Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 1458–1468
https://doi.org/10.54941/ahfe1008374

Qualifying NEV User Experience: LLM-Based Dimensional Encoding and Hierarchical Clustering of Online Comments

Siyun Zhang

Department of AI & Data Science, School of Computing and Data Science, The University of Hong Kong

ABSTRACT

As competition in the new-energy-vehicle (NEV) market increasingly shifts from technical specifications to holistic user experience, understanding user-generated feedback has become essential for product optimisation. This report investigates NEV user experience across 12 dimensions, including powertrain performance, handling, space and comfort, intelligent driving, smart cabin functions, charging convenience, and after-sales service, with the aim of identifying distinct user segments from online reviews. Based on a sample of 500 user reviews, the study applies a text-mining pipeline that combines rule-based term standardisation, LLM-based summarisation and classification, semantic encoding with Sentence-BERT, and hierarchical clustering. This workflow converts large-scale, unstructured comments into structured experience features. It also enables the comparison of users according to the similarity of their expressed concerns and evaluations. The analysis identifies five user clusters with clearly differentiated priorities: smart-cabin enthusiasts, comfort-driven users, pragmatic technology users, range-anxious drivers, a space-oriented majority. Research indicates significant diversity among NEV consumers regarding their priorities – in more details, that is intelligent capabilities, seat comfort, driving range, charging accessibility, and powertrain dependability. These results provide data-driven support for targeted product development, user-segment-specific marketing, and strategic expectation management, while also demonstrating the value of text mining for extracting actionable insights from NEV user reviews.

Keywords: LLM, User-experience, Text-mining, Semantic-vector

INTRODUCTION

With the rapid progress of global energy transformation together with "Carbon Peaking & Carbon Neutrality" goals, NEV industry has grown significantly. The focus of market competition has gradually transformed, from earlier technology parameters as well as driving range, to user-centered comprehensive experience competition. Under this background, deeper understanding of real and fine-grained experience together with demand of users plays a significant role in designing optimised products of vehicle enterprises, enhancing service quality and establishing brand loyalty.

Meanwhile, internet and social media has induced vast number of user-generated contents (UGC) – personal feelings, ratings and pains that users

Received August 17, 2026; Revised September 22, 2026; Accepted September 27, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

shared on social media platforms – which will be scrapped to dataset. These authentic raw datasets involve users' feedbacks which are hard to obtain using traditional research methods and valuable sources for insight into user experience. However, these raw datasets are normally unstructured and high-dimensional, so they need to be handled in some technological methods effectively and discover valuable structured information accurately – which question has become a core challenge for current study and practice.

This report will discover the core demand insights from 500 of the total review from the NEV users, including clustering of users and demand; and identify different user groups together with what they focus on.

LITERATURE REVIEW

The Concept of "User's Experience" to the Literature Review

As defined in ISO 9241-210 framework, user experience (UX) is defined as the perceptions and reactions during and after using products, systems or services. According to Kano et al. (1984), user satisfaction is not a monotonic linear dimension, and product attributes are categorised into three types: must-be, whose absence may cause strong dissatisfactions; one-dimensional, whose fulfilment is linearly related to satisfaction; attractive, whose presence may bring disproportionate delight.

Therefore, the user group insights play a vital role in product definition and design. On the one hand, such insights characterize the scale, composition, and core needs of mainstream user segments, enabling companies to accurately capture the overall profile of the target market. On the other hand, focusing on a smaller set of lead users is equally strategic – these users tend to exhibit higher technological sensitivity and more forward-looking preferences; their experiential feedback and latent needs not only reveal potential product evolution trajectories but may also signal early indicators of industry trends. This systematic understanding of diverse user groups renders user insights an indispensable component of user research. This is particularly true in the new energy vehicle sector, where user demands are increasingly heterogeneous and personalized: data-driven group insights provide empirical support for differentiated product positioning, the prioritisation of features, and the formulation of market segmentation strategies.

Based on the above UX theory, combined with industry reports, a 12-dimensional UX structure is proposed, which are detailed in the table below:

Table 1: 12-dimension table based on the UX theorem.

Dimension	Topic	Details
1	Powertrain System	driving performance and energy efficiency
2	Handling & Control	vehicle handling and driving feel
3	Space & Comfort	indicators like rear seat space, legroom, trunk capacity, ride comfort
4	Intelligent Driving	ADAS and advanced driver assistance features

(Continued)

Table 1: Continued.

Dimension	Topic	Details
5	Smart Cabin	interior intelligent interactive system
6	Interior Design	in-cabin environment and comfort
7	Exterior Styling	visual aesthetics and personalisation
8	Safety Performance	active and passive safety configurations
9	Charging Convenience	charging convenience and accessibility
10	After-sales Service	after-sales experience like service-response time, repair cost, warranty policy, remote diagnostics
11	Smart Cabin	interior intelligent interactive systems: voice recognition; central display...
12	NVH Performance	wind noise, tire noise, motor noise, sound insulation materials

All 12 aspects may affect the sales of the NEVs, and the producers may change its functions of NEVs based on the users' experience to meet the consumers' needs. Hence, transformers library is necessary for us to do "information extraction", "text classification and summarisation" on the raw dataset according to the 12 dimensions. This study also design a structured output format to ask the LLM to categorise the summarised sentences so that the results that obtained from the code are obvious for us to make conclusions.

The Large Language Models (LLM)

In the contemporary landscape of product competition, user experience (UX) has become increasingly significant, particularly within the NEV sector. Plenty of such analyses have been conducted, then this methodology is implemented through data mining including data cleansing and vector standardisation, calling large language models (LLM), clustering analysis and finally data visualisation.

The large language model that this study mainly use is Qwen3.6-Plus, and the embedding model that this study use is bge-large-zh-v1.5. This study chooses Qwen3.6-Plus as an LLM model due to its leading original Chinese model, and its outstanding semantic understandability. This study considers bge-large-zh-v1.5 as a suitable model because of its capability of understanding Chinese, pursue extreme precision, and 1024-dimensional (i.e. high dimensional).

The Architecture of Transformers

The *transformers* model is originally developed for NLP tasks, the *transformer* model has become the dominant architecture in the field – surpassing earlier neural networks like CNNs and RNNs – due to its capability of capturing long-range dependencies, supporting efficient parallel training, scaling effectively with both data and model size. These advantages provide significant supports for the information extracting and text analysis within

the project. This study adopts an open-source *transformers* library and offers a unified and user-friendly port, which enables people to apply *transformers* framework widely (Wolf et al., 2020).

METHODOLOGIES

Research Framework Overview

The process of this project is mainly based on the text-mining technique, consisting of several basic steps as well as some time-consuming computations (as shown in Fig. 1 below). To save some time, this study only selects a smaller sample of the processed data due to the large running time complexity.

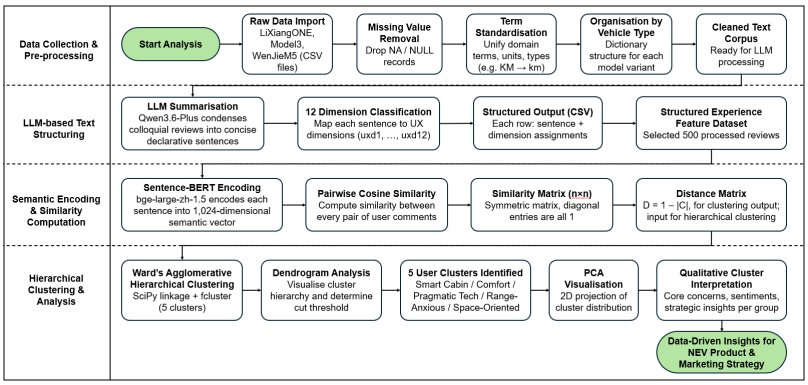


Figure1: User experience analysis model based on text-mining.

Data Acquisition

The dataset used in this study consists of user-generated product reviews collected from Chinese online automotive forums and community platforms. User reviews of three representative NEV models were acquired, each stored as a separate comma-separated values (CSV) file:

- Li Xiang ONE: a domestically produced range-extender SUV covering model years 2020 and 2021, with columns carType and userComment;
- Tesla Model 3: an imported pure-electric sedan covering model years 2019 through 2022, with columns carType and userComment;
- Wenjie M5: a Huawei-backed range-extender and pure-electric SUV covering model year 2022, with columns date, carType, num, and userComment.

Each record in these files contains the specific vehicle variant (carType) and a free-text user review (userComment) written in colloquial Chinese. The Wenjie M5 dataset additionally records the posting date and a review identifier (num). These three models were selected because they represent distinct market segments – a domestic range-extender large SUV, an imported pure-electric sedan, and a technology-branded smart SUV – thereby providing a reasonably diverse cross-section of the Chinese NEV market.

The raw reviews vary substantially in length and style, ranging from concise single-paragraph opinions to multi-section narratives exceeding 1,000 characters, and they cover a broad spectrum of experiential topics, from powertrain performance and driving comfort to intelligent cockpit features and after-sales service. This heterogeneity in both linguistic form and content richness makes the dataset well-suited for LLM-based text structuring and subsequent clustering analysis.

Data Preprocessing

The raw data consists of the user comments of three types of cars, which are `lixiang1` made in the year 2020 and 2021; `model3` made in the year from 2019 to 2022; and `wenjie` made in the year 2022. Each of them has attributes (generally user comments and car type) so that semantic analyses can be conducted more conveniently.

To prepare these raw texts for semantic analysis, the data with the missing values such as NA and NULL is first removed and then create a new dataset by utilising the `torch.utils.data` and `dict` function. The example code is shown as follows:

```
data_lx_1_dict = {}
data_lx_1_dict['2021-LiXiangONE-ExtendRange-6-Seat'] = []
data_lx_1_dict['2020-LiXiangONE-ExtendRange-7-Seat'] = []
data_lx_1_dict['2020-LiXiangONE-ExtendRange-6-Seat'] = []
for i in range(len(data_lx_1)):
    data_lx_1_dict[data_lx_1[i][1]].append(data_lx_1[i][0])

print(data_lx_1_dict)
```

A multi-step preprocessing pipeline is afterwards conducted by using the Qwen3.6-Plus natural language processing model. This involves two main stages: rule-based term standardisation and LLM-based content structuring.

Stage 1: Term Standardisation

Term standardisation refers to a segmentation of the acquired user review texts into individual sentences, thereby enabling sentence-level analysis of user comments.

To ensure consistency across the dataset and reduce noise for downstream analysis, a rigorous rule-based cleaning process is applied. This step unifies domain-specific terminologies, corrects common typos, and standardizes unit representations (e.g., unifying "KM" to "km", "KW" to "kW"). Additionally, variations in car model names and component descriptions were normalized. For instance, the Chinese term such as "增程器" (range extender) was standardised to "发动机" (engine) to maintain semantic consistency.

The standardisation process is implemented as follows:

```
  for brand, comment in data_wenjie_dict.items():
    length = len(comment)
```

```
for i in range(0, length - 1):
    # Standardize component names
    comment[i] = comment[i].replace("增程器", "发动机")
    ...
    comment[i] = comment[i].replace("KM", "km")
    ...
    # Unify terminology
    comment[i] = comment[i].replace("超充", "超级充电")
```

Stage 2: Term Standardisation

After standardisation, the raw user comments often remain colloquial and unstructured. To extract meaningful features for clustering, this study utilised Large Language Models (LLMs) such as Qwen3.6-Plus to process the text.

Therefore, this study designed a two-step prompt strategy: summarisation and classification. In terms of summarisation, the model first distills perplex user reviews into a concise and declarative single sentence, where the redundant information besides the core user feedback is removed. Regarding to classification, the summarised sentences are then mapped to the 12 pre-defined user experience dimensions (e.g., Charging Convenience, Intelligent Driving, etc).

We can discover that this systematic process has converted unstructured textual data into a comprehensive, attribute-rich dataset. In this format, individual user comments are both classified and measured, establishing a solid groundwork for the upcoming cluster analysis.

Text Encoding & Similarity Modelling

To uncover distinct user segments and their characteristic concerns, this study applied unsupervised representation learning followed by hierarchical clustering. Specifically, our approach consists of the following steps with first reducing the dimensionality of the SBERT embeddings: text encoding, similarity calculation, hierarchical clustering, cluster identification and visualisation.

We utilise the `sentence-transformers` library with the pre-trained `bge-large-zh-v1.5` model to encode each user comment into a high-dimensional semantic vector (Reimers & Gurevych, 2021). This model is not only optimised for generating meaningful sentence embeddings that capture the semantic nuances of the text, but also utilised to calculate the cosine similarities of the sentences. For a pair of sentence embeddings $\vec{u}, \vec{v} \in R^d$, their semantic similarity is measured by the cosine similarity:

$$sim(\vec{u} \cdot \vec{v}) = cos\theta = \frac{(\vec{u} \cdot \vec{v})}{\|\vec{u}\|\|\vec{v}\|} = \frac{\sum_{k=1}^{d} u_k v_k}{\sqrt{\sum_{k=1}^{d} u_k^2}\sqrt{\sum_{k=1}^{d} v_k^2}}$$

For every processed comment i, its 12-dimensional attribution vector is obtained. For each i, it satisfies that $\vec{a}_i = (a_{i1}, a_{i2}, \ldots, a_{i12}) \in [0, 1]^{12}$, whose entries quantify the relevance of the comment to each of the 12 dimensions.

A similarity matrix A with size $n \times n$ is then constructed, where $A_{ij} = \frac{1}{\sqrt{12}} \left| \vec{a}_i \cdot \vec{a}_j \right|$ and $A_{ii} = 1$. All diagonal entries are 1 and the missing values are set to be 0. Besides, the normalisation factor $\frac{1}{\sqrt{12}}$ ensures that $A_{ij} = 1$ when corresponding components of the two vectors are identical.

Hierarchical Clustering

Hierarchical clustering is used to group users based on semantic similarities, identifying differences in their demands. By analysing correlation matrices and generating user profiles, this method categorises samples into distinct groups, allowing for the summarisation of common characteristics within each group.

Selection of the Clustering Algorithm

Since the number of user groups in the raw data is not known, an algorithm requiring a specific k-value (i.e. K-means) is unsuitable. Comparisons are made between the three families of methods:

- K-means: requiring the number of clusters, k, in advance, which is sensitive to cluster shape as well as initial centroids, and providing no hierarchical relationship
- DBSCAN: depending on density parameters (ε, minPts) that are hard to tune in high-dimensional, sparse semantic space
- Hierarchical clustering: does not require a pre-specified k, exposes the merging process through a dendrogram, and yields a consistent view across different granularities

In this case, this study adopts agglomerative hierarchical clustering considering unknown group structure and the desire for interpretable hierarchy. Ward's method is chosen for it merges clusters to minimise the increments of within-cluster variance, producing compact, roughly spherical, and balanced clusters – contrasting with single linkage that tends to produce chained clusters and complete linkage, sensitive to outliers.

Agglomerative Hierarchical Clustering With Ward's Linkage

This study employed agglomerative hierarchical clustering using Ward's linkage method (Ward, 1963), implemented via SciPy (Virtanen et al., 2020). The algorithm proceeds as follows:

Step1. Initialisation: Treat each of the n comments as a singleton cluster.
Step2. Cluster merging: At each step, compute the merging cost for every pair of clusters, and merge the pair that incurs the smallest increment of within-cluster variance.
Step3. Repeat: Continue merging until all the samples are grouped into a single cluster, recording each merge to build the dendrogram.

We then compute the sum of square error (SSE) within each cluster and then minimise this value.

Determining the Number of Clusters

The similarity matrix was converted to a distance matrix using $D = 1 - |C|$, a standard approach in cluster analysis (Johnson & Wichern, 2007). The following code interprets how it works:

```
df = pd.read_csv('similarity_matrix/sm_mx_wenjie_norm.csv',
index_col=0)
corr_matrix = df.values
distance_matrix = 1 - np.abs(corr_matrix)
```

The data is segmented into five distinct clusters based on the dendrogram using the `fcluster` function (As shown in Fig. 2), and these clusters represent different user groups with shared experiences and concerns. Based on the observation of dendrogram and alignment of S/CH/DB (i.e. Silhouette Score/Calinski–Harabasz Index/Davies–Bouldin Index) which are optimized at the same time, the Bootstrap stability of 0.99, as well as the explainability of five clusters, the value of k is ultimately set to be 5; the computation process is omitted due to the restricted length of the article. Afterwards, it is followed by summarising the characteristics of each of the five clusters.

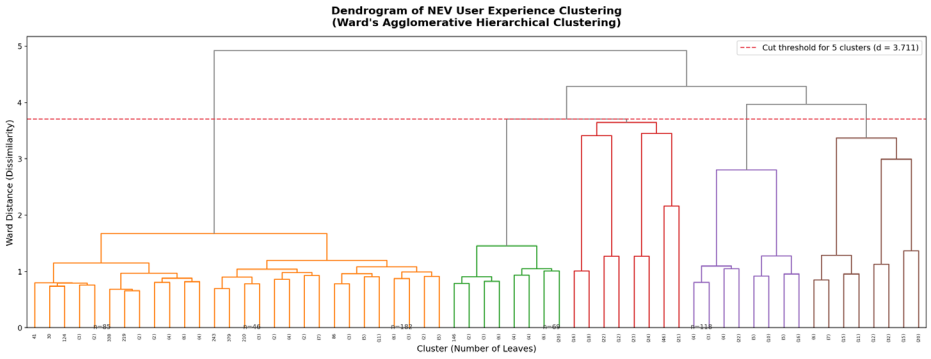


Figure 2: Dendrogram of NEV user experience clustering.

ANALYSIS & DISCUSSIONS

Followed by the graph illustrated and interpreted in the Section 3.2.2 using the Based on the hierarchical clustering results, the 500 user reviews were segmented into five distinct clusters, each reflecting a specific pattern of user experience and evaluative focus.

Table 2: Five user segments (clusters) derived from hierarchical clustering.

Cluster	Size	Key Characteristics	Strategic Implication
Cluster 1: Space-oriented Majority	221 (44.2%)	largest segment; highly satisfied with front-row space; powertrain feedback is largely absent, implying reliable but under-recognised performance	maintain spatial advantages and communicate powertrain reliability as a clearer selling point
Cluster 2: Smart Cabin Enthusiasts	90 (18.0%)	strong appreciation of smart cabin features, especially voice assistant accuracy and infotainment responsiveness	expand the smart cabin ecosystem to strengthen user engagement and competitive differentiation
Cluster 3: Comfort-driven Users	80 (16.0%)	seat comfort is the main concern; limited discussion of interior aesthetics suggests functional comfort outweighs visual design	prioritize seat ergonomics and use design-oriented communication to raise awareness of interior value
Cluster 4: Pragmatic Tech Users	57 (11.4%)	values range-extender powertrain for policy benefits and reduced range anxiety; intelligent features are acceptable but not outstanding	preserve powertrain reliability while improving the experience of intelligent driving and cabin functions
Cluster 5: Range-anxious Drivers	52 (10.4%)	Concerned about insufficient real-world EV range, high-speed and high-temperature energy consumption, and charging inconvenience	Improve thermal management, expand charging infrastructure, and provide more transparent range communication

Overall Strategic Insights

The clustering results demonstrate that modern NEV consumers are no longer a homogeneous group. Their demands are highly fragmented, ranging from foundational needs like space and comfort to advanced requirements like smart cabin fluency and accurate range estimation. A one-size-fits-all product strategy is no longer viable; instead, automakers must adopt targeted optimizations for different user profiles (Johnson & Wichern, 2007). From the graph in the Section 3.2.2, it is obvious that the samples are evenly distributed in each of the five clusters, since colours are used as discriminators of each of the five clusters.

CONCLUSIONS & RECOMMENDATIONS

Text mining and unsupervised hierarchical clustering are applied to analyse NEV user experiences. User reviews were encoded into semantic vectors using sentence-transformers and grouped based on cosine similarity, enabling subjective feedback to be transformed into structured and actionable insights.

Key Recommendations for NEV Development

- **Strengthen smart cabin advantages:** Continue improving voice assistants and system responsiveness, as these features have become key differentiators.
- **Address range and charging limitations:** Enhance battery thermal management to stabilize energy consumption under extreme conditions and accelerate charging infrastructure deployment.
- **Improve core comfort attributes:** Optimize seat ergonomics and spatial design, which remain decisive for the largest user segment.
- **Communicate "silent advantages":** Increase user awareness of reliable but under-recognized features, such as powertrain stability.
- **Manage user expectations:** Align marketing claims, especially regarding EV range, with real-world performance to reduce post-purchase dissatisfaction.

Limitations and Future Work

Being restricted by the computing power, this essay only analyses and probes the feasibilities of the proposed methodology based on the small dataset. The discovery of the diversity structure of real-world user experience must also be based on UGC data at a larger scale, and the processing of static text data in this study makes it impossible for us to capture the changes in user emotions over time. Future research should expand the dataset, incorporate time-series analysis, and develop a more domain-specific framework for the evolving NEV industry.

ACKNOWLEDGMENT

I would like to acknowledge the Design Management Institute of Shanghai Jiao Tong University for providing basic UX datasets.

REFERENCES

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

Gómez Díaz, K. Y., De León Aldaco, S. E., Aguayo Alquicira, J., Ponce Silva, M., Portillo Contreras, S., & Sánchez Vargas, O. (2025). Thermal Management Systems for Lithium-Ion Batteries for Electric Vehicles: A Review. World Electric Vehicle Journal, 16(7), 346. https://doi.org/10.3390/wevj16070346

Johnson, R. A., & Wichern, D. W. (2007). Applied multivariate statistical analysis (6th ed.). Pearson Prentice Hall.

Nayak, G. H. H., Alam, M. W., Avinash, G., Kumar, R. R., Ray, M., Barman, S., Singh, K. N., Naik, B. S., Alam, N. M., Pal, P., Rathod, S., & Bisen, J. (2024). Transformer-based deep learning architecture for time series forecasting. Software Impacts, 22, Article 100716. https://doi.org/10.1016/j.simpa.2024.100716

Reimers, N., & Gurevych, I. (2021). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 398–409. https://doi.org/10.18653/v1/2021.emnlp-main.39

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D. & SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods, 17(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2020, October). Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations (pp. 38–45). https://doi.org/10.18653/v1/2020.emnlp-demos.6