



Mapping consumer voice into engineering insight: a structured language model-driven design support framework for electric vehicles

Qingyang Jin, Luyao Wang, Wenyu Yuan & Danni Chang

To cite this article: Qingyang Jin, Luyao Wang, Wenyu Yuan & Danni Chang (09 Mar 2026): Mapping consumer voice into engineering insight: a structured language model-driven design support framework for electric vehicles, Journal of Engineering Design, DOI: [10.1080/09544828.2026.2639933](https://doi.org/10.1080/09544828.2026.2639933)

To link to this article: <https://doi.org/10.1080/09544828.2026.2639933>



Published online: 09 Mar 2026.



Submit your article to this journal [↗](#)



Article views: 142



View related articles [↗](#)



View Crossmark data [↗](#)



Mapping consumer voice into engineering insight: a structured language model-driven design support framework for electric vehicles

Qingyang Jin^{a*}, Luyao Wang^{a*}, Wenyu Yuan^b and Danni Chang ^a

^aSchool of Design, Shanghai Jiao Tong University, Shanghai, People's Republic of China; ^bSchool of Digital Technology & Innovation Design, Jiangnan University, Wuxi, People's Republic of China

ABSTRACT

Consumer-generated narratives contain rich experiential knowledge for product design, yet their perceptual and weakly structured nature limits direct translation into actionable engineering insight. This challenge is particularly evident in EV design, where user insight plays a critical role but remains weakly grounded in formal engineering representations. To address this gap, this study proposes a Structured consumer semantic insight small language model (SCSI-SLM). First, Structured Semantic Encoding (SSE) maps weakly structured user narratives into a multi-level engineering semantic space, transforming perceptual expressions into comparable and reusable semantic tokens. Second, Product-User Dynamic Mapping (PUDM) decouples product-centric performance assessment from user-centric preference representation, enabling importance-performance analysis and preference heterogeneity to be modelled from the same corpus. Third, an Engineering Design Knowledge Graph (EDKG) integrates user profiles, product attributes, perceptual features, and textual evidence into an interpretable reasoning space. Building on this representation, a hybrid retrieval-reasoning engine combines semantic similarity search with graph-constrained inference to generate evidence-backed and engineer-readable insight chains. A case study in the EV domain demonstrates that the proposed framework can generate traceable design insights from large-scale consumer feedback, suggesting structured language modelling as a viable paradigm for data-driven engineering design.

ARTICLE HISTORY

Received 25 December 2025
Accepted 27 February 2026

KEYWORDS

Data-driven engineering design; structured language modeling; engineering design knowledge graph; consumer feedback; electric vehicles

1. Introduction

Consumer-generated narratives have become an increasingly important source of information in product design and development (Wang et al. 2023; Zhang et al. 2023b). Online reviews and user feedback capture how products are perceived, experienced, and evaluated in real-world usage contexts, often revealing experiential issues and user preferences

CONTACT Wenyu Yuan  yuanwenyu@jiangnan.edu.cn  School of Digital Technology & Innovation Design, Jiangnan University, No. 18000, Lihu Avenue, Binhu District, Wuxi 214122, People's Republic of China; Danni Chang  dchang1@sjtu.edu.cn  School of Design, Shanghai Jiao Tong University, No. 800, Dongchuan Road, Minhang District, Shanghai 200240, People's Republic of China

*co-first authors and contributed equally to this work.

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

that are difficult to obtain through formal testing or specification documents (He, Xue, and Luo 2025; Li et al. 2023; Ma et al. 2019; Zhao, Gao, and Liu 2023). In domains such as EV design, where user experience plays a decisive role, experiential factors including comfort, handling, human–machine interaction, and perceived quality strongly influence product acceptance and market success (Gomez et al. 2024; Liao, Molin, and Van Wee 2017; Liu et al. 2023; Xu et al. 2024).

Despite their potential value, these narratives remain difficult to integrate systematically into engineering design processes. User feedback is typically expressed in natural language that is subjective, metaphorical, and weakly structured (Delle Monache, Misdariis, and Özcan 2022; Wang, Li, and Mo 2021). Such expressions often lack stable dimensionality and a direct mapping to specific engineering parameters, rendering systematic cross-product comparison and quantitative analysis exceptionally challenging (Ettrich et al. 2024; Wang, Luo, and Liu 2020). On the other hand, engineering design relies on structured representations, such as functional models and knowledge graphs, that support precise reasoning but are often too rigid to ingest the perceptual and ambiguous nature of human language (Chen and Luo 2019; Huang et al. 2023; Ji et al. 2021). Recent breakthroughs in large language models (LLMs) have significantly enhanced the capacity to analyse and synthesise vast textual corpora (Achiam et al. 2023; Brown et al. 2020; Nguyen-Duc et al. 2025; Qiu and Jin 2024). However, many existing approaches treat user narratives as unstructured inputs to predictive models, offering limited grounding in engineering semantics and insufficient interpretability for critical design decision-making (Doshi-Velez and Kim 2017; Nikolic et al. 2023; Ribeiro, Singh, and Guestrin 2016). Although general artificial intelligence is pre-trained on large-scale unstructured data, it remains limited in its ability to deeply mine domain-specific knowledge in areas such as user research and engineering design. This limitation is manifested in its insufficient sensitivity to user characteristics, incomplete modelling of domain knowledge, and the absence of a robust framework to ensure the reliability and validity of large language models (LLMs) in extracting user insights.

To bridge this gap, what is required is not merely more data, but a structured mechanism capable of transforming fragmented consumer language into explainable engineering knowledge. Specifically, such a mechanism must satisfy three core requirements: First, **semantic alignment**: perceptual expressions must be mapped onto a stable engineering semantic space to enable aggregation and cross-product comparison (Demoly, Kim, and Horváth 2019; Han et al. 2021). Second, **decoupling performance and preference**: product-level assessments should be separated from user-level preference representations to preserve trade-offs and model heterogeneity (Sun et al. 2023; Wang et al. 2025; Yang et al. 2024). Third, **evidence traceability**: inferred insights must remain linked to original textual evidence to ensure transparency and accountability (Gao et al. 2024; Li et al. 2025; Mustapha 2025).

The objective of this research is to establish an explainable and scalable framework that enables consumer-generated narratives to function as a first-class source of engineering design intelligence. Focusing on the EV domain as a representative case, this study demonstrates how weakly structured experiential language can be systematically transformed into actionable and traceable design insight. The main contributions of this work are as follows:

- (1) **Structured consumer semantic insight small language model (SCSI-SLM)**. We propose SCSI-SLM, a coherent design intelligence framework that transforms

weakly structured consumer narratives into grounded, comparable, and traceable engineering representations through structured semantic encoding, product-user dynamic mapping, and knowledge-grounded reasoning.

- (2) **Empirical validation in the EV domain.** A real-world case study in the EV domain demonstrates the feasibility and practical value of the proposed framework, showing that SCSI-SLM can generate evidence-backed and explainable design insights from large-scale consumer feedback.
- (3) **Design-oriented perspective on language models.** This work offers a design-oriented perspective that repositions language models as structured semantic mediators embedded within engineering knowledge representations, rather than end-to-end decision makers for design tasks.

The remainder of this paper is organised as follows. Section 2 reviews related work on consumer voice analysis and engineering knowledge representation. Section 3 presents the proposed SCSI-SLM framework, detailing the three core steps of structured semantic encoding, product-user dynamic mapping, and knowledge-grounded reasoning. Section 4 reports a case study in the EV domain to validate the proposed approach. Section 5 discusses implications and limitations, and Section 6 concludes the paper.

2. Related work

2.1. Structured semantic encoding of consumer voice for design

Consumer-generated narratives have been widely recognised as a pivotal source of experiential knowledge in product development. Traditional methodologies primarily utilised sentiment analysis, keyword extraction, and topic modelling to aggregate user opinions from massive textual corpora (Chang et al. 2025; He, Xue, and Luo 2025; Joung and Kim 2021; Li et al. 2023). With the recent paradigm shift towards Large Language Models (LLMs), research has demonstrated remarkable capabilities in the semantic understanding, summarisation, and high-level abstraction of unstructured data (Achiam et al. 2023; Brown et al. 2020; Nguyen-Duc et al. 2025; Qiu and Jin 2024). These computational advances have catalysed a burgeoning interest in deploying generative AI for design support and the systematic analysis of consumer insights (Chamola et al. 2024; Chen et al. 2025; Vats et al. 2024).

However, existing language-based approaches encounter fundamental obstacles when integrated into formal engineering design contexts. User feedback is typically characterised by perceptual and metaphorical descriptors (e.g. ‘comfort,’ ‘smoothness,’ or ‘heavy steering’) that lack a one-to-one mapping with engineering parameters or physical variables (Delle Monache, Misdariis, and Özcan 2022; Wang, Li, and Mo 2021). While LLMs excel at capturing linguistic nuances, they often lack the robust mechanisms required to align these expressions with multi-level engineering semantic spaces, including design attributes, subsystem behaviours, and functional hierarchies (Ettrich et al. 2024; Kannan 2021; Wang, Luo, and Liu 2020). As highlighted in engineering informatics, this lack of grounding results in the potential misinterpretation of technical relevance and severely restricts the reusability of insights across heterogeneous design cases, effectively forming a persistent semantic-epistemic gap (Chen and Luo 2019; Demoly, Kim, and Horváth 2019; Han et al. 2021).

This limitation is particularly pronounced in EV design, where user experience emerges from the tight coupling of powertrain response, chassis dynamics, and human-machine interaction (HMI) (Gomez et al. 2024; Liu et al. 2023; Vermesan et al. 2021; Yang et al. 2021). Current methods often oversimplify rich, multi-dimensional experiential descriptions into coarse sentiment indicators, thereby precluding a rigorous, systematic comparison across diverse product iterations (Wang et al. 2025; Xu et al. 2024; Yang et al. 2024). These limitations underscore the urgent necessity for Structured Semantic Encoding (SSE). Such a mechanism aims to explicitly map weakly structured consumer narratives onto a stable, multi-level engineering framework, enabling perceptual language to be transformed into comparable and traceable semantic tokens for high-fidelity downstream design analysis (Huang et al. 2023; Li et al. 2025; Meng et al. 2024; Mustapha 2025).

2.2. Integrating knowledge structures for deep engineering reasoning

In automotive and EV studies, Importance-Performance Analysis (IPA), satisfaction modelling, and competitive benchmarking have been widely adopted to identify design weaknesses and relative market advantages (Jia et al. 2025; Joung and Kim 2021; Shi et al. 2025; Wang et al. 2023). These approaches typically utilise probabilistic or statistical aggregation of user opinions to derive high-level product indicators that support high-stakes design evaluation and strategic decision-making (Martilla and James 1977; Wang et al. 2025).

In parallel, research on user modelling and preference analysis emphasises the heterogeneity of user needs. Advanced clustering algorithms, persona construction, and latent preference profiling methods have been applied to segment users based on their expressed interests or longitudinal behavioural patterns (He, Xue, and Luo 2025; Ma et al. 2019; Sun et al. 2023). In the EV domain specifically, these methods are instrumental in distinguishing between distinct archetypes, such as performance-oriented early adopters, comfort-focused urban commuters, and cost-sensitive pragmatic users (Gupta and Kumar 2022; Liao, Molin, and Van Wee 2017; Qian et al. 2023; Uy et al. 2024).

Despite these advances, existing studies often suffer from the conflation of product-centric performance assessment and user-centric preference representation within a monolithic analytical process. Product-level aggregation tends to obscure preference diversity by averaging out divergent voices, while user-level models are rarely anchored to stable, comparable engineering dimensions (Luo et al. 2024; Vermesan et al. 2021). As a result, it remains difficult to interpret how different user segments evaluate identical design features or to support multi-objective trade-off analysis under conflicting preferences. This issue is particularly critical in EV design, where a single vehicle architecture must often reconcile contradictory evaluations regarding range, performance, and cabin technology from diverse market segments (Gomez et al. 2024; Huang et al. 2021; Liu et al. 2023).

This theoretical gap highlights the necessity for Product-User Dynamic Mapping (PUDM), which explicitly decouples product-side importance-performance modelling from user-side preference representation while maintaining a shared semantic basis (Ettrich et al. 2024; Wang, Luo, and Liu 2020). By bifurcating these two analytical perspectives, PUDM enables both rigorous product benchmarking and preference-aware design reasoning to be derived from the same corpus of consumer-generated narratives. This decoupling ensures that design insights remain interpretable, traceable, and adaptable to changing

market segments without losing their grounding in engineering reality (Gao et al. 2024; Li et al. 2025; Mustapha 2025).

2.3. Hybrid paradigm for insight retrieval and recommendation

Engineering design has long relied on structured knowledge representations to maintain reasoning, traceability, and rigorous decision-making. Foundational frameworks, such as the Function-Behaviour-Structure (FBS) model and associated formal ontologies, provide the essential vocabularies and causal hierarchies required for reasoning across various design abstractions (Chen and Luo 2019; Demoly, Kim, and Horváth 2019; Gero 1990). Recent advancements in Engineering Knowledge Graphs (EKGs) have further demonstrated their efficacy in representing complex component interdependencies, design constraints, and multi-domain relationships, particularly within the intricate systems of modern vehicles (Han et al. 2021; Huang et al. 2023; Ji et al. 2021).

However, these knowledge-based approaches are primarily optimised for structured inputs, such as CAD metadata, parametric tables, or formal requirement documents, and remain ill-equipped to ingest the perceptual and ambiguous nature of consumer language. Currently, linguistic processing and engineering reasoning often operate as decoupled systems: NLP pipelines (e.g. BERT-based architectures) extract latent themes or sentiments, while knowledge graphs perform symbolic inference on rigid data structures (Chen et al. 2024; Devlin et al. 2019; Qiu and Jin 2024). Without an explicit architectural link between these layers, design insights remain ‘shallow,’ and reasoning cannot be traced back to the original textual evidence, creating a transparency crisis in automated design support (Doshi-Velez and Kim 2017; Li et al. 2025; Ribeiro, Singh, and Guestrin 2016).

To address this, recent research has pivoted towards hybrid retrieval and reasoning paradigms that fuse the semantic flexibility of vector embeddings with the structural rigour of symbolic logic (Bordes et al. 2013; Lin, Socher, and Xiong 2018; Mehra 2024; Wang et al. 2017). While promising, most existing systems, including early Retrieval-Augmented Generation (RAG) frameworks, often rely on shallow alignment between text and knowledge entities, frequently neglecting the nuanced modelling of user profiles and the rationalisation of recommendations (Lin et al. 2024; Vats et al. 2024; Wu et al. 2024). Recent breakthroughs have attempted to address this by introducing self-supervised rationalisation and context-enhanced reasoning over knowledge graphs (Yuan, Chang, and Han 2023) to ensure that design recommendations are not only accurate but also structurally justified (Liu et al. 2024; Yang et al. 2023; Zhu et al. 2025). In the context of EV design, where decisions must be justified across coupled subsystems (e.g. thermal management vs. cabin comfort) and diverse stakeholder perspectives, such limitations hinder the adoption of AI in professional workflows (Gupta and Kumar 2022; Uy et al. 2024; Vermesan et al. 2021).

These challenges necessitate the development of an Engineering Design Knowledge Graph (EDKG) that integrates user archetypes, product attributes, and perceptual features into a unified reasoning space (Li et al. 2021; Meng et al. 2024; Wang et al. 2021). By supporting graph-constrained inference and hybrid retrieval-reasoning (HRR), such a framework enables the generation of evidence-backed insight chains that are both computationally robust and engineer-readable. This approach directly addresses the urgent demand for explainable and traceable design intelligence in experience-driven automotive engineering (Gao et al. 2024; Li et al. 2025; Mustapha 2025).

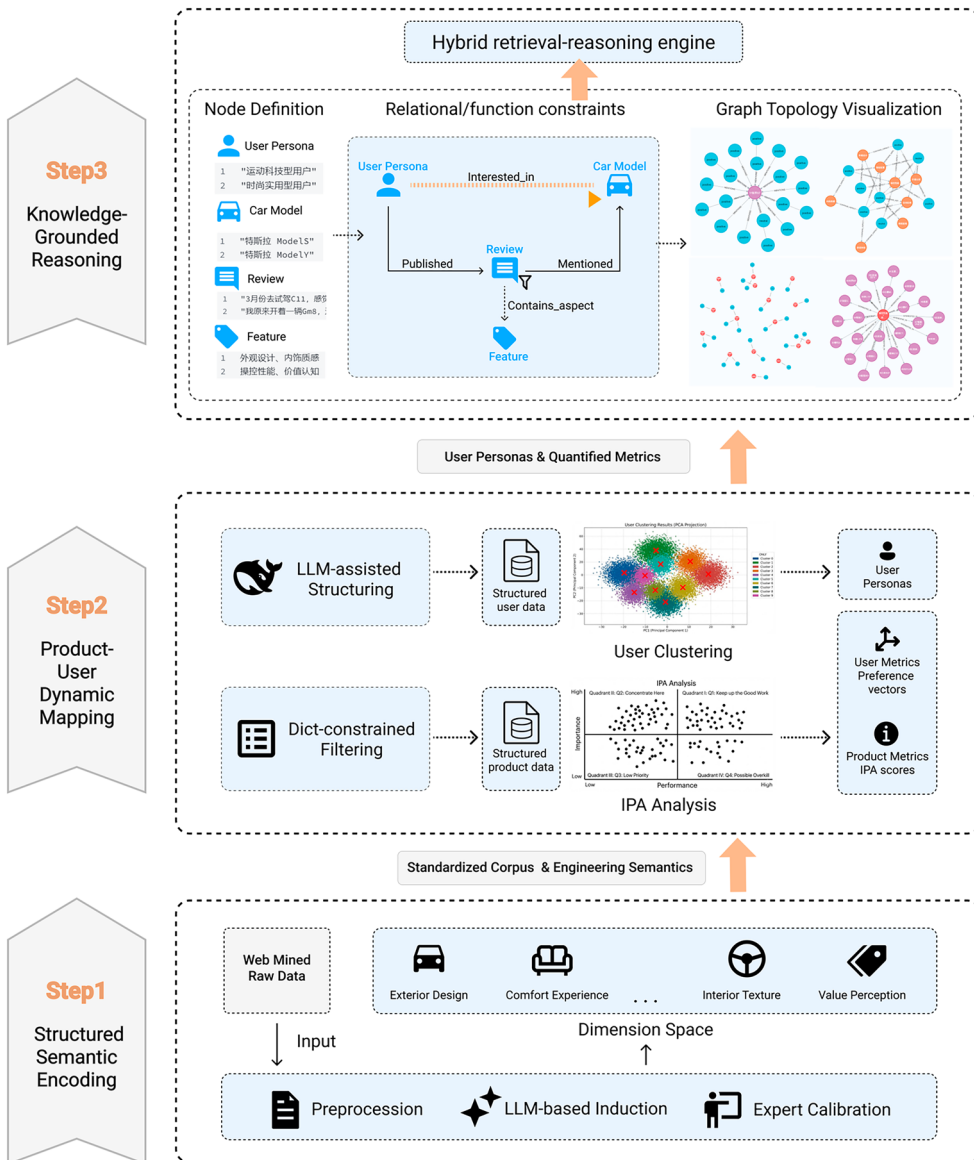


Figure 1. The SCSI-SLM framework.

3. Structured consumer semantic insight small language model (SCSI-SLM) framework

This section presents the Structured Consumer Semantic Insight Small Language Model (SCSI-SLM), an LLM-augmented design intelligence framework (Figure 1) that systematically transforms weakly structured consumer narratives into grounded, comparable, and traceable engineering insights.

The framework consists of three tightly coupled components: (1) Structured Semantic Encoding (SSE), (2) Product-User Dynamic Mapping (PUDM), and (3) Knowledge-Grounded

Reasoning (EDKG). Together, these components establish a traceable pipeline from user-generated narratives to design-relevant insights, while preserving both linguistic richness and engineering interpretability. To ensure theoretical clarity, the formal definitions of key semantic constructs utilised throughout this framework are summarised in Appendix Table A1.

3.1. Structured semantic encoding (SSE)

Structured Semantic Encoding aims to bridge the semantic-cognitive gap between perceptual user language and formal engineering representations. By constructing a systematic mapping mechanism, vague and subjective user expressions are transformed into computable, reusable, and comparable semantic tokens, which serve as the unified input for subsequent product-user mapping and knowledge reasoning.

3.1.1. Data acquisition and preprocessing

To ensure the reliability of downstream analysis, a standardised data acquisition and preprocessing pipeline is established to transform online user-generated content (UGC) into high-quality, computable corpora.

In the data acquisition stage, an automated crawling workflow is implemented using a visual DOM parsing tool (Octoparse). The crawler simulates client-side browsing behaviour to handle AJAX-based dynamic loading and employs predefined DOM path parsers to map unstructured webpage content into structured data fields.

In the preprocessing stage, raw data typically contain redundant attributes and substantial noise. To improve semantic quality and vectorisation reliability, the following procedures are applied:

- (1) **Regex Denoising:** Utilising regular expressions to remove URL links and non-alphanumeric special characters.
- (2) **Feature Selection:** Eliminating irrelevant items and retaining core dimensions.
- (3) **Short Review Filtering:** Removing low-information reviews with a character length less than M or effective words fewer than N .
- (4) **Tokenisation and Stop-word Filtering:** Adopting the **jieba** algorithm for precise-mode tokenisation, combined with a custom stop-word list to remove meaningless high-frequency function words (such as ‘的’, ‘了’).

After preprocessing, each review is represented as a standardised data object:

$$D = E_{id}, C_{raw}, C_{clean}, S_{tokens}$$

where E_{id} denotes the entity identifier, C_{raw} the original text, C_{clean} the cleaned corpus, and S_{tokens} the tokenised sequence. This representation ensures traceability from engineering inference back to original user evidence.

3.1.2. Engineering semantic space construction

User narratives are typically expressed through perceptual descriptors (e.g. ‘comfortable,’ ‘smooth,’ ‘heavy steering’) that lack direct correspondence to engineering parameters. To

align such expressions with design reasoning, this study constructs a multi-level engineering semantic space through a hybrid workflow combining LLM-based induction and expert calibration.

The construction process consists of three stages:

- (1) LLM-based open feature extraction: Using a zero-shot prompting strategy, the LLM is guided to extract core perceptual features and supporting keywords from a sampled review set, emulating a domain expert's perspective.
- (2) Semantic induction and consolidation: Extracted heterogeneous labels are clustered in a bottom-up manner. Through batch processing, the LLM identifies semantic similarity among labels, consolidating synonymous or overlapping expressions into refined feature categories with unambiguous descriptions.
- (3) Expert calibration: To ensure engineering validity and cross-product comparability, refined labels are aligned with industry evaluation standards (e.g. CPCA indices) and reduced to a stable set of core dimensions through expert review.

The result is an engineering semantic space:

$$D = d_1, d_2, \dots, d_k$$

which serves as a shared coordinate system for both product-level evaluation and user-level preference modelling.

3.2. Product-user dynamic mapping (PUDM)

While SSE provides a unified semantic foundation, design reasoning requires distinguishing what is evaluated from who evaluates it. PUDM addresses this need by decoupling the same UGC corpus into two complementary information streams:

- Product-oriented stream, which treats reviews as crowdsourced measurements of product performance;
- User-oriented stream, which interprets reviews as projections of individual and group preferences.

This decoupling enables simultaneous modelling of design trade-offs and preference heterogeneity without information loss.

3.2.1. Product-side: importance-performance quantification

For the product perspective, the core objective is to transform unstructured text into structured 'Performance-Importance' binary indicators. This approach is theoretically grounded in the orthogonality assumption of Importance-Performance Analysis (IPA) (Martilla and James 1977), which posits that consumer attention (frequency) and satisfaction (sentiment) are distinct cognitive dimensions that must be decoupled to accurately prioritise design improvements. Where performance values represent user sentiment and importance represents user attention weight. Considering the domain stability and computational efficiency of feature descriptions, a hybrid NLP workflow is adopted. First, a feature localisation mechanism based on domain knowledge is established. A vertical domain lexicon is constructed

through a combination of LLM mining and manual review. This is paired with the **jieba** algorithm and a sliding window approach to intercept context segments of specific lengths centred on feature words, achieving semantic decoupling of multi-feature evaluations in long texts. Subsequently, a pre-trained **RoBERTa** model optimised for Chinese contexts is used to mine the sentiment polarity of each feature segment, which is finally aggregated at the object level to output macro product performance and importance values.

(1) Perceived Importance I

To reflect differences in attention across feature dimensions, perceived importance is defined as a function of feature mention rates. First, the raw mention proportion w_k for feature k is calculated:

$$w_k = \frac{n_k}{\sum_{j=1}^K n_j}$$

where n_k denotes the mention frequency of the k -th dimension in the corpus, and K denotes the total number of dimensions. To address the long-tail polarisation issue in raw distributions, a truncation and re-normalisation strategy is introduced to truncate the raw proportion within the threshold interval $[\alpha, \beta]$, followed by a secondary normalisation to ensure the sum equals 1, yielding the final importance I_k .

(2) Perceived Performance Index P

Perceived performance represents the user satisfaction level of a specific model c on dimension k . First, the sentiment mapping function $S(x)$ for a single segment is defined:

$$S(x) = \begin{cases} 0.5 + 0.5 \cdot p_{prob}, & \text{if } L = \text{Positive} \\ 0.5 - 0.5 \cdot p_{prob}, & \text{if } L = \text{Negative} \end{cases}$$

where p_{prob} represents the confidence probability predicted by the model, and L represents the sentiment classification label. This mapping ensures that sentiment scores are monotonic within the $[0, 1]$ continuous interval, with 0.5 representing absolute neutrality.

Based on this, the dimension performance for a specific vehicle model is defined as the score aggregation of the set of all relevant feature segments $M_{c,k}$:

$$P_{c,k} = \frac{1}{|M_{c,k}|} \sum_{x \in M_{c,k}} S(x)$$

(3) IPA Decision Space Mapping

Based on the two indicators, the IPA decision matrix is constructed. Utilising the decision framework established by Martilla and James (Martilla and James 1977), each feature is represented as a coordinate $(P_{c,k}, I_k)$ in a 2D space, and four semantic quadrants are divided based on the median values of the dataset:

- **Q1 (Keep up the good work):** High Importance, High Performance;
- **Q2 (Concentrate here):** High Importance, Low Performance;

- **Q3 (Low priority):** Low Importance, Low Performance;
- **Q4 (Possible overkill):** Low Importance, High Performance.

3.2.2. User-side: preference vectorisation and profiling

For the user perspective, the core challenge lies in extracting common group characteristics from discrete and polysemous individual expressions. Theoretically, direct clustering of such raw textual data is hindered by the ‘curse of dimensionality’ and high sparsity, where semantically similar but lexically distinct descriptors are treated as orthogonal (Aggarwal and Zhai 2012). To overcome this, we leverage the LLM as a semantic projector to map discrete linguistic variances into a continuous, dense preference manifold (Reimers and Gurevych 2019). First, LLM-based contextual semantic disambiguation is introduced. For metaphors and vague references in user expressions, prompt templates (shown in Figure 2) are constructed to guide the LLM to map non-standard descriptions precisely to the 8 core dimensions, breaking the limitations of traditional dictionary matching. Second, multi-dimensional intent vectorisation is performed. A ‘Sentiment-Frequency’ hybrid weighting strategy is adopted to transform qualitative reviews into continuous numerical vectors in a high-dimensional space to preserve preference granularity. Finally, unsupervised group induction is executed. Clustering algorithms are used to partition the user vector space, converging massive discrete users into several typical personas with significant commonalities, thereby reducing the computational complexity of subsequent graph construction and retrieval.

(1) User Attention Vector Construction

The attention intensity $I_{u,d}$ for user u on dimension d is defined, with the calculation process as follows:

```
[System Instruction] You are a professional automotive industry analyst, specializing in extracting feature dimensions of user interest from user reviews. Your task is to analyze each review, identify specific dimensions of user concern, and extract relevant keywords. Please ensure accuracy and consistency in your analysis.
```

```
[User Input Template] Please analyze the following automotive user reviews and extract the dimensions of user concern and relevant keywords for each review.
```

```
Optional Dimension Types: {dimension_desc}
```

```
Review List: {batch_comments_json}
```

```
For each review, please:
```

1. Identify the user focus dimensions reflected in the review.
2. Extract keywords corresponding to each dimension.
3. Evaluate the user's sentiment tendency for that dimension (Positive/Negative/Neutral).

```
Please output strictly according to the following JSON format:
```

```
{
  "results": [
    {
      "comment_index": 0,
      "dimensions": {
        "Exterior Design": {
          "keywords": ["Keyword1", "Keyword2"],
          "sentiment": "Positive"
        },
        "Interior Texture": {
          "keywords": ["Keyword3"],
          "sentiment": "Neutral"
        }
      }
    }
  ]
}
```

Figure 2. Automated persona naming prompt template.

- (i) **Base Frequency Intensity** I_{base} : Calculated based on the number of extracted keywords N_{kw} , with a threshold Ω set for saturation:

$$I_{base} = \min\left(\frac{N_{kw}}{\Omega}, 1.0\right)$$

- (ii) **Sentiment Weight** M_{senti} : This study posits that positive evaluations better reflect users' active selection criteria and identity, while negative evaluations often stem from experience gaps. To enhance the sensitivity of persona clustering to core value orientations, asymmetric weights are set:

$$M_{senti} = \begin{cases} 1.2, & \text{if sentiment is Positive} \\ 1.0, & \text{if sentiment is Neutral} \\ 0.8, & \text{if sentiment is Negative} \end{cases}$$

- (iii) **Weighted Dimensional Intensity** $I_{u,d}$:

$$I_{u,d} = \min(I_{base} \times M_{senti}, 1.0)$$

Ultimately, each user is represented as a D -dimensional vector (In this project, $D = 8$):

$$V_u = [I_{u,1}, I_{u,2}, \dots, I_{u,8}]$$

(2) Clustering and Abstraction of Typical Personas

To address the sparsity of individual vectors, this study constructs a group feature extraction framework based on unsupervised learning.

- (i) **Dimensionality Reduction and Initialisation**: First, PCA (Principal Component Analysis) is utilised to reduce the 8D vector space to extract principal components. Subsequently, the K-means++ algorithm is used to initialise cluster centres to accelerate convergence and avoid local optima.
- (ii) **Optimal Cluster Number Decision Mechanism**: Determining the cluster number K is a result of balancing feature coverage and semantic interpretability. The search space is set as a defined range $\Omega_K = [K_{min}, K_{max}]$. The lower bound K_{min} ensures basic persona diversity; the upper bound K_{max} suppresses over-segmentation to prevent the graph topology from becoming too fragmented. The silhouette coefficient is used as the evaluation metric. The final K_{opt} is selected as the value within the search space that has a high silhouette coefficient and is close to K_{max} .
- (iii) **Persona Generation**: This strategy aims to maximise the retention of long-tail features in the niche market while ensuring cluster tightness and avoiding meaningless micro-noise clusters. The final set of cluster centre vectors C is defined as typical user personas.

3.3. Knowledge-grounded reasoning (EDKG)

This section achieves a transition from static data storage to dynamic knowledge reasoning by mapping user intents and product attributes into a unified structured graph, thereby

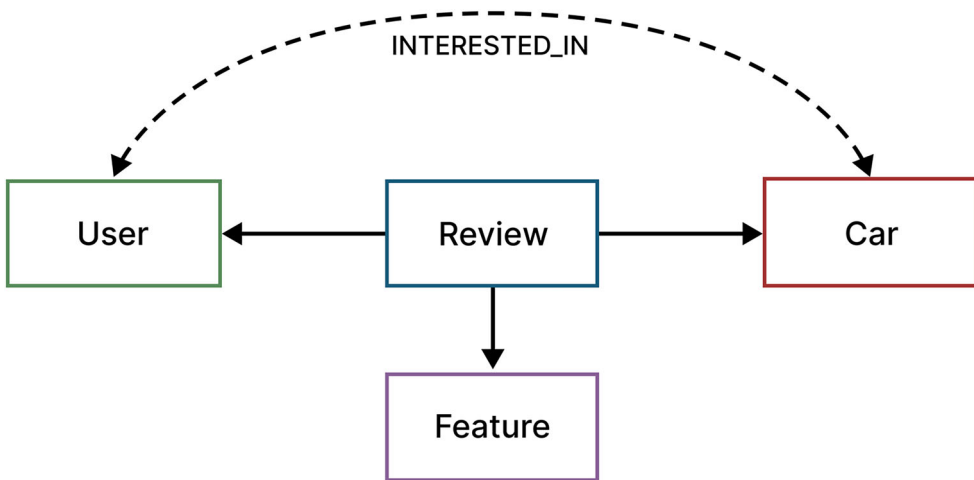


Figure 3. Logical entity-relationship schema of the experience-grounded design knowledge graph.

discovering latent association paths between users and products. Neo4j is selected as the native graph database engine, utilising its graph storage architecture and efficient path traversal capabilities to support complex computational reasoning tasks.

This section first defines the graph schema as $G = (E, R, P)$, where E denotes the set of entities, R denotes the set of relations, and P denotes the set of attributes attached to edges or nodes.

As shown in Figure 3, this study builds a logical structure with Review as the hub, connecting the user subject (User), product object (Car), and Feature.

3.3.1. Knowledge graph node schema

Entity and Associated Attribute Type Definitions

The entity set E contains four core categories of nodes:

- (1) **Car Node:** Derived from M mainstream NEV models acquired in Section 3.1.1. Each node represents a specific model as the evaluated object. Attributes include modelId, name, brand, and reviewCount.
- (2) **User Persona Node:** Derived from the 30 typical user personas clustered in Section 3.2.2. Each node represents a group of users with similar preference characteristics, serving as the cognitive subject of graph reasoning. Attributes include profileId, name, description, userCount, percentage, mainFeatures, and dimension Strengths.
- (3) **Review Node:** Derived from filtered and cleaned review data. Each node represents a specific projection of user intent, serving as the key connection point between users and products and the primary explicit evidence for reasoning. Attributes include reviewId, content, overallSentiment, and importance.

2 key indicators are quantified as follows:

Table 1. Definition of node attributes and associated properties.

| Node type | Semantic description | Attribute | Explanation |
|-------------|--------------------------------|---|---|
| CarModel | Vehicle model entity | modelId name brand reviewCount | Unique identifier of the vehicle model Official display name of the vehicle Manufacturer or brand Total number of associated user reviews |
| UserProfile | User persona | profileId name description userCount percentage mainFeatures dimensionStrengths | Unique identifier of the persona Persona name Textual description of the persona Estimated number of users in this persona Proportion of users in the overall population Dominant perceptual features Attention weights across engineering dimensions |
| Review | User review evidence | reviewId content overallSentiment importance | Unique identifier of the review Original user-generated text Aggregated sentiment polarity Evidence importance score (0–1) |
| Feature | Engineering semantic dimension | name category description | Dimension name Dimension category Detailed semantic description |

- (i) **Quality Weighting (importance):** Calculated based on text length and keyword density:

$$\text{Importance} = S_{len} \times M + S_{kw} \times N$$

where the normalised scores are $S_{len} = \min(\text{contentLength}/X, 1.0)$ and $S_{kw} = \min(\text{totalKeywords}/Y, 1.0)$. X, Y is the saturation benchmark; M, N are weighting coefficients satisfying $M + N = 1$.

- (ii) **Sentiment Aggregation (overallSentiment):** Weighted sum based on fine-grained sentiment polarity:

$$\text{Score} = \frac{\sum_{d \in D} V_{senti} \times I_d}{\sum_{d \in D} I_d}$$

where $V_{senti,d}$ represents the discrete sentiment polarity value (assigned as 1.0 for Positive, -1.0 for Negative, and 0.0 for Neutral), and I_d denotes the focus intensity defined in Section 3.1.2. A filtering threshold Q is introduced to ensure that a specific dimension is included in the calculation only if $I_d > Q$.

- (4) **Feature Node:** Derived from the K -dimensional engineering semantic space defined in Section 3.1.2. Each node represents a fine-grained evaluation dimension. Attributes include name, category, and description. Specific node-attribute associations are detailed in Table 1.

3.3.2. Relation schema and preference inference

Relation and Associated Attribute Type Definitions

The relation set R includes three explicit associations and one implicit inference.

- (1) **PUBLISHED** (*User* → *Review*): This relation links user personas to their associated review evidence. It includes a userMatchScore attribute, quantifying how representative a given review is for the persona:

$$S_{match} = \cos(\theta) = \frac{\mathbf{V}_{rev} \cdot \mathbf{C}_{prof}}{\|\mathbf{V}_{rev}\| \|\mathbf{C}_{prof}\|}$$

Where V_{rev} represents the D dimensional user attention vector extracted from the review; C_{prof} denotes the central feature vector of the corresponding persona obtained through clustering. The resulting score $\in [-1, 1]$.

- (2) **MENTIONS** (*Review* → *Car*): This relation specifies the evaluated product and carries sentimentScore and importance attributes. The sentiment score is computed as a weighted average over dimension intensities:

$$S_{senti} = \frac{\sum_{d=1}^D v_{senti,d} \cdot I_d}{\sum_{d=1}^D I_d}$$

where $v_{senti,d}$ represents the discrete sentiment mapping value. This mechanism ensures that feature dimensions with high user focus (i.e. a large I_d) dominate the overall score.

- (3) **CONTAINS_ASPECT** (*Review* → *Feature*): Represents the specific dimensions involved. Attributes include intensity (0-1) and aspectSentiment (POS/NEG/NEU).
- (4) **INTERESTED_IN** (*User* → *Car*): An implicit inference edge derived through path reasoning algorithms, representing the group's comprehensive preference for specific products. Attributes include correlationScore, topAspects, positiveMentions, and negativeMentions. The correlation score is computed as:

$$S_{corr} = \frac{1}{|R|} \sum_{r \in R} S_{match}^{(r)} \cdot S_{senti}^{(r)} \cdot w_{imp}^{(r)}$$

where R is the set of reviews under the specific user persona that mention the vehicle model¹. This formula reinforces the confidence of evidence that is highly representative, emotionally significant, and of high quality.

To ensure interpretability, each inferred relation also outputs topAspects, positiveMentions, and negativeMentions, providing explicit explanations for the inferred preference.

Based on the above representations and inference mechanisms, an LLM-augmented hybrid RAG system is constructed to support the full insight retrieval pipeline, from vague user queries to professional, evidence-backed, and interpretable design insights. The system implementation and evaluation are detailed in Section 4.3.2.

4. Case study: electric vehicle design

Building on the SCSI-SLM framework proposed in Section 3, this section transitions from methodological formulation to engineering implementation and empirical validation. Using real user-generated content (UGC) from Dongchedi, a leading Chinese automotive vertical media platform, we develop a prototype system for knowledge-graph-augmented

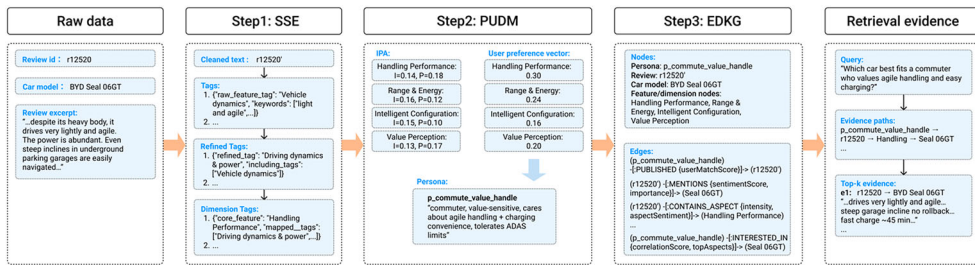


Figure 4. A minimal working example (MWE) of the knowledge-grounded reasoning pipeline in the EV recommendation system.

retrieval in the new energy vehicle (NEV) domain. The case study follows the one-to-one correspondence defined in Section 3. A minimal working example is shown in Figure 4, illustrating the step-by-step process from raw data to retrieval evidence.

4.1. Structured semantic encoding (SSE)

4.1.1. Data acquisition and preprocessing

To ensure both data quality and industrial representativeness, Dongchedi was selected as the primary data source. As one of China's mainstream automotive information platforms, it hosts a large active user base and extensive, experience-rich UGC. Data collection was restricted to the top K NEV models by sales and user attention; in this study, $K = 50$ to cover mainstream market offerings.

The acquisition pipeline proceeded as follows. First, the crawler navigated to the 'Dongche Score (Public Praise)' pages for each target model and parsed the page DOM structure to map key fields. Second, an automated execution flow was configured to simulate user browsing behaviour, including loop-based pagination and full-volume crawling. Finally, batch jobs parsed raw HTML into a structured CSV format. As of July 25, 2025, the pipeline collected 53,219 raw reviews covering 50 vehicle models and 17 auxiliary attributes (e.g. user level, likes, replies).

Raw crawled data contained redundant fields (e.g. nicknames, user levels) and substantial unstructured noise. Following the preprocessing procedure in Section 3.1.1, we applied feature selection, regex-based denoising, short-text filtering, and tokenisation. The initial screening thresholds were set to $M = 10$ (minimum character length) and $N = 3$ (minimum effective tokens). Appendix Table A2 illustrates a representative cleaning pipeline for a noisy sample.

After preprocessing, 14,153 high-confidence reviews were retained from the original 53,219 records. The resulting semi-structured dataset follows the data model and contains four core fields:

- **Vehicle Entity:** Serves as the primary node identifier for Knowledge Graph construction and is used to associate vehicle features.
- **Raw Corpus:** Retains the complete original user expressions for traceability and verification.

- **Cleaned Corpus:** Noise-free text primarily used for vectorisation and retrieval within the RAG system.
- **Token Sequence:** Tokenised sequences used for sentiment calculation and feature word extraction.

The storage schema and sample records are shown in Appendix Table A3.

4.1.2. Engineering semantic space construction

This section follows the feature extraction workflow proposed in Section 3.1.2, employing a strategy that combines ‘bottom-up’ data mining with ‘top-down’ expert knowledge:

- (1) **Seed Dataset Construction:** To balance computational efficiency and feature coverage, the preprocessed data from Section 4.1.1 was further filtered (length > 15) and randomly sampled to obtain a seed dataset of 500 samples.
- (2) **Open Feature Extraction:** LLM has been shown to effectively perform semantic labelling and annotation tasks in related works (Ding et al. 2023; Zhang et al. 2023a). A **zero-shot prompting** (shown in Appendix Figure 1) template was constructed to guide the LLM, acting as an ‘automotive market insight expert’, to mine core attention dimensions (‘raw feature tags’) using a batch processing strategy (Batch Size = 10). The system integrated result verification and auto-retry mechanisms to ensure stability. **DeepSeek-V2.5** was selected for this step, extracting 2,307 raw feature tags. Examples of the extracted raw feature tags are presented in Appendix Table A4, including both successful and failed cases. All identified failure cases were manually corrected before entering the downstream normalisation pipeline.
- (3) **Semantic Induction and Mapping:** The nested raw tags were flattened and input into the LLM using the semantic normalisation prompt (shown in Appendix Figure 2) in batches of 50. Acting as a ‘Senior Market Analyst’, the model mapped heterogeneous labels (e.g. ‘power consumption’ and ‘range anxiety’) into unified **Refined Tags** and generated descriptive categories. This process yielded 90 unique refined tags.
- (4) **Expert Calibration:** Referencing **CPCA** industry indices and **Dongchedi** evaluation systems, the 90 refined tags were mapped to 8 core feature dimensions: Exterior Design, Interior Texture, Intelligent Configuration, Spatial Utility, Comfort Experience, Handling Performance, Range & Energy, and Value Perception. The final mapping of the 8 dimensions was manually verified to ensure accuracy and consistency. Mapping examples are shown in Table 2.

4.2. Product-user dynamic mapping (PUDM)

4.2.1. Product-side: importance-performance quantification

This section validates the effectiveness of the product-side ‘Performance-Importance’ binary quantification model (Martilla and James 1977). Applying the sentiment mapping function $S(x)$ and importance weighting formula (with a truncation interval of [0.05, 0.25]), engineering indices for 8 core dimensions were calculated for various models. An **IPA** decision matrix was then constructed by weighted aggregation.

Table 2. Core engineering feature dimensions and representative refined tags.

| Core feature dimension | Representative refined tags (examples) |
|---------------------------|---|
| Exterior Design | Exterior design, exterior & styling, body size and design, design & personalisation, vehicle styling, exterior evaluation, ... |
| Interior Texture | Interior design and texture, interior comfort, interior configuration, workmanship and reliability, material quality, interior style, minimalist interior, ... |
| Intelligent Configuration | Intelligent technology, intelligent driving and technology, infotainment system and connectivity, human-machine interaction, intelligent configuration, driver assistance and safety, ... |
| Spatial Utility | Space performance, space and usability, storage space, space layout, trunk space, seating space, insufficient space, ... |
| Comfort Experience | Comfort, noise insulation, audio system, rear-seat comfort, NVH performance, seat comfort, ride comfort, ... |
| Handling Performance | Power performance, driving experience, power and handling, chassis tuning, suspension performance, braking performance, steering performance, ... |
| Range & Energy | Driving range performance, range and energy consumption, charging and thermal management, charging speed, energy recovery, power consumption, ... |
| Value Perception | Price and cost, cost-effectiveness, purchase and service experience, after-sales service, brand perception, maintenance cost, resale value, ... |

Table 3. Fine-grained performance quantisation statistics for Zeekr 001.

| Core feature dimension | Perceived importance | Perceived performance | mention rate | IPA quadrant |
|---------------------------|----------------------|-----------------------|--------------|----------------------------|
| Handling Performance | 0.1508 | 0.1107 | 87.55% | Q1 – Keep up the good work |
| Range & Energy | 0.1440 | 0.1127 | 84.13% | Q1 – Keep up the good work |
| Intelligent Configuration | 0.1439 | 0.1019 | 89.06% | Q2 – Concentrate here |
| Spatial Utility | 0.1363 | 0.1099 | 71.09% | Q2 – Concentrate here |
| Exterior Design | 0.1239 | 0.1115 | 80.20% | Q4 – Possible overkill |
| Interior Texture | 0.1235 | 0.0980 | 67.08% | Q3 – Low priority |
| Value Perception | 0.0879 | 0.1110 | 45.11% | Q4 – Possible overkill |
| Comfort Experience | 0.0896 | 0.1013 | 56.64% | Q3 – Low priority |

Taking the **Zeekr 001** as an example, results indicate a significantly unbalanced distribution across dimensions. Using the median perceived importance $\alpha \approx 0.1301$ and median perceived performance $\beta \approx 0.1103$ as thresholds, the IPA decision space was constructed. Detailed quantitative indices and intermediate variables are specified in Table 3. The constructed IPA decision space is illustrated in Figure 5.

Quadrant Analysis:

- **Q1 (Keep up the good work):** ‘Handling Performance’ ($I = 0.1508, P = 0.1107$) and ‘Range & Energy’ ($I = 0.1440, P = 0.1127$) are located here, representing core competitive advantages. High frequency mentions such as ‘Six-sided warrior’ confirm the market positioning of **Zeekr 001** as a ‘driver’s car.’
- **Q2 (Concentrate here):** ‘Spatial Utility’ ($I = 0.1363, P = 0.1099$) and ‘Intelligent Configuration’ ($I = 0.1439, P = 0.1019$) are located here. Although the car is physically large, negative feedback regarding the difficulty of parking due to its width identifies a design pain point. Software lag (‘laggy system’) is also highlighted as a core shortcoming.

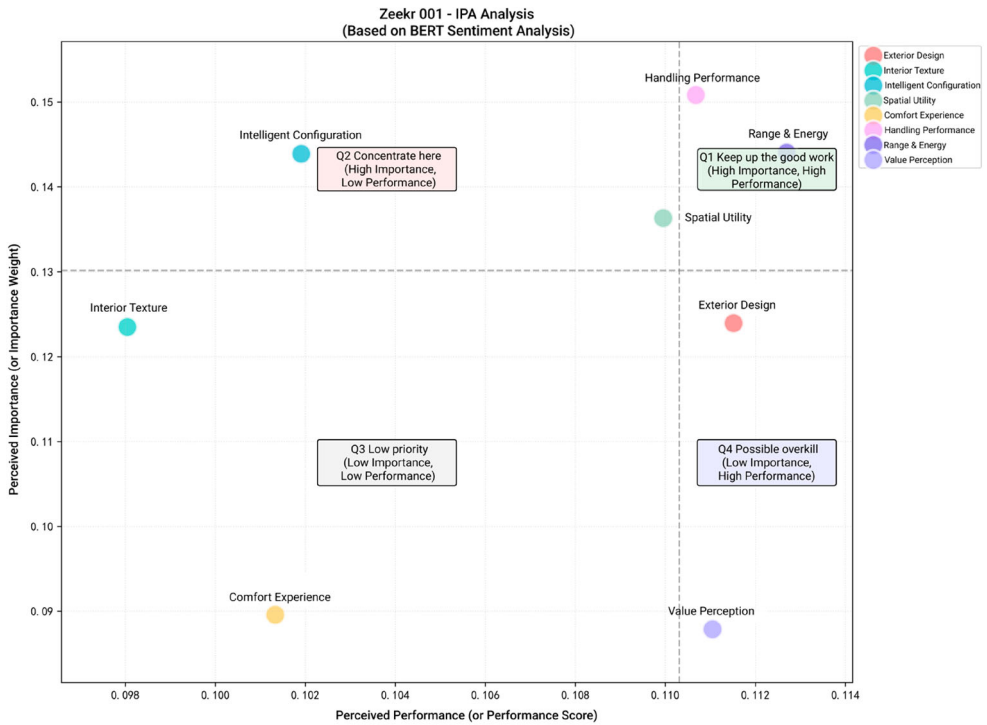


Figure 5. Importance-performance analysis (IPA) decision matrix for Zeekr 001.

- **Q3 (Low priority):** ‘Interior Texture’ ($I = 0.1235, P = 0.0980$) shows full-dimension minimum performance. Despite high-quality materials, complaints about ‘seat bulging’ and ‘abnormal noises’ lowered the overall rating, suggesting a need to prioritise basic assembly craftsmanship.
- **Q4 (Possible overkill):** ‘Exterior Design’ and ‘Value Perception’ are hidden advantages where performance is high but user discussion is relatively low due to established consensus or aesthetic saturation.

4.2.2. User-side: preference vectorisation and profiling

This section applies multi-dimensional intent vectorisation and unsupervised clustering to abstract 14,153 discrete users into typical group personas.

(1) Clustering Process and Decision-making

Consistent with the search-space guidelines in Section 3.2.2, we restricted the candidate range to $\Omega_K = [10, 30]$. After PCA-based dimensionality reduction and K-means++ initialisation, the silhouette coefficient peaked at $K = 30$, within this interval, accompanied by improved intra-cluster coherence (Figure 6, top-left). We therefore select $K_{opt} = 30$ as a pragmatic design choice that balances representational granularity against downstream retrieval requirements, rather than as a claim of a global mathematical optimum. The absolute silhouette values are relatively low; in the high-dimensional user-semantic manifold



Figure 6. Results of K-means++ clustering analysis on the user semantic space.

considered here such modest scores are interpretable for two reasons: **(i) non-orthogonal semantic overlap** – real car owners’ needs exhibit multi-dimensional coupling (Figure 6, top-right), which blurs cluster boundaries and suppresses distance-based separation metrics; and **(ii) continuity of the intent space** – user intents form a largely continuous distribution in embedding space rather than well-separated discrete islands, so compactness/separation metrics can underestimate the representational usefulness of a given discretisation. Accordingly, we treat clustering principally as a vector-quantisation and indexing strategy that trades off interpretability, long-tail coverage, and retrieval granularity for the requirements of our RAG and graph-based retrieval components. The resulting 30 clusters thus serve as practical persona anchors for handling complex queries near blurred boundaries and produce representative user personas for the EV market (Table 4).

(2) Persona Definitions

Based on $K_{opt} = 30$, the proposed approach yields 30 distinct user personas representing heterogeneous preference structures in the EV market. Table 4 summarises four representative personas that illustrate the diversity of user priorities captured by the model.

- **C26 (Trendy Performance)**: Focuses on ‘Handling Performance’ (weight 0.846) and ‘Exterior Design’ (0.765). Represents 3.5% of users.
- **C22 (Comfort Experience)**: Highly concentrated on ‘Comfort Experience’ (0.629). Represents 4.3% of users.

Table 4. Typical user persona clustering characteristics.

| Cluster ID | Persona description | User ratio | Rank | Key feature dimension weights |
|------------|-----------------------------|------------|------|---|
| C1 | Value-Oriented | 4.10% | 5 | Value Perception (0.80) |
| C22 | Comfort-Oriented | 4.30% | 4 | Comfort Experience (0.63) |
| C26 | Trendy Performance-Oriented | 3.50% | 11 | Handling Performance (0.85), Exterior Design (0.77), Range & Energy (0.74) |
| C0 | All-Round Balanced | 3.30% | 15 | Handling Performance (0.86), Range & Energy (0.81), Comfort Experience (0.72), Intelligent Configuration (0.69) |

- **C1 (Value Cognitive):** Primarily concerned with price and cost-effectiveness ('Value Perception' 0.796). Represents 4.1% of users.
- **C0 (All-round Balanced):** High intensity across Handling, Range, Comfort, and Intelligence. Represents 3.3% of users.

4.3. Knowledge-grounded reasoning (EDKG)

4.3.1. Domain knowledge graph construction

By aligning entities from structured personas, multi-dimensional evaluation features, and massive review texts based on the graph schema $G = (E, R, P)$, the **EV-KG** was instantiated. The graph contains 13,768 nodes and 65,248 edges. Figure 7 illustrates its local topological structure, with specific statistics detailed in Table 5.

4.3.2. Hybrid retrieval-reasoning

This section describes the construction of an **LLM-augmented Hybrid RAG (Retrieval-Augmented Generation) system** based on the **LangChain** framework and the **GLM-4-9B-Chat** model. Designed to bridge the gap between vague user intentions and professional data insights, the system's core architecture (as illustrated in Figure 8) follows the **Modular RAG** (Gao et al. 2024) paradigm. It implements a complete pipeline of 'intent recognition – dynamic routing – hybrid retrieval – explainable generation' tailored to the data processed in previous sections.

Differentiated Storage and Tool Construction

Targeting the structured tables from Section 4.2 and the graph-text from Section 4.3.1, the system constructs a differentiated vector storage architecture. Three categories of retrieval tools are encapsulated to adapt to various types of user queries:

- **Vector Retrieval Tool:** Powered by the **Chroma** database, this tool utilises **Embedding cosine similarity** to process 'vague semantic' queries (such as subjective experience descriptions) by recalling unstructured document fragments through fuzzy matching.
- **Graph Cypher Tool:** Powered by the **Neo4j** graph database, this tool addresses 'statistical and ranking' queries (such as safety rating rankings). It leverages the LLM to translate natural language into **Cypher** queries using a schema-constrained generation template (prompts shown in Appendix Figure 3), which are then executed by the database engine for precise aggregation and sorting operations.
- **Hybrid Retrieval Tool:** Designed for 'complex reasoning and comparative' queries, this tool adopts a 'vector anchor + graph traversal' retrieval mode. The system uses vector



Figure 7. EV knowledge graph.

Table 5. Scale statistics of the EV knowledge graph (EV-KG).

| Node label | Count | Key properties | Incoming/Outgoing relationships |
|-------------|--------|------------------------------------|--|
| CarModel | 48 | brand, reviewCount, handling_score | (:Review)-[:ABOUT]-> (:CarModel) |
| UserProfile | 30 | percentage, mainFeatures | (:Review)-[:BELONGS_TO]-> (:UserProfile) |
| Feature | 8 | category, description | (:Review)-[:MENTIONS]-> (:Feature) |
| Review | 13,682 | importance, overallSentiment | Linking UserProfile, CarModel, Feature |
| Total | 13,768 | - | 65,248 (Edges) |

indexing to locate review nodes as semantic anchors within the graph and expands context along explicit relationship paths. This captures both semantic details and structured relational information, overcoming the limitations of single-mode retrieval.

Intent Parsing and Dynamic Scheduling

The system deploys an intent parsing module as the central scheduler. First, user inputs are preprocessed via regular expressions and keyword extraction. Subsequently, the LLM is

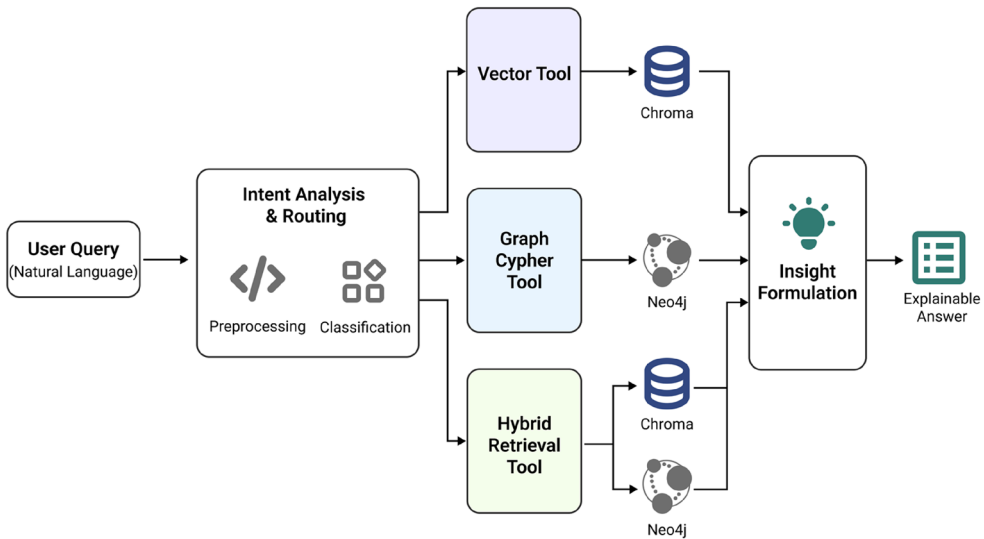


Figure 8. Framework of hybrid insight retrieval layer.

guided to perform **intent classification** (statistical / recommendation / comparison, etc.) and **entity extraction** on the cleaned query (prompts shown in Appendix Figure 4). Based on the classification results, the system executes a **dynamic routing strategy**, automatically matching the appropriate retrieval tools and loading adaptive parameters to achieve a self-adjusting balance between retrieval breadth and precision.

Insight Generation and Explainability

Upon obtaining retrieval results, the system standardises text fragments from Chroma and structured data from Neo4j into a unified sequence, injecting **citation index tags**. Finally, the LLM is constrained to perform reasoning and answering strictly within the retrieved context, explicitly labelling information sources to produce logically rigorous and traceable professional insights. The specific context injection format and instruction constraint templates are shown in Appendix Figure 5.

4.3.3. Validation

Case-Based Functional Verification

We introduce three specific query instances of different types to demonstrate and verify the actual operational mechanism and feasibility of the system pipeline. Three representative query cases are summarised in Table 6, illustrating how the system dynamically selects vector-based, graph-based, or hybrid retrieval strategies according to query intent.

Together, these cases demonstrate the flexibility and interpretability of the proposed hybrid retrieval-reasoning pipeline.

RAG Triad Evaluation

We employed the **RAG Triad framework** (Es et al. 2024; Ru et al. 2024; Saad-Falcon et al. 2024) to validate the superiority of the proposed architecture (Es et al. 2024; Ru et al. 2024; Saad-Falcon et al. 2024). The evaluation explicitly compares the **Hybrid Retrieval (Vector + Graph)** strategy with a standard **Vector-only Baseline** to demonstrate significant performance uplift.

Table 6. Representative design cases and corresponding retrieval strategies.

| Case | Query type | User request (simplified) | Retrieval strategy | Key outcome |
|---|---------------------------|---|-----------------------------------|---|
| A | Subjective recommendation | Recommend cars with a sporty exterior and a high-tech interior, along with user reviews | Vector-based semantic retrieval | Retrieved models such as NIO ET7 and Porsche Taycan, capturing design-related perceptions (e.g. fastback styling, streamlined body, large-screen interaction) |
| <p>C 请比较在车辆的颜值。请推荐几款外观设计偏向运动感且内饰具有科技感的车型。并列举用户的具体评价</p> <p>C 根据您的车辆颜值的要求，为您推荐以下几款外观设计偏向运动感且内饰具有科技感的车型，并列上用户的具体评价：</p> <p>1. 蔚来ET7</p> <ul style="list-style-type: none"> 外观设计：蔚来ET7的外观设计时间，前脸两侧设计了狭长的大灯，底部使用黑色饰板，整体辨识度高。侧面采用溜背式设计，车身饱满，尾部贯穿式尾灯提升了运动感，设计个性。 用户评价：“前几天去逛商场，在等朋友的时候，就看到了蔚来ET7这台车，感觉整体还不错，并且直接试驾了一下，外观设计颜值比较高，两侧还设计了狭长的大灯，底部使用了黑色的格栅设计，前脸设计的辨识度。”——来自用户的正面评价。 <p>2. 保时捷Taycan</p> <ul style="list-style-type: none"> 外观设计：保时捷Taycan的外观设计延续了保时捷家族化风格，流畅的车身线条、低矮的车身、标志性的四点式LED大灯，以及车头两侧和下方的空气动力学套件，影响了整体的运动感。 用户评价：“前几天去朋友家看车，他这几年攒了点钱就买了，非要去看保时捷，我就陪他去了一下，并且还试驾了保时捷Taycan，总体感觉还可以。总感觉比亚迪的有点区别，不过动力也还不错。作为保时捷，那外观方面自然不必多谈，犀利的车身，流畅的线条，极具科技感的车型，相信就算是老司机你也会知道是保时捷。”——来自用户的正面评价。 <p>3. 零跑C16</p> <ul style="list-style-type: none"> 外观设计：零跑C16的车身长度适中，外观设计时间，接近五米的车长即给人一种紧凑的感觉。车身线条流畅，颜值高。 用户评价：“颜值能打，接近五米的车长。”——来自用户的正面评价。 <p>4. 智界S7</p> <ul style="list-style-type: none"> 外观设计：智界S7作为华为科技赋能的高端纯电轿车，外观设计独特，延续了运动感和21寸轮毂，与赛道版拉开了差距。 用户评价：“在拿到这款智界S7 Max RS 自动驾驶版的时候，仅从外观我就感觉到不太一样。因为之前试驾过比亚迪版，这辆搭载了运动套件和21寸轮毂的智界S7一下就和普通版拉开了差距。”——来自用户的正面评价。 <p>以上车型均具有较高的颜值和科技感，您可以根据自己的需求和喜好进行选择。</p> | | | | |
| B | Statistical ranking | List top 5 models in space and comfort, with user keywords | Graph-based Cypher querying | Generated ranked results based on structured importance-performance scores, avoiding numerical hallucination |
| <p>C 列出空间与舒适性评分最高的5款车型，并给出用户对其“动力体验”及“内饰质感”的综合评分</p> <p>C 根据图中用户口碑数据的聚合统计，在“空间与舒适性”维度评分最高的5款车型及其关联体验评分如下：</p> <ol style="list-style-type: none"> 小米SU7 Ultra (空间舒适: 4.92) <ul style="list-style-type: none"> 动力体验评分: 4.82 (用户评价: 动力极其狂暴, 赛道级调校) 内饰质感评分: 4.90 (用户评价: 满屏碳纤维, 战斗感拉满) 宝马i5 M60 (空间舒适: 4.89) <ul style="list-style-type: none"> 动力体验评分: 4.90 (用户评价: 底盘非常有韧性, 碾碎十足) 内饰质感评分: 4.58 (用户评价: 静谧性做得很好, 老牌豪华) 宝马7 M70 (空间舒适: 4.86) <ul style="list-style-type: none"> 动力体验评分: 4.74 (用户评价: 加速深不见底, 从容优雅) 内饰质感评分: 4.66 (用户评价: 后排大屏很震撼, 移动行宫) 小鹏MONA M03 (空间舒适: 4.82) <ul style="list-style-type: none"> 动力体验评分: 4.66 (用户评价: 转向轻便, 城市代步很舒服) 内饰质感评分: 4.82 (用户评价: 内饰虽然简单, 但该有的都有) 小米YU7 (空间舒适: 4.80) <ul style="list-style-type: none"> 动力体验评分: 4.58 (用户评价: 动力输出线性, 跟脚) 内饰质感评分: 4.74 (用户评价: 座舱设计很像SU7, 细节有优化) <p>数据基于 Neo4j 图数据库 Cypher 多维聚合查询结果</p> | | | | |
| C | Comparative reasoning | Compare negative feedback on the intelligent cockpits of Xiaomi SU7 and Zeekr 001 | Hybrid retrieval (vector + graph) | Identified differentiated pain points in infotainment systems with traceable review evidence |
| <p>C 对比小米SU7和极氪001在智能座舱方面的用户反馈差异，重点分析用户对两款车车机系统的负面吐槽主要集中在哪里？</p> <p>C 对比小米SU7和极氪001在智能座舱方面的用户反馈，我们可以从以下几个方面进行分析：</p> <p>1. 小米SU7:</p> <ul style="list-style-type: none"> 根据用户评论，小米SU7的智能座舱系统在用户体验上得到了一定的认可，用户提到，小米SU7的智能配置丰富，科技感强，语音智能功能能够满足日常使用需求。 负面反馈主要集中在以下几个方面： <ul style="list-style-type: none"> 系统响应速度：部分用户反映车机系统的反应速度不够快，有时会出现卡顿现象。 人机交互：部分用户对语音识别的准确性有待提高，有时无法正确理解指令。 <p>2. 极氪001:</p> <ul style="list-style-type: none"> 极氪001的智能座舱同样获得了用户的青睐，用户评论中提到，极氪001的内饰质感好，智能配置先进，车机系统流畅，交互体验良好。 负面反馈主要集中在以下几个方面： <ul style="list-style-type: none"> 系统复杂性：部分用户认为车机系统的操作复杂，对于不熟悉智能设备的用户来说，上手难度较高。 系统稳定性：有用户反映，在特定情况下，车机系统会出现不稳定的情况，如偶尔死机或无法连接。 <p>总结：小米SU7和极氪001在智能座舱方面都表现出色，但用户反馈显示，小米SU7在系统响应速度和语音识别准确性上存在不足，而极氪001则在系统操作复杂度和稳定性方面有待提升。这两款车型的智能座舱系统各有优势，用户在选择时可以根据自己的需求和偏好进行考虑。</p> | | | | |

Table 7. Comparative performance on RAG triad metrics.

| Metric | Vector-only baseline (Mean \pm SD) | Hybrid retrieval (Mean \pm SD) | Performance uplift (Δ) | <i>p</i> -Value |
|---------------------|---|-------------------------------------|------------------------------------|-----------------|
| Context Relevance | 2.13 \pm 0.59 | 3.60 \pm 0.71 | +1.47 | < 0.001 |
| Answer Relevance | 3.57 \pm 0.54 | 4.07 \pm 0.40 | +0.50 | < 0.001 |
| Answer Faithfulness | 2.57 \pm 0.40 | 3.73 \pm 0.65 | +1.17 | < 0.001 |

A test set consisting of **100 domain-specific queries** was constructed based on the high-frequency intent categories identified in Section 4.2. The queries covered diverse engineering scenarios, including complex feature comparisons and persona-based recommendations.

For each query, answers were generated using both the **Vector-only baseline** (α_v) and the **Hybrid strategy** (α_h). To ensure objective assessment, a blinded evaluation process was implemented. Two independent automotive domain experts scored the anonymised outputs on a 5-point Likert scale (0.5 step intervals) across three standard RAG metrics:

Context Relevance (CR): Measures the precision of retrieved evidence. It assesses whether the retrieved document chunks or graph triples contain necessary information without introducing irrelevant noise.

Answer Faithfulness (AF): Measures the grounding of the generated response. It assesses whether the answer is strictly derived from the retrieved context, serving as a critical indicator for preventing hallucinations in engineering applications.

Answer Relevance (AR): Measures the alignment between the generated response and the user's intent, assessing completeness and directness.

The comparative results are summarised in Table 7.

The analysis highlights that **the proposed Hybrid Retrieval consistently outperforms the Vector-only baseline across all three RAG Triad dimensions**, confirming a clear performance uplift attributable to graph-enhanced reasoning. The largest gains are observed in Context Relevance and Answer Faithfulness ($\Delta CR = +1.47$, $p < 0.001$; $\Delta AF = +1.17$, $p < 0.001$), indicating that the KG constraints substantially improve evidence precision and reduce irrelevant retrieval noise, while also strengthening grounding by encouraging responses to stay within the bounds of retrieved evidence. In contrast, the improvement in Answer Relevance is more moderate ($\Delta AR = +0.50$, $p < 0.001$), suggesting that while a strong LLM can already generate fluent and seemingly on-topic answers under vector retrieval, the Hybrid strategy yields responses that are more consistently aligned with the intended engineering dimensions and better supported by verifiable context.

Evaluation of the service quality

This study adopts the SERVQUAL instrument. Following the methodology described by Chiu et al. (2021), the service quality (Q) is defined as the gap between user expectations (E) and their actual perception (P) of the system's performance.

A total of 30 subjects, consisting of professional designers, automotive engineers, and senior STEM students, were recruited to assess the service quality of the SCSI-SLM system. Prior to the experiment, all subjects were provided with a standardised orientation regarding the system's core modules, including Structured Semantic Encoding (SSE), Product-User

Table 8. Statistical Results of SERVQUAL Questionnaire for SCSi-SLM.

| Statement | Expectation (E) | Perception (P) | $Q = P - E$ | Stat. power ($1 - \beta$) |
|--------------------------------------|-----------------|----------------|-------------|-----------------------------|
| 1. Tangibles | | | | |
| SCSi-SLM's charts are clear | 6.13 | 4.67 | -1.46 | 0.996 |
| The interface is attractive | 6.23 | 3.83 | -2.40 | 1.00 |
| The outputs are well-organised | 6.03 | 5.17 | -0.86 | 0.998 |
| 2. Reliability | | | | |
| SCSi-SLM provides accurate details | 6.63 | 5.23 | -1.40 | 1.00 |
| SCSi-SLM performs as promised | 6.47 | 4.73 | -1.74 | 1.00 |
| The analysis results are error-free | 6.70 | 5.37 | -1.33 | 1.00 |
| 3. Responsiveness | | | | |
| SCSi-SLM responds quickly | 5.83 | 4.77 | -1.06 | 0.999 |
| SCSi-SLM understands vague queries | 5.87 | 6.23 | 0.36 | 0.719 |
| SCSi-SLM addresses my specific needs | 6.00 | 5.63 | -0.37 | 0.999 |
| 4. Assurance | | | | |
| I trust SCSi-SLM's advice | 6.57 | 5.27 | -1.30 | 1.00 |
| SCSi-SLM shows expert knowledge | 6.20 | 5.53 | -0.67 | 0.926 |
| It supports answers with evidence | 6.40 | 4.80 | -1.60 | 0.994 |
| 5. Empathy | | | | |
| SCSi-SLM understands my preferences | 5.77 | 5.67 | -0.10 | 0.327 |
| It adapts to my specific situation | 5.80 | 5.70 | -0.10 | 1.00 |
| It prioritises my key needs | 5.83 | 5.37 | -0.46 | 0.516 |

Dynamic Mapping (PUDM), and the Engineering Design Knowledge Graph (EDKG). Participants performed retrieval tasks using the prototype system and subsequently completed the SERVQUAL questionnaire.

A statistical power analysis was conducted to ensure inferential confidence. The results (detailed in Table 8) indicate that the statistical power ($1 - \beta$) for detecting significant gaps in critical dimensions – specifically Tangibles, Reliability, and Assurance – consistently exceeds 0.90, surpassing the conventional threshold of 0.80. This confirms that the test provides sufficient sensitivity to validate the observed service quality deficits.

The statistical findings indicate that the hybrid retrieval-reasoning engine effectively addresses the perceptual and weakly structured nature of consumer language, with the statement 'SCSi-SLM understands vague queries' achieving a positive gap. Furthermore, minimal negative gaps in the Empathy dimension demonstrate the efficacy of the PUDM module and user persona nodes in capturing preference heterogeneity and aligning with engineering semantics.

Significant gaps in Tangibles, Reliability, and Assurance identify specific prototype constraints and areas for future optimisation. The lowest Q value for interface attractiveness is attributed to the current Streamlit-based front-end, where the utilitarian layout lacks the distinct visual hierarchy and interactive feedback typical of commercial software. Within the Reliability dimension, negative gaps regarding promised performance and accurate details suggest a divergence in information expectation: while the system effectively structures subjective User-Generated Content (UGC), expert users anticipate deterministic, metric-based engineering conclusions, differing from the qualitative nature of consumer reviews. Additionally, the trust deficit regarding Assurance reflects both response latency and a transparency issue in the inference process. The observed delays stem from the computational cost of real-time Cypher query generation and graph traversal; meanwhile, since the front-end did not explicitly visualise these backend graph constraints, users – accustomed to the hallucination risks of standard LLMs – perceived a lack of verification mechanisms, highlighting a gap between system capability and user perception. These results confirm

that while the framework excels at semantic mediation, strengthening the interface aesthetics, reducing inference latency and visualising the reasoning path remain priorities for professional adoption.

5. Discussion

5.1. Methodological value of SCSI-SLM for engineering design intelligence

This study contributes to engineering design research by advancing consumer voice analysis from descriptive text analytics towards structured design intelligence. While prior work has demonstrated the usefulness of sentiment analysis, topic modelling, and, more recently, large language models (LLMs) for summarising user feedback (Li et al. 2023; Wang et al. 2023; Zhang et al. 2023b), these approaches predominantly treat language as an unstructured signal to be predicted, clustered, or generated. Such treatments remain insufficient for engineering design, where decisions must be justified, compared, and traced across design alternatives and iterations (Ettrich et al. 2024; Wang, Luo, and Liu 2020).

SCSI-SLM introduces a methodological shift by explicitly separating language understanding from design reasoning. In this framework, language models are not used as end-to-end decision makers but as structured semantic mediators embedded within an engineered reasoning pipeline. Structured semantic encoding establishes a stable semantic interface between perceptual expressions and engineering dimensions, product-user dynamic mapping disentangles evaluation and preference, and knowledge-grounded reasoning ensures traceability and accountability. This modularisation directly addresses longstanding concerns in engineering informatics regarding non-grounded semantics, lack of causal transparency, and limited reusability of language-based insights (Doshi-Velez and Kim 2017; Nikolic et al. 2023; Qiu and Jin 2024; Ribeiro, Singh, and Guestrin 2016).

From a broader perspective, this work aligns with calls in design research to move beyond performance-oriented AI tools towards design intelligence systems that support reasoning, interpretation, and decision accountability. Rather than optimising prediction accuracy, SCSI-SLM prioritises representational clarity and inferential structure, which are critical for adoption in real-world design processes. Taken together, these characteristics distinguish SCSI-SLM from prior language-based design analytics by enabling structured, comparable, and accountable reasoning rather than descriptive summarisation alone.

5.2. Practical implications and future applicability in EV design

The EV case study provides insight into how structured consumer voice analysis can support design reasoning in experience-driven and systemically complex products. EV user experience is shaped by tightly coupled subsystems, including powertrain behaviour, chassis dynamics, software responsiveness, and human-machine interaction, making it difficult to isolate causes from subjective evaluations (Gomez et al. 2024; Liu et al. 2023; Yang et al. 2021). The case results demonstrate that when experiential language is structured and decoupled appropriately, consumer feedback can reveal meaningful design patterns rather than contradictory opinions (Liang et al. 2024; Xu et al. 2024).

Product-side importance-performance patterns highlight dimensions where perceived performance fails to match user expectations, offering a data-driven basis for design

prioritisation. Meanwhile, user-side preference profiling reveals that these gaps are not uniformly perceived across the market (Joung and Kim 2021; Sun et al. 2023; Wang et al. 2025; Yang et al. 2024). This finding underscores a critical insight: design weaknesses are often preference-dependent rather than absolute. In EV development, where resource constraints force trade-offs, understanding which user segments are affected by which shortcomings is as important as identifying the shortcomings themselves (He, Xue, and Luo 2025; Ma et al. 2019; Qian et al. 2023).

Importantly, the case study illustrates that SCSI-SLM supports early-stage and strategic design intelligence rather than post-hoc evaluation alone. By preserving heterogeneity and evidence structure, the framework enables designers to reason about alternative improvement paths, anticipate market response diversity, and align technical decisions with experiential priorities, capabilities that are difficult to achieve with aggregate sentiment scores or single-index benchmarks.

From an application perspective, the proposed framework is particularly suitable for early-stage design exploration, concept evaluation, and iterative refinement, where formal engineering data may be incomplete but consumer feedback is abundant. As EV products increasingly rely on continuous software updates and rapid iteration cycles, SCSI-SLM provides a scalable mechanism for integrating evolving consumer voice into ongoing design decision-making.

5.3. Implications for design knowledge representation and intelligence

A deeper contribution of this work lies in its expansion of what constitutes design knowledge. Classical engineering knowledge representations emphasise functions, behaviours, structures, and physical relationships (Demoly, Kim, and Horváth 2019; Gero 1990). While indispensable, these representations largely exclude experiential, perceptual, and preference-driven knowledge because such knowledge lacks formal structure and measurable parameters. As a result, a significant portion of real-world design rationale, why users prefer one design over another, remains external to formal design models (Chen and Luo 2019; Kang 2021; Wang, Li, and Mo 2021; Wei, Hao, and Wang 2025).

The Engineering Design Knowledge Graph (EDKG) proposed in this study offers a complementary form of design knowledge representation. By integrating user profiles, perceptual features, product attributes, and textual evidence into a unified graph, the framework captures experience-grounded design knowledge that is typically absent from engineering repositories (Han et al. 2021; Huang et al. 2023; Ji et al. 2021; Yuan et al. 2025). Crucially, this integration introduces a parallel layer of evidence-backed relational reasoning that reflects how experiential judgments and smart product-service interactions are formed and aggregated (Yuan et al. 2024). In complex domains such as EVs, this hybrid knowledge representation provides a more complete foundation for design intelligence.

This integration does not aim to replace physics-based causality with linguistic reasoning. Instead, it introduces a parallel layer of evidence-backed relational reasoning that reflects how experiential judgments are formed and aggregated. In complex design domains such as EVs, where user acceptance depends on both objective performance and subjective perception, such hybrid knowledge representation provides a more complete foundation for design intelligence (Gao et al. 2024; Mustapha 2025; Vermesan et al. 2021).

In this sense, SCSI-SLM does not merely integrate language and knowledge, but reframes experiential feedback as a legitimate and structured component of engineering design intelligence.

5.4. Limitations and future work

Despite its contributions, this study has several limitations that define its appropriate scope of application. First, the engineering semantic space relies on LLM-assisted extraction combined with expert calibration, implying that the ontology is not zero-shot transferable to unrelated fields. However, this methodology demonstrates high reliability for complex engineered systems characterised by explicit functional definitions (e.g. consumer electronics, smart home appliances). In these vertical domains, the reliance on expert knowledge serves as a critical quality guarantee, achieving significantly higher semantic precision compared to generic LLM approaches.

Second, semantic definitions may evolve over time as technologies and user discourse change. Without active maintenance, this could reduce comparability. Nevertheless, within the timeframe of a standard product development cycle, core engineering semantics remain relatively stable. The framework maintains longitudinal reliability for iterative design, provided that the semantic space is periodically refreshed to incorporate emerging features.

Third, regarding data bias and algorithmic reliability, we acknowledge that metrics derived from mention frequency may be subject to user review bias, as users tend to overemphasize easily verbalised features (e.g. visual design) while overlooking latent needs (Joung and Kim 2021; Shi et al. 2025). Consequently, this framework is best scoped as a quantitative filter for explicit requirements and high-frequency pain points, rather than a discovery tool for tacit or subconscious needs. While it provides a statistically robust signal for prioritising articulated complaints, it cannot replace qualitative methods required to uncover unvoiced user desires.

Furthermore, regarding LLM hallucination risks, LLMs inherently face the challenge of generating plausible but non-factual content, which poses a validity threat to engineering decision-making. To mitigate this, our framework employs Graph-Constrained Inference, anchoring reasoning strictly to retrieved evidence chains. This architectural constraint significantly reduces fabrication risks compared to open-ended generation, ensuring that insights remain traceable and factually grounded.

Fourth, while the reasoning supported by EDKG is structurally interpretable, it does not establish physical causality. Although the framework provides a transparent and accountable way to infer design insights from user feedback and product attributes, it cannot replace the need for engineering simulations, real-world testing, or empirical validation. Its primary role is to support prioritisation, hypothesis generation, and decision framing rather than final technical verification.

Fifth, we also acknowledge certain limitations regarding the quantitative validation of the proposed computational modules. Specifically, with regard to **persona validation**, since our user preference archetypes are derived exclusively from public UGC without accessible demographic metadata, they function as engineering intent clusters rather than demographically anchored market segments. In terms of **semantic encoding**, while we relied on manual expert verification to ensure engineering rigour, we did not conduct a large-scale benchmark comparing our **Structured Semantic Encoding (SSE)** against

traditional keyword-matching baselines. Furthermore, although the system-generated insights were validated via a double-blind expert review (**RAG Triad Evaluation**), the scale of human evaluation was constrained by available resources. A full-scale **A/B testing experiment** within an active design cycle – comparing AI-generated insights directly against human-curated reports – remains a critical direction for future research to further quantify the framework's practical utility.

Future research can address these limitations by incorporating adaptive semantic space management, integrating human-in-the-loop validation for high-impact insights, and strengthening links between experiential knowledge and structured engineering data. Furthermore, expanding the scope of the case study to include multiple EV models, conducting cross-brand comparisons, and incorporating longitudinal analysis to capture sentiment drift over time would provide a more comprehensive evaluation. Applying the same framework to other experience-intensive engineered products, such as consumer electronics or service robotics, would further test its generality and reveal domain-specific adaptation requirements. Moreover, future research could explore how SCSI-SLM could inform business and design decision-making by guiding feature prioritisation, supporting early-stage concept screening, predicting consumer acceptance, and aiding in engineering trade-offs.

6. Conclusion

This paper proposes SCSI-SLM, an explainable design intelligence framework that enables consumer-generated narratives to be systematically transformed into actionable engineering insight. By integrating structured semantic encoding, product-user dynamic mapping, and knowledge-grounded reasoning, the framework establishes a traceable pathway from weakly structured experiential language to design-relevant representations.

A real-world case study in the EV domain demonstrates that SCSI-SLM can support preference-aware analysis and evidence-backed insight generation from large-scale user feedback, while preserving interpretability and accountability. Rather than treating language models as end-to-end decision makers, this work positions them as structured semantic mediators embedded within an engineering reasoning process. Overall, this study shows that structured language modelling provides a viable paradigm for incorporating large-scale consumer voice into data-driven engineering design. Future work will focus on semantic space evolution and broader application across experience-driven engineered systems.

Acknowledgements

The authors would like to thank the anonymous reviewers for their constructive comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This paper was supported by the Shanghai Planning Office of Philosophy and Social Science (grant number 2022BSH001), and the Ministry of Education in China Humanities and Social Sciences project (grant number 23YJA760006).

ORCID

Danni Chang  <http://orcid.org/0000-0002-6230-5796>

References

- Achiam, J., S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, et al. 2023. Gpt-4 Technical Report. Preprint, arXiv:2303.08774.
- Aggarwal, C. C., and C. X. Zhai. 2012. "A Survey of Text Clustering Algorithms." In *Mining Text Data*, edited by C. C. Aggarwal and C. X. Zhai, 77–128. New York, NY: Springer.
- Bordes, A., N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. 2013. "Translating Embeddings for Modeling Multi-relational Data." *Advances in Neural Information Processing Systems* 26: 2787–2795.
- Brown, T., B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, et al. 2020. "Language Models Are Few-Shot Learners." *Advances in Neural Information Processing Systems* 33:1877–1901.
- Chamola, V., S. Sai, R. Sai, A. Hussain, and B. Sikdar. 2024. "Generative AI for Consumer Electronics: Enhancing User Experience with Cognitive and Semantic Computing." *IEEE Consumer Electronics Magazine* 14 (2): 10–19. <https://doi.org/10.1109/MCE.2024.3387049>.
- Chang, D., L. Wang, Y. Xiang, X. Zhu, and C. H. Lee. 2025. "Exploring Human Perception in Interactive Digital Advertising: A Genetic-Kansei Engineering Approach with Human-AI Collaboration." *Knowledge-Based Systems* 311:113072. <https://doi.org/10.1016/j.knosys.2025.113072>.
- Chen, H., and X. Luo. 2019. "An Automatic Literature Knowledge Graph and Reasoning Network Modeling Framework Based on Ontology and Natural Language Processing." *Advanced Engineering Informatics* 42:100959. <https://doi.org/10.1016/j.aei.2019.100959>.
- Chen, J., S. Xiao, P. Zhang, K. Luo, D. Lian, and Z. Liu. 2024. "M3-Embedding: Multi-linguality, Multi-functionality, Multi-granularity Text Embeddings through Self-knowledge Distillation." In *Findings of the Association for Computational Linguistics: ACL 2024*, edited by L.-W. Ku, A. Martins, and V. Srikumar, 2318–2335. Bangkok: Association for Computational Linguistics.
- Chen, L., Z. Cai, W. Cheang, Q. Long, L. Sun, P. Childs, and H. Zuo. 2025. "AskNatureGPT: An LLM-Driven Concept Generation Method Based on bio-Inspired Design Knowledge." *Journal of Engineering Design* 42 (1): 238–272. <https://doi.org/10.1080/09544828.2025.2481536>.
- Chiu, M. C., J. H. Huang, S. Gupta, and G. Akman. 2021. "Developing a Personalized Recommendation System in a Smart Product Service System Based on Unsupervised Learning Model." *Computers in Industry* 128 (10): 103421. <https://doi.org/10.1016/j.compind.2021.103421>.
- Delle Monache, S., N. Misdaris, and E. Özcan. 2022. "Semantic Models of Sound-Driven Design: Designing with Listening in Mind." *Design Studies* 83:101134. <https://doi.org/10.1016/j.destud.2022.101134>.
- Demoly, F., K. Y. Kim, and I. Horváth. 2019. "Ontological Engineering for Supporting Semantic Reasoning in Design: Deriving Models Based on Ontologies for Supporting Engineering Design." *Journal of Engineering Design* 30 (10-12): 405–416. <https://doi.org/10.1080/09544828.2019.1633626>.
- Devlin, J., M. W. Chang, K. Lee, and K. Toutanova. 2019. "Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, edited by J. Burstein, C. Doran, and T. Solorio, 4171–4186. Minneapolis, MN: Association for Computational Linguistics.
- Ding, B., C. Qin, L. Liu, Y. K. Chia, B. Li, S. Joty, and L. Bing. 2023. "Is GPT-3 a Good Data Annotator?." In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by A. Rogers, J. Boyd-Graber, and N. Okazaki, 11173–11195. Toronto, Canada: Association for Computational Linguistics.
- Doshi-Velez, F., and B. Kim. 2017. Towards a Rigorous Science of Interpretable Machine Learning. Preprint, arXiv:1702.08608.
- Es, S., J. James, L. E. Anke, and S. Schockaert. 2024. "Ragas: Automated Evaluation of Retrieval Augmented Generation." In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, edited by N. Aletas and O. De Clercq, 150–158. St. Julian's, Malta: Association for Computational Linguistics.

- Ettrich, O., S. Stahlmann, H. Leopold, and C. Barrot. 2024. "Automatically Identifying Customer Needs in User-Generated Content Using Token Classification." *Decision Support Systems* 178:114107. <https://doi.org/10.1016/j.dss.2023.114107>.
- Gao, Y., Y. Xiong, M. Wang, and H. Wang. 2024. Modular Rag: Transforming Rag Systems into Lego-like Reconfigurable Frameworks. Preprint, arXiv:2407.21059.
- Gero, J. S. 1990. "Design Prototypes: A Knowledge Representation Schema for Design." *AI Magazine* 11 (4): 26–26.
- Gomez, R., L. Swann, P. Florentzos, and A. Singleton. 2024. "Automotive Interior Design in the Age of Electric Vehicles: User Interface Expectations, Perceptions and Preferences." In *Adjunct Proceedings of the 16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, edited by F. Steinicke, S. Loehmann, S. Sadeghian, and A. Riener, 82–87. New York, NY: Association for Computing Machinery.
- Gupta, A., and H. Kumar. 2022. "Multi-dimensional Perspectives on Electric Vehicles Design: A Mind map Approach." *Cleaner Engineering and Technology* 8:100483. <https://doi.org/10.1016/j.clet.2022.100483>.
- Han, J., S. Sarica, F. Shi, and J. Luo. 2021. "Semantic Networks for Engineering Design: A Survey." *Proceedings of the Design Society* 1:2621–2630. <https://doi.org/10.1017/pds.2021.523>.
- He, S., B. Xue, and D. Luo. 2025. "Research on the Evolution of Online User Reviews of New Energy Vehicles in China Based on LDA." *World Electric Vehicle Journal* 17 (1): 21. <https://doi.org/10.3390/wevj17010021>.
- Huang, X., Y. Lin, M. K. Lim, M. L. Tseng, and F. Zhou. 2021. "The Influence of Knowledge Management on Adoption Intention of Electric Vehicles: Perspective on Technological Knowledge." *Industrial Management & Data Systems* 121 (7): 1481–1495. <https://doi.org/10.1108/IMDS-07-2020-0411>.
- Huang, Z., X. Guo, Y. Liu, W. Zhao, and K. Zhang. 2023. "A Smart Conflict Resolution Model Using Multi-layer Knowledge Graph for Conceptual Design." *Advanced Engineering Informatics* 55:101887. <https://doi.org/10.1016/j.aei.2023.101887>.
- Ji, S., S. Pan, E. Cambria, P. Marttinen, and P. S. Yu. 2021. "A Survey on Knowledge Graphs: Representation, Acquisition, and Applications." *IEEE Transactions on Neural Networks and Learning Systems* 33 (2): 494–514. <https://doi.org/10.1109/TNNLS.2021.3070843>.
- Jia, X., M. Xu, C. Xu, and P. Zhu. 2025. "An Online Review-Driven Decision-Making Method for Electric Vehicle Selection Based on Satisfaction Analysis." *Expert Systems with Applications* 300: 130518.
- Joung, J., and H. M. Kim. 2021. "Approach for Importance–Performance Analysis of Product Attributes from Online Reviews." *Journal of Mechanical Design* 143 (8): 081705. <https://doi.org/10.1115/1.4049865>.
- Kang, X. 2021. "Combining Rough set Theory and Support Vector Regression to the Sustainable Form Design of Hybrid Electric Vehicle." *Journal of Cleaner Production* 304:127137. <https://doi.org/10.1016/j.jclepro.2021.127137>.
- Kannan, H. 2021. "Formal Reasoning of Knowledge in Systems Engineering through Epistemic Modal Logic." *Systems Engineering* 24 (1): 3–16. <https://doi.org/10.1002/sys.21563>.
- Li, H., Y. Wang, S. Zhang, Y. Song, and H. Qu. 2021. "KG4Vis: A Knowledge Graph-Based Approach for Visualization Recommendation." *IEEE Transactions on Visualization and Computer Graphics* 28 (1): 195–205. <https://doi.org/10.1109/TVCG.2021.3114863>.
- Li, Q., Y. Yang, C. Li, and G. Zhao. 2023. "Energy Vehicle User Demand Mining Method Based on Fusion of Online Reviews and Complaint Information." *Energy Reports* 9:3120–3130. <https://doi.org/10.1016/j.egy.2023.02.004>.
- Li, S., X. Zhou, Y. Liu, J. Chen, T. Guo, W. Yang, and L. Hou. 2025. "Agile Conceptual Design and Validation Based on Multi-source Product Data and Large Language Models: A Review, Framework, and Outlook." *Journal of Engineering Design* 36 (4): 473–503. <https://doi.org/10.1080/09544828.2025.2476879>.
- Liang, C., Q. Yang, H. Sun, and X. Ma. 2024. "Unveiling Consumer Satisfaction and Its Driving Factors of EVs in China Using an Explainable Artificial Intelligence Approach." *Humanities and Social Sciences Communications* 11 (1): 1575. <https://doi.org/10.1057/s41599-024-04120-z>.
- Liao, F., E. Molin, and B. Van Wee. 2017. "Consumer Preferences for Electric Vehicles: A Literature Review." *Transport Reviews* 37 (3): 252–275. <https://doi.org/10.1080/01441647.2016.1230794>.

- Lin, J., R. Shan, C. Zhu, K. Du, B. Chen, S. Quan, et al. 2024. "Rella: Retrieval-Enhanced Large Language Models for Lifelong Sequential Behavior Comprehension in Recommendation." In *Proceedings of the ACM Web Conference 2024*, edited by T. S. Chua, C. W. Ngo, R. K. Lee, R. Kumar, and H. W. Lauw, 3497–3508. New York, NY: Association for Computing Machinery.
- Lin, X. V., R. Socher, and C. Xiong. 2018. Multi-hop Knowledge Graph Reasoning with Reward Shaping. Preprint, arXiv:1808.10568.
- Liu, J., F. Wan, J. Zou, and J. Zhang. 2023. "Exploring Factors Affecting People's Willingness to use a Voice-Based in-car Assistant in Electric Cars: An Empirical Study." *World Electric Vehicle Journal* 14 (3): 73. <https://doi.org/10.3390/wevj14030073>.
- Liu, X., L. Yang, Z. Liu, M. Yang, C. Wang, H. Peng, and P. S. Yu. 2024. "Knowledge Graph Context-Enhanced Diversified Recommendation." In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, edited by L. A. Caudillo-Mata, S. Lattanzi, A. M. Medina, L. Akoglu, A. Gionis, and S. Vassilvitskii, 462–471. New York, NY: Association for Computing Machinery.
- Luo, S., P. Shan, Z. Bian, H. Lin, Y. Zhang, Z. Cui, C. Shen, and L. Wang. 2024. "Effects of Product Personalisation Degree on User Perception in car Front Design." *Journal of Engineering Design* 35 (8): 944–971. <https://doi.org/10.1080/09544828.2024.2339164>.
- Ma, S. C., Y. Fan, J. F. Guo, J. H. Xu, and J. Zhu. 2019. "Analysing Online Behaviour to Determine Chinese Consumers' Preferences for Electric Vehicles." *Journal of Cleaner Production* 229:244–255. <https://doi.org/10.1016/j.jclepro.2019.04.374>.
- Martilla, J. A., and J. C. James. 1977. "Importance-performance Analysis." *Journal of Marketing* 41 (1): 77–79. <https://doi.org/10.1177/002224297704100112>.
- Mehra, A. 2024. "Hybrid AI Models: Integrating Symbolic Reasoning with Deep Learning for Complex Decision-Making." *Journal of Emerging Technologies and Innovative Research* 11 (8): f693–f695.
- Meng, S., J. Zhou, X. Chen, Y. Liu, F. Lu, and X. Huang. 2024. "Structure-information-based Reasoning over the Knowledge Graph: A Survey of Methods and Applications." *ACM Transactions on Knowledge Discovery from Data* 18 (8): 1–42. <https://doi.org/10.1145/3671148>.
- Mustapha, K. B. 2025. "A Survey of Emerging Applications of Large Language Models for Problems in Mechanics, Product Design, and Manufacturing." *Advanced Engineering Informatics* 64:103066. <https://doi.org/10.1016/j.aei.2024.103066>.
- Nguyen-Duc, A., B. Cabrero-Daniel, A. Przybyłek, C. Arora, D. Khanna, T. Herda, et al. 2025. "Generative Artificial Intelligence for Software Engineering—A Research Agenda." *Software: Practice and Experience* 55 (11): 1806–1843.
- Nikolic, S., S. Daniel, R. Haque, M. Belkina, G. M. Hassan, S. Grundy, S. Lyden, P. Neal, and C. Sandison. 2023. "ChatGPT versus Engineering Education Assessment: A Multidisciplinary and Multi-institutional Benchmarking and Analysis of This Generative Artificial Intelligence Tool to Investigate Assessment Integrity." *European Journal of Engineering Education* 48 (4): 559–614. <https://doi.org/10.1080/03043797.2023.2213169>.
- Qian, L., Y. Huang, D. Tyfield, and D. Soopramanien. 2023. "Dynamic Consumer Preferences for Electric Vehicles in China: A Longitudinal Approach." *Transportation Research Part A: Policy and Practice* 176:103797. <https://doi.org/10.1016/j.tra.2023.103797>.
- Qiu, Y., and Y. Jin. 2024. "ChatGPT and Finetuned BERT: A Comparative Study for Developing Intelligent Design Support Systems." *Intelligent Systems with Applications* 21:200308. <https://doi.org/10.1016/j.iswa.2023.200308>.
- Reimers, N., and I. Gurevych. 2019. "Sentence-bert: Sentence Embeddings Using Siamese Bert-Networks." In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, edited by K. Inui, J. Jiang, V. Ng, and X. Wan, 3982–3992. Hong Kong, China: Association for Computational Linguistics.
- Ribeiro, M. T., S. Singh, and C. Guestrin. 2016. "“Why Should I Trust You?” Explaining the Predictions of Any Classifier." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, edited by B. Krishnapuram, B. Krishnapuram, M. Shah, A. J. Smola, C. Aggarwal, D. Shen, and R. Rastogi, 1135–1144. New York, NY: Association for Computing Machinery.

- Ru, D., L. Qiu, X. Hu, T. Zhang, P. Shi, S. Chang, et al. 2024. "Ragchecker: A Fine-Grained Framework for Diagnosing Retrieval-Augmented Generation." *Advances in Neural Information Processing Systems* 37:21999–22027.
- Saad-Falcon, J., O. Khattab, C. Potts, and M. Zaharia. 2024. "Ares: An Automated Evaluation Framework for Retrieval-Augmented Generation Systems." In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, edited by K. Duh, H. Gomez, and S. Bethard, 338–354. Mexico City, Mexico: Association for Computational Linguistics.
- Shi, L., S. S. Ou, Y. Zhou, Y. Wu, X. Tan, X. He, D. J. De Castro Gomez, and Z. Lin. 2025. "Assessing Chinese User Satisfaction with Electric Vehicle Battery Performance from Online Reviews." *Transportation Research Part D: Transport and Environment* 141:104644. <https://doi.org/10.1016/j.trd.2025.104644>.
- Sun, H., S. K. Ong, A. Y. C. Nee, and W. Guo. 2023. "A Customer Requirements Analysis Method of considering Product Scenarios for Improving Product Design." *Journal of Engineering Design* 34 (5–6): 339–362. <https://doi.org/10.1080/09544828.2023.2225843>.
- Uy, J. R. R., A. K. S. Ong, D. M. B. De Guzman, I. T. Dela Cruz, and J. C. Dela Cruz. 2024. "Consumer Segmentation and Market Analysis for Sustainable Marketing Strategy of Electric Vehicles in the Philippines." *World Electric Vehicle Journal* 15 (7): 301. <https://doi.org/10.3390/wevj15070301>.
- Vats, A., V. Jain, R. Raja, and A. Chadha. 2024. Exploring the Impact of Large Language Models on Recommender Systems: An Extensive Review. Preprint, arXiv:2402.18590.
- Vermesan, O., R. John, P. Pype, G. Daalderop, K. Kriegel, G. Mitic, V. Lorentz, et al. 2021. "Automotive Intelligence Embedded in Electric Connected Autonomous and Shared Vehicles Technology for Sustainable Green Mobility." *Frontiers in Future Transportation* 2:688482. <https://doi.org/10.3389/ffutr.2021.688482>.
- Wang, Q., Z. Mao, B. Wang, and L. Guo. 2017. "Knowledge Graph Embedding: A Survey of Approaches and Applications." *IEEE Transactions on Knowledge and Data Engineering* 29 (12): 2724–2743. <https://doi.org/10.1109/TKDE.2017.2754499>.
- Wang, X., Y. Cheng, T. Lv, and R. Cai. 2023. "Fuel Vehicles or new Energy Vehicles? A Study on the Differentiation of Vehicle Consumer Demand Based on Online Reviews." *Marketing Intelligence & Planning* 41 (8): 1236–1251. <https://doi.org/10.1108/MIP-04-2023-0173>.
- Wang, X., T. Huang, D. Wang, Y. Yuan, Z. Liu, X. He, and T. S. Chua. 2021. "Learning Intents behind Interactions with Knowledge Graph for Recommendation." In *Proceedings of the web Conference 2021*, edited by J. Leskovec, M. Grobelnik, M. Najork, J. Tang, and L. Zia, 878–887. New York, NY: Association for Computing Machinery.
- Wang, Y., X. Li, and D. Y. Mo. 2021. "Knowledge-empowered Multitask Learning to Address the Semantic gap between Customer Needs and Design Specifications." *IEEE Transactions on Industrial Informatics* 17 (12): 8397–8405. <https://doi.org/10.1109/TII.2021.3067141>.
- Wang, Y., L. Luo, and H. Liu. 2020. "Bridging the Semantic gap between Customer Needs and Design Specifications Using User-Generated Content." *IEEE Transactions on Engineering Management* 69 (4): 1622–1634. <https://doi.org/10.1109/TEM.2020.3021698>.
- Wang, Z., S. J. Hu, S. F. Niu, S. Y. Li, W. D. Liu, and L. Y. Huang. 2025. "Research on Product Design Improvement Method Based on Online Review and Improvement Importance Performance Competitor Analysis." *Expert Systems with Applications* 279:127400. <https://doi.org/10.1016/j.eswa.2025.127400>.
- Wei, W., C. Hao, and Z. Wang. 2025. "User Needs Insights from UGC Based on Large Language Model." *Advanced Engineering Informatics* 65:103268. <https://doi.org/10.1016/j.aei.2025.103268>.
- Wu, L., Z. Zheng, Z. Qiu, H. Wang, H. Gu, T. Shen, C. Qin, et al. 2024. "A Survey on Large Language Models for Recommendation." *World Wide Web* 27 (5): 60. <https://doi.org/10.1007/s11280-024-01291-2>.
- Xu, Y., X. Shan, M. Guo, W. Gao, and Y. S. Lin. 2024. "Design and Application of Experience Management Tools from the Perspective of Customer Perceived Value: A Study on the Electric Vehicle Market." *World Electric Vehicle Journal* 15 (8): 378. <https://doi.org/10.3390/wevj15080378>.
- Yang, H., J. Zhang, Y. Wang, and R. Jia. 2021. "Exploring Relationships between Design Features and System Usability of Intelligent car Human–Machine Interface." *Robotics and Autonomous Systems* 143:103829. <https://doi.org/10.1016/j.robot.2021.103829>.

- Yang, Y., C. Huang, L. Xia, and C. Huang. 2023. "Knowledge Graph Self-supervised Rationalization for Recommendation." In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, edited by K. Das, S. Raghavan, I. King, and B. Mobasher, 3046–3056. New York, NY: Association for Computing Machinery.
- Yang, Y., Q. Li, C. Li, and Q. Qin. 2024. "User Requirements Analysis of new Energy Vehicles Based on Improved Kano Model." *Energy* 309:133134. <https://doi.org/10.1016/j.energy.2024.133134>.
- Yuan, W., D. Chang, and T. Han. 2023. "A Context-Aware Smart Product-Service System Development Approach and Application Case." *Computers & Industrial Engineering* 183:109468. <https://doi.org/10.1016/j.cie.2023.109468>.
- Yuan, W., D. Chang, C. Mao, L. Wang, K. Ren, and T. Han. 2025. "A Novel User Scenario and Behavior Sequence Recognition Approach Based on Vision-Context Fusion Architecture." *Advanced Engineering Informatics* 65:103161. <https://doi.org/10.1016/j.aei.2025.103161>.
- Yuan, W., H. Zhao, X. Yang, T. Han, and D. Chang. 2024. "Toward Dynamic Rehabilitation Management: A Novel Smart Product-Service System Development Approach Based on Fine-Tuned Large Vision Model and Fuzzy-Dematel." *Advanced Engineering Informatics* 62:102616. <https://doi.org/10.1016/j.aei.2024.102616>.
- Zhang, M., L. Sun, Y. Li, G. A. Wang, and Z. He. 2023b. "Using Supplementary Reviews to Improve Customer Requirement Identification and Product Design Development." *Journal of Management Science and Engineering* 8 (4): 584–597. <https://doi.org/10.1016/j.jmse.2023.03.001>.
- Zhang, R., Y. Li, Y. Ma, M. Zhou, and L. Zou. 2023a. "Llm4aa: Making Large Language Models as Active Annotators." In *Findings of the Association for Computational Linguistics: EMNLP 2023*, edited by H. Bouamor, J. Pino, and K. Bali, 13088–13103. Singapore: Association for Computational Linguistics.
- Zhao, D., J. Gao, and N. Liu. 2023. "Exploring User Satisfaction and Improvement Priorities in Electric Vehicle Segments." *Transportation Research Part D: Transport and Environment* 125:103996. <https://doi.org/10.1016/j.trd.2023.103996>.
- Zhu, X., Y. Xie, Y. Liu, Y. Li, and W. Hu. 2025. "Knowledge Graph-Guided Retrieval Augmented Generation." In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, edited by A. Williams, J. Mueller, and S. Roller, 8912–8924. Albuquerque, New Mexico: Association for Computational Linguistics.

Appendix

Table A1. Formal definitions of key semantic constructs.

| Term | Definition |
|----------------------------|--|
| Semantic Tokens | Computable, reusable, and comparable linguistic units derived from raw user narratives through regex denoising, filtering, and tokenisation. They serve as the atomic input for semantic encoding, ensuring traceability from engineering inference back to original textual evidence. |
| Engineering Semantic Space | A multi-level, stable coordinate system consisting of K core engineering dimensions (e.g. Handling Performance, Interior Texture). It acts as a shared basis for both decoupling product performance assessment and modelling user preference heterogeneity. |
| Persona Vectors | A quantitative representation of user preference structures in the high-dimensional semantic space. It serves as a computed archetype acting as a reference anchor for a specific market segment's attention distribution. |
| Reasoning Paths | The logical traversal sequences within the Engineering Design Knowledge Graph (EDKG) connect User Personas, Reviews, Features, and Vehicle Entities. These paths support graph-constrained inference to generate evidence-backed and engineer-readable insight chains |

Table A2. Example preprocessing pipeline for a noisy review sample.

| Processing stage | Processing logic and method | Example data state |
|---------------------------------------|--|--|
| 0. Raw Input | Raw data collected by Octoparse | {User nickname} {URL link} {Vehicle model: Han EV} {Review content} ... 'Car picked up in mid-February, grey exterior with orange interior ... mainly attracted by the exterior and interior design ... driven for one month ... bought a paint protection film during the 3.15 promotion ... the only drawback is that the car is a bit large!!! scraped the car while making a U-turn yesterday, heartbreaking!' |
| 1. Feature Selection | Removal of irrelevant fields and retention of core attributes | {Vehicle model: Han EV} {Review content}'Car picked up in mid-February, grey exterior with orange interior ... (same as above)' |
| 2. Regex-based Denoising | Regular expression replacement to remove special characters and emojis; retaining Chinese/English characters, numbers, and basic punctuation | 'Car picked up in mid-February, grey exterior with orange interior ... mainly attracted by the exterior and interior design ... driven for one month ... bought a paint protection film during the 3.15 promotion ... the only drawback is that the car is a bit large!!! scraped the car while making a U-turn yesterday, heartbreaking! (Note: emoji removed) |
| 3. Short Review Filtering | Length threshold check ($Len \geq 10$) | (Length satisfied, retained) |
| 4. Tokenisation and Stop-word Removal | Jieba tokeniser (precise mode) + custom stop-word list (e.g. '的', '了', '就是', '虽然', '但是') | ['mid-February', 'car pickup', 'grey', 'exterior', 'orange', 'interior', ..., 'air-conditioning', 'stable', 'paint film', 'care', 'feels', 'good', 'drawback', 'car', 'a bit', 'large', 'U-turn', 'scrape', 'damage', 'heartbreaking'] |

Table A3. Data schema and sample records of the preprocessed corpus.

| Record ID | Storage path | Vehicle model | Raw review text | Cleaned review text | Token sequence |
|-----------|------------------------------------|---------------|---|---|---|
| 1 | ./data/processed/Tesla_Model_S.csv | Tesla Model S | [Exterior]Model S uses a coupe-style flat engine hood structure ... [Exterior]Stylish[Powertrain] ... | [Exterior] Model S adopts a coupe-style flat engine hood design ... [Exterior] Stylish[Powertrain] Three-motor system ... | [`exterior', `Model', `coupe-style', `flat', `hood', `structure', `design', `stylish', `powertrain', `three-motor', `performance', ...] |
| 2 | ./data/processed/BYD_Han_EV.csv | BYD Han EV | [Purchase Experience]I work in municipal maintenance ... [Purchase Price]288,600 + 20,000 (package) + 3,000 ... | [Purchase Experience] I work in municipal maintenance ... [Purchase Price] 288600 + 20000 (package) + 3000 (Aurora Blue exterior) ... | [`purchase', `experience', `municipal', `maintenance', `project', `price', `288600', `package', `Aurora', `blue', `exterior', ...] |
| 3 | ./data/processed/NIO_ET7.csv | NIO ET7 | [Range]Overall range is slightly lower ... [Most Satisfied]Recently test-drove the ET7 ... | [Range] Overall range is slightly lower, mainly because ... [Most Satisfied] Recently test-drove the ET7 ... | [`range', `overall', `slightly', `lower', `reason', `brand', `comparison', `test-drive', `ET7', `satisfaction', ...] |
| ... | ... | ... | ... | ... | ... |

Table A4. Example of raw feature tags.

| Cleaned comment (excerpt) | Vehicle model | Extracted feature dimensions | Status |
|---|---------------|--|---|
| 'The new car feels much more intelligent, especially functions such as assisted driving and parking. Besides standard perpendicular parking...' | Zhijie S7 | [{'raw_feature_tag': 'Intelligent Configuration', 'keywords': ['assisted driving', 'parking', 'system operation']}] | success |
| 'The exterior design is impressive, outperforming gasoline sedans such as the Audi A6L. Focusing on power and range: three driving modes are available...' | BYD Han EV | [{'raw_feature_tag': 'Exterior Design', 'keywords': ['exterior', 'outperform', 'Audi A6L']}, {'raw_feature_tag': 'Power Performance', 'keywords': ['power', 'eco mode', 'sport mode']}, {'raw_feature_tag': 'Range Performance', 'keywords': ['range', 'highway', 'urban roads']}] | success |
| 'Before purchasing the Zeekr 001, I was a loyal Volkswagen customer. After buying the 001 at the end of last year, it felt like entering a new driving...' | Zeekr 001 | [{'raw_feature_tag': 'Exterior Design', 'keywords': ['exterior', 'black', 'white']}, {'raw_feature_tag': 'Spatial Comfort', 'keywords': ['interior space', 'trunk', 'seating space']}, {'raw_feature_tag': 'Driving Experience', 'keywords': ['handling', 'ride quietness', 'acceleration performance']}] | success |
| ... | ... | ... | ... |
| 'The exterior and interior are the parts I am most satisfied with, and the price is also reasonable. It basically has all the expected features. The intelligent driving system can only be...' | BYD Qin L | [{'raw_feature_tag': 'Exterior Design', 'keywords': ['Exterior', 'Interior', 'Reasonable Price']}, {'raw_feature_tag': 'Intelligence', 'keywords': ['Intelligent Driving', 'Expressway', 'Highway']}, {'raw_feature_tag': 'Price & Cost', 'keywords': ['Range', 'City Driving', 'Fuel Consumption']}] | Failed: 'Price' incorrectly grouped under 'Exterior Design'. |
| 'I've waited for about a month and have only driven about 200 km in the city so far. Since I have no long-distance requirements, I didn't opt for the battery upgrade...' | Zeekr 007 | [{'raw_feature_tag': 'Interior Design', 'keywords': ['Interior', 'Comfortable', 'Budget']}, {'raw_feature_tag': 'Range Performance', 'keywords': ['Low Temperature', '13.2 Consumption', '17 Consumption']}, {'raw_feature_tag': 'Driving Experience', 'keywords': ['City Driving', '200km', 'Long-distance Need']}] | Failed: 'Budget' incorrectly grouped under 'Interior Design'. |

[System Instruction] You are a top-tier automotive market insight expert, specializing in identifying multiple, non-overlapping core user needs from single user reviews. Your task is to deeply analyze each review and extract all independent dimensions of concern.

[User Input Template] Please deeply analyze the following automotive user review and complete the tasks below:

1. Identify **ALL** core focus points mentioned in the review. Users may care about multiple aspects simultaneously (e.g., intelligence, power, space, exterior, price).
2. For **EACH** identified focus point: a. Summarize it into a concise, precise phrase (approx. 2-5 words) as the "Raw Feature Tag". b. Extract 2-4 "Core Keywords" from the original text that directly support this tag.
3. Ensure tags do not overlap; each tag must represent an independent dimension.

Review Text: *{review_text}*

Please output strictly in the following JSON format without extra explanation:

```
{
  "feature_dimensions": [
    {
      "raw_feature_tag": "Generated Tag 1",
      "keywords": ["Keyword 1", "Keyword 2", "Keyword 3"]
    },
    {
      "raw_feature_tag": "Generated Tag 2",
      "keywords": ["Keyword A", "Keyword B", "Keyword C"]
    }
  ]
}
```

Figure A1. Raw feature extraction prompt template.

[System Instruction] You are a senior automotive brand strategy expert, specializing in creating accurate and compelling Persona names for user groups. Your task is to generate a precise, concise, and commercially insightful Persona name for each user group based on clustering analysis results.

[User Input Template] Please generate a precise Persona name for the following automotive user group:

User Group Characteristics:

- Focus Tags: *{{report['tags']}}*
- High-Frequency Keywords: *{{report['top_keywords']}}*
- Sample Review Snippets: *{{report['sample_comments']}}*

Requirements:

1. The name must be **concise and impactful** (approx. 2-5 words).
2. It must reflect the core characteristics of the group.
3. It must possess **commercial value** and marketing significance.
4. It must be catchy and easy to remember.

Please output **strictly** in the following JSON format:

```
{{
  "persona_name": "Your Generated Persona Name",
  "rationale": "Brief explanation of the naming logic"
}}
```

Figure A2. Semantic normalisation prompt template.

```

cypher_prompt = PromptTemplate(
    input_variables=["schema", "question"],
    template="""
You are a Neo4j Cypher query expert. Based on the provided graph schema and the
user's question, generate an accurate Cypher query.

Graph Schema: {schema}

Important Instructions:
1. This is a knowledge graph for a New Energy Vehicle (NEV) recommendation system.
2. Key Node Types: CarModel, UserProfile, Review, Feature.
3. Key Relationships: PUBLISHED, MENTIONS, CONTAINS_ASPECT, INTERESTED_IN.
4. Please refer to the specific definitions in the schema for node properties.
5. The generated query must be efficient and accurate.
6. Avoid returning excessive data; use LIMIT appropriately.
7. Prioritize using indexed fields for querying.

User Question: {question}

Cypher Query:
"""
)

```

Figure A3. Text-to-Cypher generation prompt template.

```

analysis_prompt = ChatPromptTemplate.from_messages([
    ("system", """You are a query analysis expert for a new energy vehicle
recommendation system. Analyze the user's question and return the question type,
confidence score, and recommended search mode.

Question Types (Must choose one):
- statistics: Statistical queries (e.g., "How many types of EVs", "Total number of
models")
- car_recommendation: Vehicle recommendation (e.g., "Recommend a car", "Which one
is good")
- feature_comparison: Feature comparison (e.g., "Compare range", "Compare
performance")
- user_analysis: User analysis (e.g., "User preferences", "User persona")
- car_info: Vehicle information (e.g., "How is Tesla", "Vehicle parameters")
- general_search: General search (other questions)

Search Modes (Must choose one):
- cypher: Precise query, suitable for statistical questions
- graph: Knowledge graph search, suitable for recommendation, comparison, and
information retrieval
- vector: Vector search, suitable for semantic similarity search

IMPORTANT: Return ONLY the JSON format, without any other text. Example format:
{"question_type": "statistics", "confidence": 0.95, "search_mode": "cypher",
"reasoning": "This is a statistical question"}"""),
    ("user", "Please analyze this question: {question}")
])

```

Figure A4. Intent analysis and routing prompt template.

```
SYSTEM_PROMPT_TEMPLATE = """
You are a professional New Energy Vehicle (NEV) recommendation assistant. You have
access to the following knowledge:
- Rich vehicle feature data (exterior design, interior texture, intelligent
  configuration, spatial utility, comfort experience, handling performance, range &
  energy, value perception)
- Detailed user profile analysis
- Real user review data
- Knowledge graph capturing relationships among vehicle models

Please answer the user's question based strictly on the provided context.

Requirements:
1. Answers must be accurate, objective, and data-driven.
2. If providing recommendations, strictly consider the user's specific needs.
3. You may cite specific data and user reviews to support your points.
4. Maintain a friendly and professional tone.
5. If you are unsure of the answer, honestly state so.

Context:
{context}

User Question: {question}
"""
```

Figure A5. Insight generation system prompt template.