or data-collection periods.

Table 10. Sensitivity Analysis After Removal of Upper-Tail Endorsement Counts

Coefficient	Full sample	Restricted sample	Comparison
T1 main contrast	0.229	0.225	Limited change
T2 main contrast	0.489	0.496	Slightly higher
T3 main contrast	0.870	0.903	Higher
T1 × extremity	0.143	0.144	Broadly stable
T2 × extremity	0.175	0.163	Lower
T3 × extremity	0.408	0.303	More substantial reduction
Positive sentiment	0.223	0.218	Slightly lower
Sample size	16,500	16,345	155 observations excluded

Note: Restricted sample retains observations with endorsements ≤ 25 (reported 99th-percentile threshold). Standard errors not provided in the source are not imputed.

This check should be complemented, where the underlying data permit, by alternative sentiment encodings, tests of the zero-count mechanism, observations from different software versions and time windows, vehicle-specific effects, and sensitivity to topic-assignment rules. Zero-inflated or hurdle count models may be warranted if a distinct zero-generating process is supported by diagnostic evidence. The distinction between numerical stability within the current dataset and generalisability to other populations remains central to interpretation.

5. Discussion

Previous research has largely addressed NEV consumers through adoption barriers, range anxiety, charging access, and technology trust. Egbue and Long [1] identified range and infrastructure constraints as barriers to electric-vehicle adoption, while Hardman and colleagues [2] emphasised the importance of consumers' actual interaction with charging facilities. Research on automated driving has also highlighted safety concern, trust calibration, and the difficulty of transferring control appropriately between automation and human drivers [3,16,17,22,26,27]. Building on these perspectives, the present framework shifts attention from pre-purchase acceptance towards evaluations formed in actual use and differentiates four technical domains rather than relying on a single overall satisfaction score. T0 and T1 distinguish range and energy efficiency from the convenience and reliability of replenishment and propulsion; T2 and T3 separate environmental recognition and scenario adaptation from control decisions and driver takeover. This differentiation helps make consumer-perceived value more specific to the technological architecture of NEVs. From the perspectives of TCV and TAM, the contribution lies in identifying the different functional benefits, usability conditions, and perceived risks involved in users' post-adoption judgments. These domains are theoretical and semantic constructs rather than psychometrically validated equivalents of existing TCV or TAM scales.

Studies of online-review helpfulness have shown that review depth, rating extremity, and information diagnosticity can matter for reader evaluations [9,10,32–34]. The present analytical framework extends this discussion by contrasting topic frequency with observable community endorsement and by examining whether rating extremity conditions the association between technical themes and endorsement counts. The supplied descriptive pattern features a relatively high prevalence of range-related content but greater average interactions for decision-making and takeover content; the reported interaction estimates are also positive for T3. The conceptual contribution is the separation of discussion prevalence from content-specific endorsement and the explicit consideration of structured evaluative cues alongside thematic content. From an ELM perspective, this design provides a way to investigate whether substantive technical descriptions and readily observed rating cues operate jointly, although the regression results alone do not identify cognitive processing mechanisms or causal effects. For practice, the framework offers manufacturers a structured means of organising consumer feedback on range loss, charging compatibility, ADAS scenario limits, warning timing, and takeover experiences. Review-derived signals should inform engineering triage and clearer consumer communication but cannot substitute for operating-domain information, vehicle logs, safety testing, or independently verified incident investigations.

6. Conclusion

This study proposes an integrated analytical framework for interpreting user evaluations of electrified powertrains and advanced driver-assistance systems. Addressing **RQ1**, it defines four technology-specific domains: battery range and energy efficiency (T0), charging replenishment and electric-drive reliability (T1), driving-environment perception and scenario adaptation (T2), and driving decision-making and safety takeover (T3). Combining conceptual classification with LDA and BERTopic-based semantic analysis provides a systematic way to distinguish subjects that an overall star rating would otherwise compress. The framework therefore broadens the study of NEV consumer experience beyond general adoption and satisfaction to post-adoption, technology-specific perceptions.

Regarding **RQ2**, the reported topic and sentiment summaries distinguish how often a subject is discussed from how positively it is described and how frequently a comment is endorsed. T0 accounts for 29.10% of assigned topics and has an 83.01% positive-sentiment share, while T3 accounts for 19.62% but has the highest mean endorsement count, 8.753. The negative binomial approach provides a means to examine such associations while accounting for differences in ratings, sentiment, and exposure-related characteristics. Regarding **RQ3**, the positive T3-by-rating-extremity coefficient ($\beta = 0.408$) implies that the modelled relative endorsement rate of takeover-related content rises with distance from the midpoint of the star-rating scale. These findings identify conditional relationships in the analysed records, not an experimentally established rating effect or proof of reader cognition.

Overall, the study brings together consumption-value, technology-acceptance, and information-processing perspectives to connect technical review content, affective expression, and platform endorsement. Its practical

value lies in helping automotive manufacturers distinguish routine energy-use concerns from charging, perception, and safety-takeover issues that warrant targeted product and communication responses. The approach is subject to important limitations arising from review selection, unequal exposure, the possible overlap of technical themes, vehicle configuration differences, and the absence of independent objective safety records. Moreover, the numerical summaries in the supplied source were not accompanied by the original review database and full model execution logs during preparation of this English version; they are therefore translated and retained as reported rather than independently revalidated. Further research should use documented and lawfully obtained review corpora, validate topic and sentiment predictions on independently annotated Chinese text, and incorporate powertrain configuration, software version, driving circumstances, and exposure mechanisms to strengthen external validity and practical interpretation.

Declarations

Author contributions: Conceptualization, X.L. and Q.L.; methodology, X.L.; software, X.L.; validation, X.L. and Q.L.; formal analysis, X.L.; investigation, X.L. and Q.L.; resources, Q.L.; data curation, X.L.; writing−original draft preparation, X.L.; writing−review and editing, X.L. and Q.L.; visualization, X.L.; supervision, Q.L.; project administration, Q.L. Both authors have read and approved the final version of the manuscript.

Data availability: The analytical dataset comprises 16,500 review records concerning BYD Tang New Energy and AITO M9. Autohome provided the reference framework for the review categories and data structure. The dataset and associated analytical materials are available from the corresponding author upon reasonable request.

Funding: This research received no external funding.

Competing interests: The authors declare no competing interests.

Ethics approval and consent to participate: No human participants were recruited, and no personally identifiable information was collected or processed. Informed consent was therefore not applicable. The applicability of institutional ethics approval or exemption remains subject to confirmation under the relevant institutional requirements.

Clinical trial number: not applicable

Consent for publication

Both authors, Li Xinyu and Liu Qijia, have read and approved the final version of the manuscript and have agreed to its submission and publication in Discover Applied Sciences. Both authors consent to the publication of the manuscript and confirm that they have approved its content and authorship information.

References

[1] Egbue, O., & Long, S. (2012). Barriers to widespread adoption of electric vehicles: An analysis of consumer attitudes and perceptions. Energy Policy, 48, 717–729. https://doi.org/10.1016/j.enpol.2012.06.009

[2] Hardman, S., et al. (2018). A review of consumer preferences of and interactions with electric vehicle charging infrastructure. Transportation Research Part D: Transport and Environment, 62, 508–523. https://doi.org/10.1016/j.trd.2018.04.002

[3] Kyriakidis, M., Happee, R., & de Winter, J. C. F. (2015). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. Transportation Research Part F: Traffic Psychology and Behaviour, 32, 127–140. https://doi.org/10.1016/j.trf.2015.04.014

[4] SAE International. (2021). Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles (SAE J3016_202104). https://saemobilus.sae.org/standards/j3016_202104

[5] Sheth, J. N., Newman, B. I., & Gross, B. L. (1991). Why we buy what we buy: A theory of consumption values. Journal of Business Research, 22(2), 159–170. https://doi.org/10.1016/0148-2963(91)90050-8

[6] Sweeney, J. C., & Soutar, G. N. (2001). Consumer perceived value: The development of a multiple item scale. Journal of Retailing, 77(2), 203–220. https://doi.org/10.1016/S0022-4359(01)00041-0

[7] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

[8] Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. Advances in Experimental Social Psychology, 19, 123–205. https://doi.org/10.1016/S0065-2601(08)60214-2

[9] Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. MIS Quarterly, 34(1), 185–200. https://doi.org/10.2307/20721420

[10] Filieri, R. (2015). What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. Journal of Business Research, 68(6), 1261–1270. https://doi.org/10.1016/j.jbusres.2014.11.006

[11] Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv:2203.05794. https://doi.org/10.48550/arXiv.2203.05794

[12] Liu, Y., Ott, M., Goyal, N., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv:1907.11692. https://arxiv.org/abs/1907.11692

[13] Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. Proceedings of EMNLP-IJCNLP, 3982–3992. https://doi.org/10.18653/v1/D19-1410

[14] Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data (2nd ed.). Cambridge University Press.

[15] Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. Transportation Research Part A: Policy and Practice, 77, 167–181. https://doi.org/10.1016/j.tra.2015.04.003

[16] Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

[17] Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics—Part A, 30(3), 286–297. https://doi.org/10.1109/3468.844354

[18] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. Journal of Machine Learning Research, 3, 993–1022.

[19] Funke, S. Á., Sprei, F., Gnann, T., & Plötz, P. (2019). How much charging infrastructure do electric vehicles need? A review of the evidence and international comparison. Transportation Research Part D: Transport and Environment, 77, 224–242. https://doi.org/10.1016/j.trd.2019.10.024

[20] Hao, X., Wang, H., Lin, Z., & Ouyang, M. (2020). Seasonal effects on electric vehicle energy consumption and driving range: A case study on personal, taxi, and ridesharing vehicles. Journal of Cleaner Production, 249, 119403. https://doi.org/10.1016/j.jclepro.2019.119403

[21] Nazari, F., Mohammadian, A. K., & Stephens, T. (2019). Modeling electric vehicle adoption considering a latent travel pattern construct and charging infrastructure. Transportation Research Part D: Transport and Environment, 72, 65–82. https://doi.org/10.1016/j.trd.2019.04.010

[22] Zhang, T., Tao, D., Qu, X., Zhang, X., Lin, R., & Zhang, W. (2019). The roles of initial trust and perceived risk in public's acceptance of automated vehicles. Transportation Research Part C: Emerging Technologies, 98, 207–220. https://doi.org/10.1016/j.trc.2018.11.018

[23] Iogansen, X., Gore, C., Kneifel, J., Ranganath, S., & Helveston, J. P. (2026). Consumers' perceptions, knowledge, and adoption patterns of battery electric vehicles. Transportation Research Part D: Transport and Environment, 105194. https://doi.org/10.1016/j.trd.2025.105194

[24] Ma, X., Xie, R., Guo, L., Niu, S., Cheng, L., & Hu, R. (2025). Range anxiety of battery electric vehicles: Quantification and determinants using real-world data. Transportation Research Part D: Transport and Environment, 146, 104837. https://doi.org/10.1016/j.trd.2025.104837

[25] Ma, X., Li, J., Cui, H., Cheng, L., Ji, Y., & Wang, J. (2025). Electric vehicle range prediction considering real-time driving factors and battery capacity index. Transportation Research Part D: Transport and Environment, 104795. https://doi.org/10.1016/j.trd.2025.104795

[26] Zhang, J., Ma, Y., Xing, G., Xing, Y., Ong, G. P., & Chen, S. (2026). Effect of system operational transparency on driver trust and takeover behavior in a conditional automated driving environment. Transportation Research Part F: Traffic Psychology and Behaviour, 121, 103683. https://doi.org/10.1016/j.trf.2026.103683

[27] Pan, H., Xu, K., Qin, Y., & Wang, Y. (2023). How does drivers' trust in vehicle automation affect non-driving-related task engagement, vigilance, and initiative takeover performance after experiencing system failure? Transportation Research Part F: Traffic Psychology and Behaviour, 98, 73–90. https://doi.org/10.1016/j.trf.2023.09.001

[28] Marcinkiewicz, V., Zhang, Q., Asada, M., et al. (2026). Trust and blame in autonomous vehicles: Examining the effects of cyber readiness and response in the UK and Japan. Transportation Research Part F: Traffic Psychology and Behaviour, 103589. https://doi.org/10.1016/j.trf.2026.103589

[29] Parker, N. C., Kuby, M., Liu, J., et al. (2025). Extreme heat effects on electric vehicle energy consumption and driving range. Applied Energy, 380, 125051. https://doi.org/10.1016/j.apenergy.2024.125051

[30] Seo, J., Vijayagopal, R., Kim, N., Rousseau, A., & Stutenberg, K. (2025). Effects of ambient temperature on electric vehicle range considering battery performance, powertrain efficiency, and HVAC load. Energy Conversion and Management, 326, 119493. https://doi.org/10.1016/j.enconman.2025.119493

[31] Lv, S., Gan, W., Lai, K.-H., & Xu, Z. (2026). The impact of charging station's electronic word-of-mouth on the electric vehicle adoption. Transportation Research Part D: Transport and Environment, 155, 105306. https://doi.org/10.1016/j.trd.2026.105306

[32] Majumder, M. G., Gupta, S. D., & Paul, J. (2022). Perceived usefulness of online customer reviews: A review mining approach using machine learning and exploratory data analysis. Journal of Business Research, 150, 147–164. https://doi.org/10.1016/j.jbusres.2022.06.012

[33] Choi, H. S., & Leon, S. (2020). An empirical investigation of online review helpfulness: A big data perspective. Decision Support Systems, 139, 113403. https://doi.org/10.1016/j.dss.2020.113403

[34] Zheng, L. (2021). The classification of online consumer reviews: A systematic literature review and integrative framework. Journal of Business Research, 135, 226–251. https://doi.org/10.1016/j.jbusres.2021.06.038

[35] Hu, W., Zhang, T., Zhang, Y., & Chan, A. H. S. (2024). Non-driving-related tasks and drivers' takeover time: A meta-analysis. Transportation Research Part F: Traffic Psychology and Behaviour, 623–637. https://doi.org/10.1016/j.trf.2024.05.012

[36] Yi, B., Cao, H., Song, X., Zhao, S., Guo, W., & Li, M. (2022). How to identify the take-over criticality in conditionally automated driving? An examination using drivers' physiological parameters and situational factors. Transportation Research Part F: Traffic Psychology and Behaviour, 85, 161–178. https://doi.org/10.1016/j.trf.2021.12.007

[37] Melliger, M. A., van Vliet, O. P. R., & Liimatainen, H. (2018). Anxiety vs reality—Sufficiency of battery electric vehicle range in Switzerland and Finland. Transportation Research Part D: Transport and Environment, 65, 101–115. https://doi.org/10.1016/j.trd.2018.08.011

[38] Zhang, B., Niu, N., Li, H., Wang, Z., & He, W. (2021). Could fast battery charging effectively mitigate range anxiety in electric vehicle usage? Evidence from large-scale data on travel and charging in Beijing. Transportation Research Part D: Transport and Environment, 95, 102840. https://doi.org/10.1016/j.trd.2021.102840

[39] Kumar, V., Kaushik, A. K., Noravesh, F., Sindhwani, R., & Mathiyazhagan, K. (2025). Green drives: Understanding how environmental propensity, range and technological anxiety shape electric vehicle adoption intentions. Technological Forecasting and Social Change, 210, 123859. https://doi.org/10.1016/j.techfore.2024.123859

[40] Lin, B., & Yang, M. (2024). Changes in consumer satisfaction with electric vehicle charging infrastructure: Evidence from two cross-sectional surveys in 2019 and 2023. Energy Policy, 185, 113924. https://doi.org/10.1016/j.enpol.2023.113924

[41] Featherman, M., Jia, S., Califf, C. B., & Hajli, N. (2021). The impact of new technologies on consumers beliefs: Reducing the perceived risks of electric vehicle adoption. Technological Forecasting and Social Change, 169, 120847. https://doi.org/10.1016/j.techfore.2021.120847

[42] Junquera, B., Moreno, B., & Álvarez, R. (2016). Analyzing consumer attitudes towards electric vehicle purchasing intentions in Spain: Technological limitations and vehicle confidence. Technological Forecasting and Social Change, 109, 6–14. https://doi.org/10.1016/j.techfore.2016.05.006

[43] Adu-Gyamfi, G., Asamoah, A. N., Obuobi, B., Nketiah, E., & Zhang, M. (2024). Electric mobility in an oil-producing developing nation: Empirical assessment of electric vehicle adoption. Technological Forecasting and Social Change, 123173. https://doi.org/10.1016/j.techfore.2023.123173

[44] Cui, D., Liu, P., & Wang, Z. (2025). Assess changes in electric vehicle usage behaviour: Comparison between 2018 and 2021. Transportation Research Part D: Transport and Environment, 141, 104661. https://doi.org/10.1016/j.trd.2025.104661

[45] Kolarova, V., & Cherchi, E. (2021). Impact of trust and travel experiences on the value of travel time savings for autonomous driving. Transportation Research Part C: Emerging Technologies, 131, 103354. https://doi.org/10.1016/j.trc.2021.103354

[46] Xu, Z., Zhang, K., Min, H., Wang, Z., Zhao, X., & Liu, P. (2018). What drives people to accept automated vehicles? Findings from a field experiment. Transportation Research Part C: Emerging Technologies, 95, 320–334. https://doi.org/10.1016/j.trc.2018.07.024

[47] Singh, R., Quinn, C., & MacKenzie, D. (2025). Poor reliability of public charging stations can impede the growth of the electric vehicle market. Transport Policy. https://doi.org/10.1016/j.tranpol.2025.06.026

[48] Jaiswal, D., Kaushal, V., Kant, R., & Singh, P. K. (2021). Consumer adoption intention for electric vehicles: Insights and evidence from Indian sustainable transportation. Technological Forecasting and Social Change, 173, 121089. https://doi.org/10.1016/j.techfore.2021.121089

[49] Paddeu, D., Parkhurst, G., & Shergold, I. (2020). Passenger comfort and trust on first-time use of a shared autonomous shuttle vehicle. Transportation Research Part C: Emerging Technologies, 102604. https://doi.org/10.1016/j.trc.2020.02.026

[50] Elayan, M., & Kontar, W. (2026). Can automated vehicles have it all? A consensus framework for diagnosing behavioral trade-offs in mixed traffic. Transportation Research Part C: Emerging Technologies, 190, 105782. https://doi.org/10.1016/j.trc.2026.105782

[51] Wang, A., Wang, Z., Yan, S., He, D., & Yang, H. (2026). Enhancing novice drivers' takeover behavior in conditional automated driving: A KAP-based training approach. Transportation Research Part C: Emerging Technologies, 191, 105828. https://doi.org/10.1016/j.trc.2026.105828

[52] Du, N., Zhou, F., Pulver, E. M., Tilbury, D. M., Robert, L. P., Pradhan, A. K., & Yang, X. J. (2020). Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving. Transportation Research Part C: Emerging Technologies, 112, 78–87. https://doi.org/10.1016/j.trc.2020.01.006

[53] Müller, W., Leyer, M., Degirmenci, K., & Barros, A. (2025). Designing EV charging networks from a city perspective: A conceptual analysis. Energy Policy, 198, 114516. https://doi.org/10.1016/j.enpol.2025.114516

[54] Baumgarte, F., Kaiser, M., & Keller, R. (2021). Policy support measures for widespread expansion of fast charging infrastructure for electric vehicles. Energy Policy, 156, 112372. https://doi.org/10.1016/j.enpol.2021.112372

[55] Anselma, P. G., Kollmeyer, P., Lempert, J., Zhao, Z., Belingardi, G., & Emadi, A. (2021). Battery state-of-health sensitive energy management of hybrid electric vehicles: Lifetime prediction and ageing experimental validation. Applied Energy, 285, 116440. https://doi.org/10.1016/j.apenergy.2021.116440

[56] Anandkumar, R., Sathiyaseelan, V., Sindhu Priya, G., et al. (2026). A comprehensive review on electric vehicle battery technologies, thermal management systems, and emerging nanofluid applications. Discover Applied Sciences, 8, 943. https://doi.org/10.1007/s42452-026-08963-w

[57] Bhatti, A. R., Tamoor, M., Liaqat, R., Rasool, A., Salam, Z., Ali, A., et al. (2024). Electric vehicle charging stations and the employed energy management schemes: A classification based comparative survey. Discover Applied Sciences, 6, 503. https://doi.org/10.1007/s42452-024-06190-9

[58] Ma, J., & Li, Y. (2026). Environmental anomaly detection and intelligent response for autonomous driving based on deep neural networks. Discover Applied Sciences, 8, 960. https://doi.org/10.1007/s42452-026-08919-0

[59] Gunasekaran, R., Manjunatha, B., Anand, S., Pareek, P. K., Gupta, S., et al. (2024). Prediction of electric vehicle charging demand using enhanced gated recurrent units with RKOA based graph convolutional network. Discover Applied Sciences, 6, 605. https://doi.org/10.1007/s42452-024-06326-x

[60] Louis, G. A., & Sampathkumar, S. (2025). Predicting the state of charge of lithium ion battery in e-vehicles using Box-Jenkins combined artificial neural network model. Discover Applied Sciences, 7, 129. https://doi.org/10.1007/s42452-024-06445-5

[61] Pande, A. S., Sharma, A. K., & Soni, B. P. (2025). Deep learning based state of charge estimation of lithium ion battery for electric vehicle application. Discover Internet of Things, 5, 150. https://doi.org/10.1007/s43926-025-00148-y

[62] Han, W., & Wang, D. (2025). New energy vehicle fast charging reservation algorithm based on internet of things coordination. Discover Internet of Things, 5, 124. https://doi.org/10.1007/s43926-025-00212-7

[63] Das, S. R., Mohapatra, S. K., Kamila, N. K., et al. (2025). Smart urban transportation: Advancing IoT-based thermal management for lithium-ion EV battery packs. Discover Internet of Things, 5, 20. https://doi.org/10.1007/s43926-025-00121-9

[64] Bao, L. (2026). Distributed imitation learning and adversarial path tracking strategies for intelligent connected vehicles oriented to dynamic interaction scenarios. Discover Internet of Things, 6, 223. https://doi.org/10.1007/s43926-026-00532-2

[65] Pal, S., Khalifa, S., Miller, D., et al. (2024). Uncertainty propagation in the internet of things. Discover Internet of Things, 4, 32. https://doi.org/10.1007/s43926-024-00085-2