

ORIGINAL RESEARCH ARTICLE

AutoGen-Insight: Translating consumer reviews
into probabilistic visual design parameters for
generative electric vehicle concept generationLuyao Wang^{1,2} , Chun-Hsien Chen² , and Danni Chang^{1*} ¹Department of Design, School of Design, Shanghai Jiao Tong University, Shanghai, China²Design and Human Factors Lab, School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore

Abstract

Large-scale consumer reviews provide a valuable but weakly structured source of market insight for early-stage automotive design. However, existing generative design workflows often rely on direct prompting or manual interpretation, making it difficult to translate heterogeneous consumer preferences into inspectable visual design conditions. This paper presents AutoGen-Insight, a probabilistic semantic-to-visual parameter mapping framework for consumer-driven electric vehicle concept generation. Building on a retained corpus of 14,153 high-confidence electric vehicle reviews, the proposed framework reorganizes consumer feedback into four representative persona-level semantic preference profiles. These profiles are further transformed into probabilistic design vectors that encode both central tendencies and uncertainty across visually observable exterior concept descriptors, including vehicle proportion, styling, color, lighting, wheel design, and surface-finish parameters. A conflict-aware reasoning module integrates knowledge graph traversal with parameter-level trade-off aggregation to handle competing perceptual requirements, such as sportiness, comfort, compactness, and spatial utility. The resulting probabilistic design vectors are used to construct parameterized prompts for diffusion-based multi-view concept image generation, while the intermediate reasoning chains and parameter vectors remain available for inspection. The framework is evaluated through quantitative metrics, focused ablation analysis, qualitative conflict-resolution analysis, and expert/user assessment. Compared with direct prompting, keyword-based prompting, and large language model-generated prompting baselines, AutoGen-Insight achieves higher semantic alignment, plausibility of visual proportions, and cross-view consistency. The results suggest that probabilistic intermediate representations can improve the controllability and interpretability of early-stage image-based design exploration. This study contributes a structured and extensible approach for transforming large-scale consumer feedback into inspectable visual design knowledge for early-stage electric vehicle concept development.

***Corresponding author:**Danni Chang
(dchang1@sjtu.edu.cn)

Citation: Wang L, Chen C-H, Chang D. AutoGen-Insight: Translating consumer reviews into probabilistic visual design parameters for generative electric vehicle concept generation. *Int J AI Mater Design*. 2026;3(2):026190015. doi: 10.36922/IJAMD026190015

Received: May 10, 2026**Revised:** June 15, 2026**Accepted:** June 22, 2026**Published online:** June 30, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Keywords: Generative design; Electric vehicle design; Consumer insights; Knowledge graph; Diffusion models

1. Introduction

The global electric vehicle (EV) market is experiencing unprecedented growth, with design differentiation becoming a decisive factor in consumer purchase decisions. Unlike mechanical performance, which has reached a level of homogenization among mainstream competitors, visual design remains a primary brand identifier and value driver. Traditional automotive design processes, however, are characterized by long cycles, high costs, and heavy reliance on the subjective expertise of design teams. Market research inputs, typically gathered through focus groups or limited surveys, often lag and fail to capture the dynamic, nuanced preferences expressed in the vast volume of online consumer narratives.

In an era where generative artificial intelligence (AI) is redefining engineering processes and accelerating innovation cycles,^{1,2} this reliance on static, low-resolution data represents a significant competitive disadvantage. While the digital era offers unprecedented access to user feedback, its perceptual and weakly structured nature creates a formidable barrier: how to translate unstructured “voice-of-customer” (VOC) data into actionable, design-relevant representations for early-stage concept generation?

Concurrently, the advent of generative AI,³ particularly text-to-image diffusion models, has opened new avenues for conceptual design exploration.⁴ These models can rapidly produce high-fidelity visual concepts from textual prompts.⁵ However, their application in professional industrial design remains superficial. Directly feeding raw consumer feedback or simplified keywords into these models often yields results that are visually striking but weakly controlled, inconsistent with brand identity, or misaligned with specific market segments. The core challenge lies in the semantic gap: the lack of a systematic mechanism to transform unstructured, perceptual market insights into structured, visual parameter representations that can guide generative models in a traceable manner.

Current approaches to data-driven design face a critical disconnect. On one end, VOC methodologies excel at extracting thematic requirements but stop short of generating tangible design solutions.⁶ On the other end, generative AI tools offer powerful synthesis capabilities but lack the contextual grounding to respond precisely to complex market data.⁷ There is a pressing need for an integrated framework that bridges this gap, enabling an interpretable workflow from raw consumer data to early-stage visual concept generation.

Specifically, three technical challenges must be addressed:

(i) Semantic structuring: How to map weakly structured

natural language descriptions of user preferences into a structured, design-relevant semantic space that preserves heterogeneous consumer expectations?

(ii) Conflict resolution: How to infer actionable design parameters when user requirements are inherently contradictory (e.g., “sporty” versus “comfortable,” “compact” versus “spacious”)? Existing systems often ignore these conflicts or produce averaged, mediocre results.

(iii) Parameterized generation and evaluation: How to guide generative models with these parameters to ensure multi-view consistency, visual proportion plausibility, and alignment with the inferred insights? Furthermore, how to quantitatively measure this alignment beyond subjective surveys?

To address these challenges, this study proposes AutoGen-Insight, a consumer-driven framework for early-stage EV concept generation. Rather than directly using consumer reviews as prompts, the proposed framework introduces an intermediate visual parameter representation that transforms weakly structured semantic preferences into inspectable exterior concept descriptors. In this study, design parameters are visually observable concept-level descriptors, including vehicle proportion, styling, lighting, color, wheel design, and surface finish.

This study builds upon a previously established EV consumer review corpus.⁸ The earlier work focused on consumer-insight extraction, structured modeling, and knowledge-based retrieval. In contrast, the present study addresses a distinct downstream problem: how to reorganize retained high-confidence consumer reviews into persona-level semantic preferences and further transform them into probabilistic visual design parameters for diffusion-based concept image generation.

The main contributions of this work are as follows:

(i) Persona-level reorganization of consumer feedback. We reorganize a retained EV consumer review corpus into persona-level semantic preference profiles, enabling large-scale user feedback to serve as structured input for generative design.

(ii) Probabilistic semantic-to-visual parameter mapping. We introduce a probabilistic design vector (PDV) that converts semantic preferences into distributions over visually observable exterior concept descriptors, capturing both dominant preferences and uncertainty.

(iii) Conflict-aware design reasoning. We develop a reasoning mechanism that combines knowledge graph (KG) traversal and parameter-level trade-off aggregation to model competing perceptual requirements, such as sportiness, comfort, compactness, and spatial utility, as interpretable visual parameter trade-offs.

The remainder of this paper is organized as follows: Section 2 reviews related work. Section 3 details the system architecture and proposed algorithms. Section 4 describes the implementation. Section 5 presents the evaluation results. Section 6 discusses implications, and Section 7 concludes the study.

2. Related work

2.1. From statistical voice-of-customer analysis to Kansei-driven generative synthesis

The foundational challenge of translating user perception into design form has traditionally been addressed through VOC analysis and Kansei engineering (KE). Early data-driven approaches focused on extracting thematic requirements from text. Park *et al.*⁹ developed automated frameworks to adjust product strategies based on review trends, while Yu *et al.*¹⁰ utilized analytics to optimize product assortment. In the automotive domain, Tian *et al.*¹¹ employed deep learning to predict cabin odor perceptions, and Wang *et al.*¹² mapped user emotions to three-dimensional (3D) surface data. However, these methods predominantly yield statistical summaries. As noted by Briard *et al.*¹³, a significant challenge remains in translating such raw data into actionable 3D forms during early conceptual phases.

To address this, KE emerged to systematically map affective responses to design elements. Classical KE relies on statistical regression; Liu and Yang¹⁴ combined KE with virtual reality to quantify form–affect relationships, while Lian *et al.*¹⁵ proposed an integral KE framework for “affective-blue” design. Recent works have integrated machine learning to handle non-linearity; Jin *et al.*¹⁶ mined online reviews with a Kansei-integrated Kano model, and Bing *et al.*¹⁷ combined KE with generative adversarial networks (GANs) for shape innovation. Most notably, Yang *et al.*¹⁸ pioneered the integration of KE directly with diffusion models, enabling emotional adjectives to guide image synthesis.

Despite this evolution from VOC to generative Kansei-oriented methods, a semantic-to-visual parameter discontinuity persists. Existing methods typically map adjectives directly to image space or to fixed-shape coefficients, leaving limited room for inspectable intermediate representations. They lack a probabilistic visual parameter space in which fuzzy consumer semantics can be translated into controllable exterior concept descriptors, such as proportion tendency, stance, lighting emphasis, surface complexity, or color/finish variation. Consequently, current systems tend to produce static visual outputs rather than adaptable design-condition vectors that can support iterative early-stage concept exploration.

2.2. Conflict resolution: From rule-based knowledge graphs to Pareto optimization

Resolving conflicting requirements is a fundamental challenge in engineering design. Traditional approaches have leveraged KGs to detect and resolve contradictions. Huang *et al.*¹⁹ proposed a smart conflict-resolution model that uses multi-layer KGs for conceptual design, focusing on technical constraints. This aligns with broader surveys by Peng *et al.*²⁰ and Xue and Zou²¹ on KG quality and reasoning, while recent efforts have begun integrating large language models (LLMs) to enhance KG conflict detection.²² However, these KG-based methods primarily address hard technical constraints (e.g., material compatibility) using rule-based logic, struggling with soft perceptual conflicts (e.g., “aggressive” versus “safe”).

In parallel, multi-objective optimization (MOO) offers a mathematical framework for trade-offs. In physical design, Li *et al.*²³ used evolutionary algorithms to balance assembly, cost, and manufacturing constraints, generating uniform Pareto fronts. Marler and Arora²⁴ provided a foundational survey of MOO methods for engineering design, while Harkare *et al.*²⁵ and Rashed *et al.*²⁶ further summarized evolutionary approaches and algorithmic trends for solving complex multi-objective problems. In the EC domain, Dharma *et al.*²⁷ reviewed MOO algorithms for design and safety optimization, providing a more domain-relevant basis for discussing design trade-offs. Zhang *et al.*²⁸ explored gradient orthogonality for domain adaptation, and Zeng *et al.*²⁹ applied multi-agent coordination to molecular optimization.

A significant dichotomy exists between technical and perceptual conflict resolution. Current MOO applications focus mainly on hard engineering constraints and are rarely applied to subjective aesthetic or perceptual conflicts. Conversely, while AI-based conflict resolution shows promise, it remains largely textual or domain-specific. There is still limited research on how conflicting consumer sentiments can be represented as competing tendencies in visual parameters and resolved through interpretable parameter-level trade-off aggregation. This gap is particularly important for generative concept design, where unresolved conflicts may lead to averaged, incoherent, or one-sided visual outputs.

2.3. Controllable generation and the void in design-specific evaluation

The capability to steer generative models for industrial applications has improved dramatically. Early automotive applications by Lu *et al.*³⁰ and Wu *et al.*³¹ demonstrated the feasibility of using GANs and diffusion models for frontal form generation, though these relied heavily on manual

prompt engineering. Subsequent works explored multi-agent frameworks for aesthetic negotiation³² and the role of AI in design education.^{33, 34} To enhance controllability, Zheng *et al.*³⁵ introduced LayoutDiffusion for precise object placement, while Lian *et al.*³⁶ enhanced prompt adherence using LLM-grounded diffusion. Surveys by Jiang *et al.*³⁷ and Hartwig *et al.*³⁸ categorize these multimodal control techniques (e.g., ControlNet) as essential for industrial use.

Simultaneously, the field has sought robust objective metrics to replace subjective evaluation. Lin *et al.*³⁹ proposed evaluating generation quality via image-to-text consistency, surpassing traditional alignment metrics. Comprehensive surveys by Tian *et al.*⁴⁰ and Hartwig *et al.*³⁸ document dozens of metrics (Fréchet inception distance, contrastive language–image pretraining [CLIP] score), while recent studies have critically examined CLIP’s perceptual limitations and questioned whether CLIP-based similarity fully captures human visual judgment.^{41,42}

Despite progress in controllability and metrics, an evaluation gap remains in data-driven design. General-purpose metrics such as CLIP score measure broad text–image similarity, but they do not directly assess whether a generated concept visually reflects aggregated consumer preferences, satisfies basic visual proportion priors, or remains consistent across multiple views. Moreover, image-level structural conditioning methods improve spatial control, but they do not, by themselves, explain how weakly structured consumer reviews should be translated into inspectable design conditions. The present study, therefore, focuses on the upstream semantic-to-visual parameter transformation problem: converting consumer-derived persona preferences into probabilistic visual design vectors and parameterized prompts for early-stage EV concept generation.

3. System architecture and methodology

The proposed AutoGen-Insight framework is structured as a four-stage computational pipeline comprising data foundation, conflict-aware reasoning, semantic-to-visual parameter mapping, and parameterized prompt-based visual generation. The overall process transforms a retained consumer review corpus into persona-level semantic profiles, then into probabilistic visual design vectors, and finally into multi-view EV concept images via parameterized prompt construction.

The overall process can be summarized as:

$$R \rightarrow U \rightarrow S \rightarrow P \rightarrow T \rightarrow V \quad (1)$$

where R denotes the retained consumer review corpus, U denotes persona-level user profiles, S denotes semantic

preference representations, P denotes probabilistic visual design vectors, T denotes PDV-derived parameterized prompts, and V denotes generated multi-view concept images.

The framework is implemented as a modular pipeline (Figure 1): data foundation layer transforms unstructured reviews into structured attribute–sentiment representations; conflict-aware reasoning layer identifies competing perceptual requirements and resolves them into visual parameter trade-offs through KG traversal and parameter-level aggregation. Semantic-to-visual parameter mapping layer converts qualitative user preferences into probabilistic visual design vectors, and parameterized prompt-based visual generation layer translates these vectors into prompt-level generation conditions for diffusion-based multi-view concept image generation.

3.1. Layer 1: Data foundation

The foundation layer transforms unstructured consumer reviews into structured representations suitable for preference aggregation and downstream concept generation.

3.1.1. Comment–attribute–sentiment representation

The empirical data used in this study are drawn from a previously established EV consumer review corpus. As of July, 2025, the original crawler had collected 53,219 raw reviews from Dongche Score’s public praise pages, along with auxiliary attributes such as user level, likes, and replies. After preprocessing, including noise removal, short-text filtering, feature selection, and tokenization, 14,153 high-confidence reviews were retained for downstream analysis.

In this study, “high-confidence reviews” refer to retained reviews that contain sufficient textual content, at least one identifiable design- or usage-related attribute, and attribute-level sentiment information. Reviews consisting only of emojis, repeated slogans, extremely short comments, duplicated content, or unrelated transactional information were excluded.

The present study uses the retained corpus as the empirical basis for generative concept design, while the detailed review-mining procedure follows the previous corpus–construction study. A subset of retained and excluded reviews was manually inspected to verify whether the filtering rules removed irrelevant or low-information entries without eliminating design-relevant feedback.

For the present generative design task, these retained reviews were reorganized into an attribute–sentiment representation. Each review c_i is decomposed into a set of attribute-specific triplets:

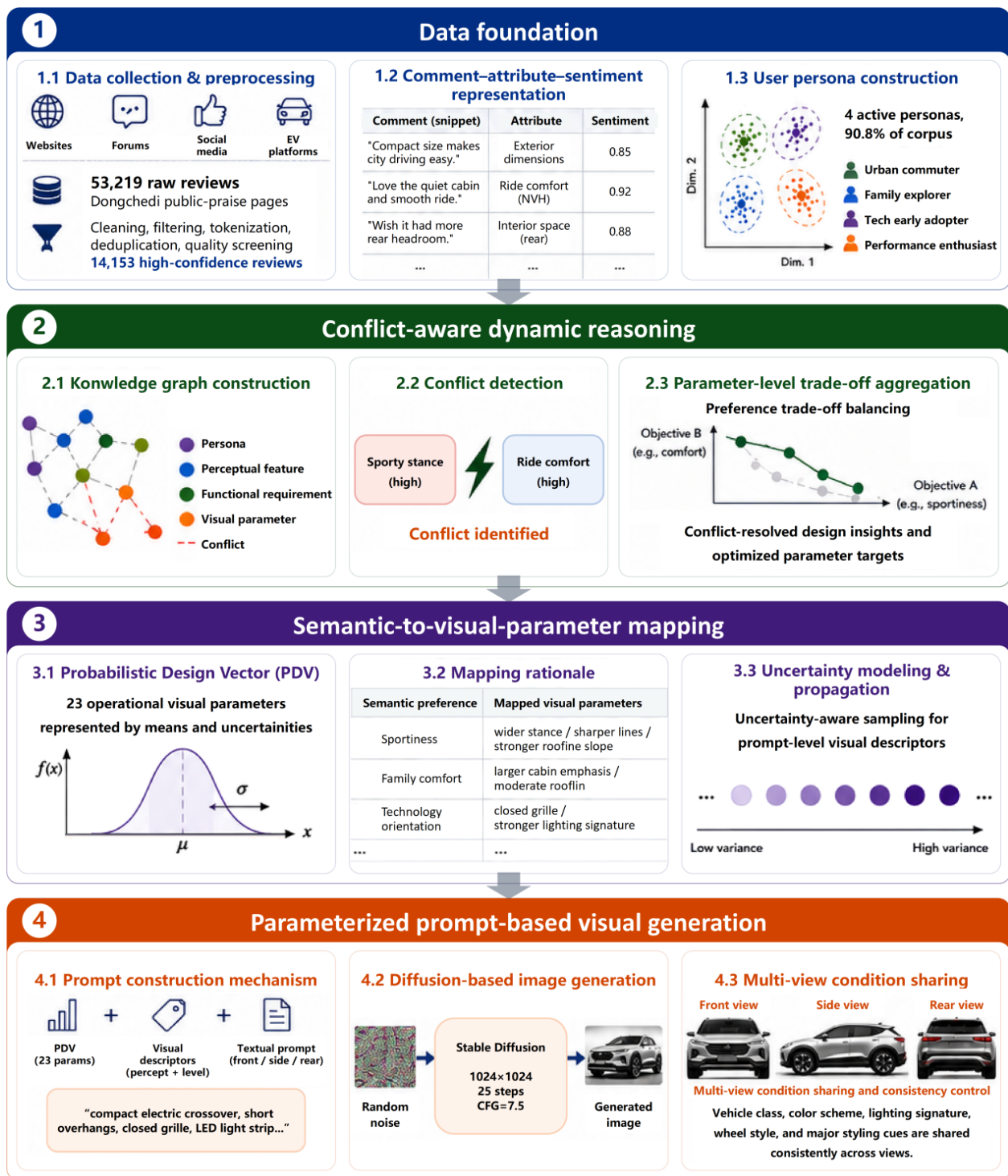


Figure 1. AutoGen-Insight architecture

$$c_{ij} = (t_{ij}, a_{ij}, S_{ij})_{j=1}^{n_i} \quad (2)$$

where t_{ij} denotes the text span associated with the j -th extracted expression, a_{ij} is the corresponding design-related attribute (e.g., “exterior”, “lighting”, “proportions”), and $s_{ij} \in [-1, 1]$ represents its sentiment polarity.

The review level attribute–sentiment matrix is defined as:

$$M_{ik} = \frac{1}{|J_{ik}|} \sum_{j \in J_{ik}} S_{ij}, J_{ik} = \{j | a_{ij} = k\} \quad (3)$$

where M_{ik} denotes the aggregated sentiment intensity of attribute k in review c_i , and J_{ik} denotes the set of extracted expressions in c_i associated with attribute k . If no expression in a review is associated with attribute k , M_{ik} is set to zero. This formulation converts weakly structured textual feedback into a continuous attribute–sentiment feature space.

Design-relevant semantic attributes were then grouped into higher-level preference dimensions, including exterior design, space utility, comfort, intelligence, performance, range, and value perception. These dimensions provide the basis for constructing persona-level semantic preference profiles. The attribute–sentiment matrix is used as an intermediate preference representation rather than as a direct image-generation prompt.

3.1.2. User persona construction

The current study requires persona profiles that are interpretable, design-relevant, and sufficiently stable to map to visually observable exterior concept descriptors. Therefore, user personas were constructed by clustering review-level feature representations using a Gaussian mixture model. The input feature space combines term frequency–inverse document frequency textual embeddings with sentiment-weighted attribute frequencies. The number of mixture components was determined by minimizing the Bayesian information criterion, while also considering interpretability and design relevance. The model was initialized using k -means ++ with diagonal covariance matrices.

The initial clustering process produced 12 preliminary persona clusters from the retained corpus of 14,153 high-confidence reviews. For the purpose of generative design evaluation, this study focuses on four high-frequency and design-representative personas: urban commuter, family explorer, tech early adopter, and performance enthusiast. These four personas cover 12,847 reviews, accounting for

90.8% of the retained corpus. The remaining 1,306 reviews were distributed across eight smaller clusters and retained only as auxiliary background nodes in the full reasoning graph.

In Table 1, values in parentheses indicate normalized persona-level attribute importance weights computed from sentiment-weighted attribute frequency and rescaled to (0,1). Priority attributes are defined as ($w \geq 0.8$), while secondary attributes are defined as ($0.6 \leq w < 0.8$).

Representative review excerpts and persona-level attribute summaries were inspected to verify whether the clusters reflected distinguishable design-relevant preference patterns. Together, these four personas account for 90.8% of the retained corpus. Because the present study focuses on downstream concept–generation evaluation, the four active personas are used as interpretable experimental conditions rather than definitive market segmentation categories, while broader demographic validation and stability testing of the persona taxonomy are left for future work.

3.2. Layer 2: Conflict-aware dynamic reasoning

This layer connects persona-level preferences with visual design parameters and identifies potential conflicts among competing preference signals. It is used to make the preference-to-parameter transformation traceable and to support parameter-level trade-off aggregation.

3.2.1. Knowledge graph construction

This study builds on the experience-grounded EV–KG established in prior work.⁸ The previous EV–KG organized user personas, vehicle models, review evidence, and evaluation features to support consumer-insight retrieval and hybrid graph-based reasoning. In contrast, the present study adapts this retrieval-oriented graph into a design-oriented reasoning graph to map from semantic to visual parameters and to support generative concept exploration.

The adapted graph is defined as $G = (V, E)$. The node set (V) includes four types of nodes: persona profiles, perceptual features, functional requirements, and design parameters. The edge set (E) includes four relationship types: preference, implication, realization, and conflict (Table 2).

The adaptation followed a hybrid procedure. Persona and perceptual-feature nodes were inherited from the prior EV–KG and its user-side preference-vectorization results. Review-derived evaluation features were reorganized into design-relevant perceptual features, such as sportiness, elegance, spaciousness, technological impression, comfort, and robustness. Functional requirement nodes

Table 1. Four representative user personas with priority attributes and cluster statistics

Persona	Priority attributes ($w \geq 0.8$)	Secondary ($0.6 \leq w < 0.8$)	Size	Percentage
Urban commuter	Compact dimensions (0.92), maneuverability (0.85)	Futuristic aesthetics (0.73), efficiency (0.68)	3,421	24.2%
Family explorer	Passenger space (0.95), cargo capacity (0.88), safety (0.91)	Comfort (0.79), quietness (0.71)	4,103	29.0%
Tech early adopter	Connectivity (0.94), autonomous driving (0.89)	Digital cockpit (0.86), over-the-air updates (0.77)	2,891	20.4%
Performance enthusiast	Acceleration (0.93), handling (0.88), sporty design (0.84)	Brand prestige (0.72), sound tuning (0.65)	2,432	17.2%

Table 2. Full design-oriented reasoning graph schema for semantic-to-visual parameter mapping

Component	Type	Symbol	Count	Example
Nodes	User persona	V_{persona}	12 (4 active)	“Urban commuter”
	Perceptual feature	V_{percept}	54	“Sporty,” “elegant”
	Functional requirement	V_{func}	87	“Aerodynamic efficiency”
	Design parameter	V_{param}	216	“Ground clearance,” “grille type”
Edges	Preference	Prefers	148	Persona \rightarrow feature
	Implication	Implies	203	Feature \rightarrow requirement
	Realization	Realized_by	267	Requirement \rightarrow parameter
	Conflict	Conflicts_with	34	Parameter \leftrightarrow parameter

were introduced to bridge perceptual features and visual design operations, including spatial utility, aerodynamic impression, perceived safety, entry/exit convenience, technology orientation, and driving experience. Visual parameter nodes were curated from exterior vehicle design descriptors and early-stage visual concept-generation controls. Conflict edges were added when two requirements implied incompatible or weakly overlapping tendencies for the same visual parameter, such as sportiness versus easy entry/exit or compactness versus cargo capacity.

The adapted graph contains 12 persona nodes, 54 perceptual feature nodes, 87 functional requirement nodes, and 216 visual parameter nodes. These 216 visual parameter nodes represent the broader design-reasoning space and are not all directly used for image generation.

Parameters related to cabin comfort, over-the-air services, battery management, or intelligent cockpit functions may influence semantic reasoning but cannot be directly expressed in a two-dimensional (2D) exterior concept-generation task. Therefore, the full graph is preserved for reasoning, while 23 visually observable parameters are selected later for the PDV.

In the main generation experiment, four representative personas are used as active persona nodes: urban commuter, family explorer, tech early adopter, and performance enthusiast. The remaining eight preliminary persona clusters are retained as auxiliary background nodes.

The resulting graph is therefore a hybrid, inspectable reasoning artifact: inherited from the prior EV-KG where possible, reorganized using corpus-level evidence,

and curated according to exterior design interpretability. Edge importance is computed by combining corpus-level co-occurrence evidence and sentiment intensity:

$$w_e = \lambda \cdot \text{freq}(e) + (1 - \lambda) \cdot |\text{sent}(e)| \quad (4)$$

where w_e is the weight of edge e , $\text{freq}(e)$ is the normalized co-occurrence frequency, $|\text{sent}(e)|$ is the average sentiment magnitude associated with the edge, and $\lambda \in (0,1]$ is a balancing coefficient. Both $\text{freq}(e)$ and $|\text{sent}(e)|$ are min-max normalized to $(0,1]$ within each edge type. In the current implementation, λ is set to 0.5 to assign equal importance to frequency evidence and sentiment magnitude. The same fixed value is used across all personas and generation conditions to ensure comparability.

The resulting graph should be interpreted as an inspectable design-reasoning structure rather than as an automatically learned causal model of automotive form perception. Its purpose is to make the transformation from review-derived semantic preferences to visual design parameters explicit and adjustable.

3.2.2. Conflict detection and parameter-level trade-off aggregation

Design conflicts arise when different user requirements impose incompatible or weakly overlapping tendencies on the same visual parameter. For example, a sporty appearance may favor a lower stance and a steeper roofline, while ease of entry/exit and family comfort may require a higher cabin profile and less aggressive roofline. Similarly, compact dimensions may conflict with cargo capacity or rear passenger space.

For a design parameter P_i , let $I_i^a = [l_i^a, u_i^a]$ and $I_i^b = [l_i^b, u_i^b]$ denote the preferred intervals imposed by two requirements, a and b . A conflict is detected when the overlap ratio falls below a predefined threshold η :

$$\rho_i(a,b) = \frac{\max(0, \min(u_i^a, u_i^b) - \max(l_i^a, l_i^b))}{\max(u_i^a, u_i^b) - \min(l_i^a, l_i^b) + \varepsilon} < \eta \quad (5)$$

where $\rho_i(a,b)$ denotes the normalized overlap ratio between the two preferred intervals, ε is a small constant used to avoid division by zero, and η is the threshold for identifying weakly compatible visual parameter tendencies.

Instead of averaging competing requirements, the method formulates them as a parameter-level trade-off aggregation problem for early-stage visual preference balancing. For each persona u_k , the satisfaction score is defined as:

$$S_k(P) = \sum_{i=1}^m w_{ik} \exp\left(-\frac{(P_i - u_{ik})^2}{2\sigma_{ik}^2 + \epsilon}\right) \quad (6)$$

where P denotes the visual parameter vector, P_i denotes the i -th visual parameter, w_{ik} is the importance weight of parameter i for persona u_k , u_{ik} is the preferred value, and σ_{ik} represents tolerance, and ϵ is a small constant used to avoid division by zero.

The global trade-off objective is then formulated as:

$$P^* = \arg \max_P \left[\sum_{k=1}^K \alpha_k \cdot S_k(P) - \beta \cdot \sum_{q=1}^Q \max(0, g_q(P))^2 \right] \quad (7)$$

where α_k denotes the normalized importance of persona k , $S_k(P)$ denotes the satisfaction score of persona k , β controls the strength of the penalty term, and $g_q(P)$ denotes violations of the q -th visual proportion prior. In this study, $g_q(P)$ include wheelbase-to-length ratio, height-to-width ratio, overhang ratio, wheel-to-body proportion, roofline slope, and stance width. The trade-off solution P^* is then used to construct the probabilistic representation in the next stage.

3.2.3. Insight chain representation

The output of this layer includes a traceable insight chain, a directed acyclic graph linking personas to design parameters with annotated trade-offs. Example chain:

[Urban Commuter] \rightarrow (prefers, 0.92) \rightarrow [Agility]
 \rightarrow (implies, 0.85) \rightarrow [Compact Body]

\searrow (conflicts, 0.45) \rightarrow [Cargo Space] \rightarrow (trade-off aggregation) \rightarrow [moderate roofline slope, short overhangs, compact crossover proportion]

This structure explicitly encodes the transformation from user personas to visual design parameters, including intermediate perceptual and functional nodes as well as conflict-resolution steps. Each edge is annotated with its semantic type and confidence weight, enabling traceability and post hoc inspection. The insight chain documents which preference signals were selected, where conflicts occurred, and which visual parameters were adjusted as a result.

3.3. Layer 3: Semantic-to-visual parameter mapping

Based on the persona-level preference profiles and the design-oriented reasoning graph, this layer constructs a PDV. The purpose of the PDV is to translate review-derived semantic preferences into a compact, visually controllable set of exterior concept descriptors for early-stage EV concept generation.

3.3.1. Probabilistic design vector

The PDV represents each generation-operational visual parameter as a probability distribution rather than a fixed value. Although the full reasoning graph contains 216 visual parameter nodes, only parameters that are visually observable and can be expressed through prompt-level exterior concept descriptors are included in the PDV. Therefore, 23 parameters related to vehicle proportions, exterior styling, color, lighting, wheel design, and surface finish are selected to construct the PDV (Table 3 and Figure 2). This distinction allows the framework to preserve a broader semantic reasoning space while remaining compatible with diffusion-based image generation via parameterized prompts. The PDV is defined as:

$$\text{PDV} = \left\{ (P_i, u_i, \sigma_i) \right\}_{i=1}^m, m = 23 \quad (8)$$

where P_i denotes the i -th operational visual parameter, u_i denotes its expected value, and σ_i denotes its uncertainty. The expected value of each parameter is computed as:

$$u_i = \sum_{k=1}^K \alpha_k u_{ik}, \sum_{k=1}^K \alpha_k = 1 \quad (9)$$

where α_k is the normalized weight of persona k , and u_{ik} is the preferred value of parameter i for that persona.

The variance is computed as:

$$\sigma_i^2 = \sum_{k=1}^K \alpha_k \left[\sigma_{ik}^2 + (u_{ik} - u_i)^2 \right] \quad (10)$$

This formulation captures both within-persona uncertainty and between-persona preference variation. As a result, the PDV does not merely encode a single dominant design direction, but also represents the diversity and ambiguity of consumer expectations (Table 3).

The domains in Table 3 were not derived from generated images. They were defined before image generation as visual design priors based on three sources: normalized measurements of representative production and concept EVs, automotive exterior design heuristics used in early-stage proportion studies, and expert review by automotive design practitioners. These ranges are therefore used as visual proportion plausibility priors rather than as engineering-feasibility constraints. Discrete styling categories were defined using common exterior-design descriptors. Default σ values define conservative sampling ranges for visual variation, while $\tau_{\text{uncertain}}$ controls when broader exploration is allowed.

3.3.2. Mapping rationale

To make the semantic-to-visual parameter transformation

interpretable, each semantic preference is linked to one or more operational visual parameters through an explicit mapping rationale. These mappings are treated as operational design priors rather than universal automotive design laws. Their role is to convert review-derived semantic preferences into inspectable generation conditions.

For example, sportiness is associated with a larger wheel-to-body impression, wider stance, sharper character lines, and a more pronounced roofline slope. Family comfort is associated with a higher cabin-to-body ratio, moderate roofline slope, and less extreme stance. Technology orientation is associated with a closed grille, continuous lighting signature, and cleaner surface organization. Premium perception is associated with restrained surface complexity, coherent color/finish selection, and balanced proportions.

For a given persona, the expected value of a visual parameter can be estimated through relation-based aggregation:

$$u_i = \sum_{a \in A} \tilde{w}_a r_{ai} v_{ai} \quad (11)$$

where A denotes the set of semantic attributes associated with the persona, \tilde{w}_a denotes the normalized importance of semantic attribute a , r_{ai} denotes the relation strength from semantic attribute a to visual parameter i , and v_{ai} denotes the preferred parameter tendency implied by attribute a .

Table 4 summarizes representative semantic-to-parameter mappings used in the current implementation.

Table 4 is not intended to claim a universal causal relationship between perception and vehicle form. Instead, it documents the system's operational design assumptions so that the mapping process can be inspected, adjusted, and evaluated. Accordingly, the mapping table functions as a transparent and revisable design-prior layer. It is intended to support traceable concept generation rather than to provide a universally valid empirical model of how consumers perceive automotive form.

3.3.3. Uncertainty modeling and propagation

Uncertainty is explicitly quantified as the relative variance of each visual parameter and is used to control the diversity of sampled prompt-level design conditions. Parameters exceeding a predefined uncertainty threshold are sampled over expanded ranges before being converted into textual descriptors. In this study, the sampling step is used only to obtain visual parameter values for prompt construction; it is not used as a learned latent conditioning mechanism inside the diffusion model.

Table 3. Probabilistic design vector parameter schema

Category	Parameter	Symbol	Domain	Default σ	$\tau_{\text{uncertain}}$
Proportions	Wheelbase/length	θ_1	(0.55, 0.65)	0.008	0.12
	Height/width	θ_2	(0.80, 0.95)	0.012	0.15
	Wheel diameter/height	θ_3	(0.35, 0.45)	0.006	0.10
	Front overhang/wheelbase	θ_4	(0.30, 0.40)	0.009	0.12
	Rear overhang/wheelbase	θ_5	(0.25, 0.38)	0.009	0.12
	Cabin/body ratio	θ_6	(0.42, 0.58)	0.015	0.15
	Ground-clearance/height	θ_7	(0.08, 0.18)	0.010	0.12
	Roofline slope	θ_8	(0, 1)	0.080	0.18
	Beltline height ratio	θ_9	(0.45, 0.65)	0.015	0.15
	Stance width ratio	θ_{10}	(0.70, 0.90)	0.020	0.15
Style	Line sharpness	s_1	(Soft, moderate, sharp)	–	0.20
	Surface complexity	s_2	(Simple, moderate, complex)	–	0.20
	Grille prominence	s_3	(Subtle, moderate, dominant)	–	0.25
	Lighting signature intensity	s_4	(Low, moderate, high)	–	0.20
	Headlamp slimness	s_5	(Broad, moderate, slim)	–	0.18
	Wheel design sportiness	s_6	(Low, moderate, high)	–	0.20
	Body contour muscularity	s_7	(Soft, balanced, muscular)	–	0.22
	Aerodynamic smoothness	s_8	(Angular, balanced, smooth)	–	0.20
Color/finish	Primary hue	c_1	(0°, 360°) hue–saturation–value	15°	0.18
	Saturation	c_2	(0, 1)	0.08	0.15
	Lightness	c_3	(0, 1)	0.06	0.12
	Accent contrast	c_4	(0, 1)	0.10	0.18
	Surface finish	c_5	(Matte, metallic, pearl, gloss)	–	0.20

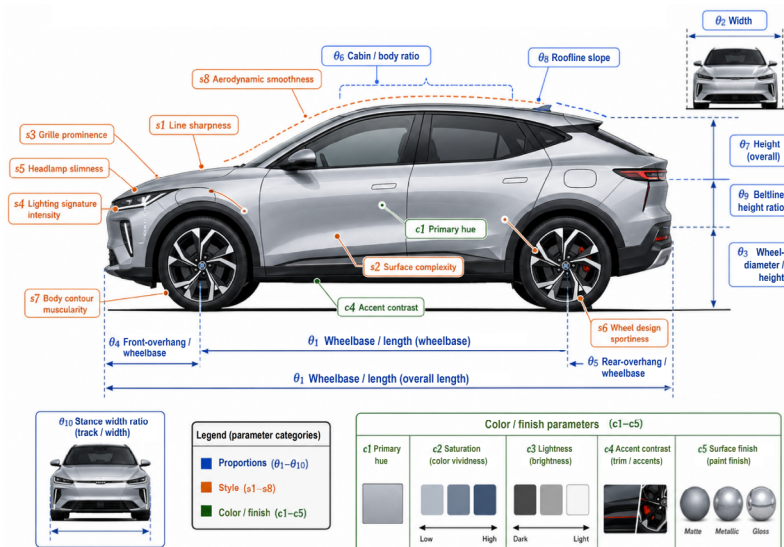

Figure 2. Electrical vehicle probabilistic design vector parameters

Table 4. Representative semantic-to-visual parameter mappings

Semantic preference	Affected design parameters	Mapping tendency	Design rationale
Sportiness	Wheel-diameter/height, stance width, roofline slope, line sharpness, wheel sportiness	Higher wheel emphasis, wider stance, sharper lines, stronger roofline slope	Sporty vehicle concepts are commonly perceived through a lower, wider, sharper, and more dynamic exterior profile
Family comfort	Cabin/body ratio, height/width, roofline slope, ground-clearance impression	Larger cabin emphasis, moderate height, less aggressive roofline	Family-oriented designs require stronger visual cues of space, accessibility, and comfort
Compact urban use	Front/rear overhang, wheelbase/length, cabin/body ratio	Shorter overhangs, compact proportions, efficient cabin packaging	Urban commuting emphasizes maneuverability, parking convenience, and efficient use of space
Technology orientation	Grille prominence, lighting intensity, headlamp slimness, surface complexity	Closed grille, stronger lighting signature, slimmer lamps, cleaner surfaces	Electric vehicle technology identity is often communicated through simplified front-end treatment and distinctive lighting
Premium perception	Surface complexity, color lightness, surface finish, accent contrast	More restrained surface treatment, coherent finish, controlled contrast	Premium appearance is often associated with visual restraint, material coherence, and balanced proportions
Robustness/safety	Beltline height, stance width, body contour, muscularity, and ground-clearance impression	Higher beltline, stable stance, stronger body volume	Robust and safe impressions are visually reinforced by stable proportions and stronger body mass
Efficiency	Aerodynamic smoothness, roofline slope, grille prominence, surface complexity	Smoother surfaces, reduced grille emphasis, streamlined profile	Efficiency is commonly associated with aerodynamic and visually clean exterior forms
Personalization	Primary hue, saturation, accent contrast, surface finish	Broader color and finish variation	Personalization-oriented users are better represented through controlled variation in color and material expression

For each visual parameter, sampled values are obtained as:

$$P_{\text{sampled},i} = \mu_i + \sigma_i \epsilon_i, \epsilon_i \sim N(0, I) \quad (12)$$

Sampled values are clipped to the predefined domain of each parameter:

$$\tilde{P}_i = \text{clip}(P_{\text{sampled},i}, L_i, U_i) \quad (13)$$

where L_i and U_i denote the lower and upper bounds of parameter (i), respectively.

Relative uncertainty is defined as:

$$u_i = \frac{\sigma_i}{U_i - L_i} \quad (14)$$

For parameters whose relative uncertainty exceeds the threshold $\tau_{\text{uncertain}}$, the sampling range is expanded as:

$$\sigma_i^{\text{expanded}} \begin{cases} \sigma_i, \sigma_i < \tau_{\text{uncertain}} \\ \gamma \sigma_i, u_i \geq \tau_{\text{uncertain}} \end{cases} \quad (15)$$

where $\gamma > 1$ controls the exploration strength in uncertain regions.

The expanded standard deviation is then used in the sampling step before conversion into textual descriptors. This uncertainty mechanism controls the parameter values used to construct prompt-level visual descriptors. Its contribution is evaluated through output-level metrics and ablation experiments rather than assumed from the formulation alone.

3.4. Layer 4: Parameterized prompt-based visual generation

The final layer synthesizes multi-view concept images using diffusion-based generation guided by PDV-derived parameterized prompts. The diffusion model receives text prompts, negative prompts, view-specific tokens, random seeds, and fixed inference settings. It does not receive PDV-derived structural maps as image-level conditioning inputs in the main experiment. Therefore, the generation control in this study operates at the prompt and descriptor level.

3.4.1. Prompt construction mechanism

The PDV is deterministically translated into textual

prompts using a template-based strategy. Continuous visual proportion parameters are converted into proportion-related descriptors, discrete styling parameters are converted into controlled adjective phrases, and color/finish parameters are converted into paint and material descriptors. This template-based strategy is used to make the prompt-construction process reproducible and inspectable, rather than relying on unconstrained LLM prompt generation.

The prompt structure integrates vehicle class, stylistic descriptors, visual features, environmental settings, and rendering quality (Table 5):

$$\text{Prompt} = T_{\text{subj}} \oplus T_{\text{style}} \oplus T_{\text{vis}} \oplus T_{\text{env}} \oplus T_{\text{quality}} \quad (16)$$

Unlike direct prompting or LLM-generated prompting, the prompt used by AutoGen-Insight is derived from the PDV. For example, for a PDV specifying compact crossover proportions, sharp line character, moderate sporty stance, a closed grille, continuous LED lighting, and a metallic silver/red color scheme, the generated prompt is:

“Professional automotive photography of a compact electric crossover, short overhangs, balanced wheelbase-to-length proportion, moderately sporty stance, sharp character lines, closed front grille with continuous LED light strip, multi-spoke alloy wheels, metallic silver paint with subtle red accents, side profile view, studio lighting, highly detailed, photorealistic.”

Representative prompts for the three baselines and the proposed method are reported in the experimental section to enable a transparent comparison.

3.4.2. Probabilistic design vector-derived visual descriptor construction

In addition to persona-level textual descriptors, the PDV is converted into a set of visual descriptors that guide prompt construction. These descriptors represent normalized exterior layout tendencies, including compactness, stance width, roofline slope, overhang impression, wheel-to-body emphasis, beltline height, lighting prominence, and surface complexity. In the current implementation, these descriptors are not used as image-level conditioning maps. Instead, they are converted into controlled textual phrases and used to construct view-specific prompts. Schematic previews of these descriptors may be used for inspection and explanation, but they are not supplied to the diffusion model as ControlNet-style structural inputs in the main experiment.

3.4.3. Multi-view condition sharing

Because the framework generates front, side, and rear views of the same vehicle concept, the same PDV is shared across all views. View-specific prompt tokens are adjusted according to the required viewpoint, while the underlying vehicle class, proportion descriptors, color scheme, lighting signature, wheel style, and major styling cues remain fixed. This shared prompt-level conditioning strategy is intended to reduce the likelihood that different views represent unrelated vehicle concepts. Its effect is evaluated empirically through cross-view visual consistency metrics and blind human assessment.

The current study allows designers to inspect persona profiles, reasoning chains, PDV, generated prompts, PDV-derived visual descriptors, and output images. However, adaptive learning from designer feedback is not evaluated in the present experiment and is therefore not presented as a core technical contribution. Future work may investigate how designer corrections can be systematically incorporated into the relation graph or parameter-mapping rules.

Together, these four layers provide a traceable workflow from consumer feedback to the controllable generation of visual concepts. The key methodological contribution lies in the intermediate translation from persona-level semantic preferences to probabilistic visual parameters, rather than in the claim of a fully automated end-to-end vehicle design system.

4. System implementation

To validate AutoGen-Insight, we implemented an experimental prototype that connects persona-level preference representation, PDV construction, prompt generation, and diffusion-based multi-view image synthesis. The prototype was developed for controlled experimental comparison rather than as a commercial automotive design platform. Its primary purpose was to record and reproduce how consumer-review-derived preferences were transformed into generated concept images under consistent generation settings.

4.1. Prototype overview

The prototype consists of three functional components: a lightweight inspection interface, a reasoning and parameter-mapping backend, and a graphics processing unit (GPU)-based image-generation service. The interface supports persona selection, reasoning-chain inspection, PDV display, prompt inspection, PDV-derived visual descriptor preview, and generated concept browsing. The backend performs KG traversal, conflict detection, parameter-level trade-off aggregation, PDV

Table 5. Prompt construction rules from probabilistic design vector (PDV) parameters

Component	Template	PDV mapping rule
T_{subj}	“(Vehicle class) electric vehicle”	$\theta_1, \theta_2 \rightarrow$ vehicle classification (compact SUV/Sedan, etc.)
T_{style}	“(Line_char) character lines, (stance) stance”	Style tokens \rightarrow adjective selection
T_{vis}	“Closed grille with (lighting_type), (wheel_size)-inch wheels”	Discrete parameters \rightarrow specific feature descriptions
T_{env}	“(View_angle) view, (lighting_setup)”	Fixed as “side/front/rear profile, studio lighting”
T_{quality}	“Automotive photography, 8k, highly detailed”	Fixed quality modifiers

construction, and deterministic prompt generation. The image-generation service receives the generated prompts, negative prompts, view-specific tokens, random seeds, and fixed diffusion settings, then produces front, side, and rear concept images.

The prototype was implemented as a client-server system. The backend was implemented in Python and used standard graph-processing and numerical-computation libraries for relation traversal, conflict detection, and parameter aggregation. The image-generation service was deployed on a GPU-enabled server. These deployment details are reported for reproducibility and are not treated as independent methodological contributions.

The interface was used only for experimental inspection and record-keeping. It allowed researchers to view persona profiles, reasoning paths, PDVs, generated prompts, PDV-derived visual descriptors, and multi-view outputs in a single workspace. If included in the manuscript, the interface figure should be placed here as [Figure 3](#) and captioned as follows:

The interface displays persona-level preferences, reasoning chains, PDVs, generated prompts, PDV-derived visual descriptors, and multi-view concept outputs. It was used for inspection and logging rather than as an evaluated design contribution.

4.2. Generation workflow

The generation workflow follows the four methodological layers described in Section 3.

First, the selected persona profile is converted into a weighted semantic preference representation. The system retrieves relevant perceptual features, functional requirements, and visual design parameters from the design-oriented KG. Each reasoning path is assigned a confidence score based on edge weights and the importance

of persona-specific attributes.

Second, potential conflicts among design requirements are detected at the parameter level. When overlapping preferred parameter intervals fall below the predefined threshold, the system formulates the problem as a parameter-level trade-off aggregation task. The resulting trade-off solution is then used to initialize the PDV.

Third, the PDV is translated into prompt-level generation conditions. Continuous visual proportion parameters, such as wheelbase-to-length ratio, height-to-width ratio, roofline slope, front and rear overhang, stance width, and wheel-to-body proportion, are converted into proportion-related textual descriptors. Discrete styling parameters, such as line sharpness, grille prominence, lighting signature intensity, surface complexity, and wheel design sportiness, are converted into controlled adjective phrases. Color and finish parameters are converted into paint, material, and accent descriptors.

Fourth, the generation service synthesizes multi-view vehicle concepts under shared parameterized prompt conditions. The same PDV is used across front, side, and rear views, while view-specific prompt tokens are adjusted according to the required viewpoint. This workflow clarifies that the diffusion model does not directly consume the KG, reasoning chains, probabilistic vectors, or structural maps. Instead, the graph and conflict-resolution steps determine the PDV, which in turn determines the textual descriptors used in prompt construction.

4.3. Diffusion-based generation settings

Visual synthesis was performed using a latent diffusion model based on the stable diffusion XL architecture. The model operated at a resolution of 1024×1024 pixels. DPM-Solver++ was used as the sampler, with 25 inference steps and a classifier-free guidance scale of 7.5. The same diffusion backbone, image resolution, sampling steps,

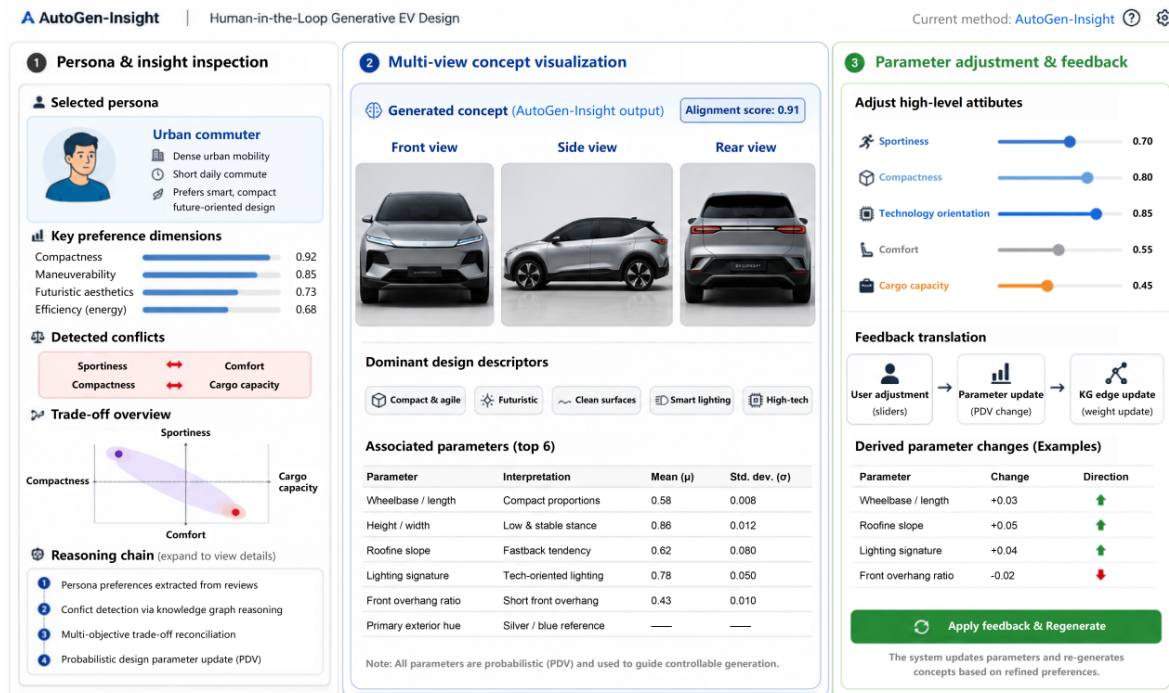


Figure 3. Prototype interface of AutoGen-Insight for experimental inspection
Abbreviations: EV: Electric vehicle; KG: Knowledge graph.

guidance scale, and negative prompt were used across all methods.

The positive prompt for AutoGen-Insight was generated deterministically from the PDV using the template described in Section 3.4. The negative prompt was fixed as:

“Worst quality, low quality, distorted proportions, unrealistic, deformed wheels, asymmetric, blurry, text, watermark, signature.”

No brand-specific fine-tuning, low-rank adaptation, additional style reference data, or image-level structural conditioning module was used in the main experiment. This setting was adopted to avoid introducing external style information or additional control mechanisms that could confound the comparison between methods.

For the proposed method, front, side, and rear views were generated under shared PDV-derived prompt conditions. The same PDV was used across all views, while the view token was adjusted according to the target viewpoint. For prompt-based baselines, the same view-specific tokens and inference settings were used, but no PDV-derived parameterization was applied.

Probabilistic design vector-derived schematic previews

may be used for inspection and explanation, but they are not supplied to the diffusion model as image-level conditioning inputs in the main experiment. These previews represent normalized exterior layout tendencies, such as roofline slope, stance width, overhang impression, wheel-to-body emphasis, and lighting placement. They are included only to help designers assess how the PDV aligns with visual concept descriptors.

5. Evaluation and results

5.1. Experimental setup

We evaluated AutoGen-Insight using four representative persona profiles derived from the retained corpus: urban commuter, family explorer, tech early adopter, and performance enthusiast. These personas cover 12,847 reviews, accounting for 90.8% of the 14,153 high-confidence reviews. The remaining reviews belonged to smaller preliminary clusters and were not used as primary experimental conditions to keep the generational comparison focused and interpretable.

For each persona, design concepts were generated using the same diffusion backbone and identical inference settings, including image resolution, sampling steps, guidance scale, and negative prompt. For each method and

each persona, 10 design concepts were generated, resulting in 160 generated samples across all comparison methods. To improve comparability, random seeds were fixed where applicable, and all generated outputs were anonymized before evaluation.

We compared AutoGen-Insight with three baseline methods:

- (i) Baseline A (direct prompt): Raw review summaries were directly used as input prompts for the diffusion model.
- (ii) Baseline B (keyword extraction): High-frequency keywords extracted from reviews were used to construct prompts.
- (iii) Baseline C (LLM-generated prompt): An LLM was used to summarize consumer feedback and generate structured prompts.
- (iv) AutoGen-Insight: The proposed framework incorporates persona-level preference aggregation, conflict-aware reasoning, probabilistic semantic-to-visual parameter mapping, and PDV-derived parameterized prompting.

All methods used the same diffusion backbone and generation settings. The difference among methods lies in how consumer review information was transformed into prompts for generation. AutoGen-Insight does not use image-level structural conditioning maps in the main experiment; its control mechanism operates through PDV-derived textual descriptors and shared prompt-level generation conditions.

For each generated output, the prototype recorded the selected persona, parameter vector, generated prompt, negative prompt, random seed, sampling settings, and

view type. These logs were used to support the quantitative evaluation of semantic alignment, visual proportion plausibility, and multi-view consistency.

To make the comparison between baselines and AutoGen-Insight transparent, Table 6 reports representative prompt examples for the urban commuter persona.

These prompt examples show that AutoGen-Insight differs from LLM-generated prompting not by using a different diffusion backbone, but by constructing prompts from an intermediate PDV. This makes the generation conditions more traceable and allows the connection between consumer preferences, visual parameters, and generated concept images to be inspected.

Participants evaluated anonymized and randomized image sets covering the four personas and four generation methods. Method labels were hidden during evaluation. Professional participants were required to have training or experience in automotive, transportation, industrial design, or product design. Consumer participants were screened for EV interest, EV ownership experience, or purchase intention.

5.2. Evaluation metrics

To evaluate the generated concepts, we used three metrics corresponding to the main objectives of this study: semantic alignment, visual proportion plausibility, and multi-view consistency.

5.2.1. Semantic alignment

Semantic alignment measures the degree to which generated designs reflect the underlying consumer preferences. It is computed as the cosine similarity between

Table 6. Representative prompt examples for the urban commuter persona

Method	Prompt construction strategy	Example prompt
Direct prompt	The review summary is directly used as a prompt	"Generate an electric vehicle (EV) concept for consumers who care about compact size, daily commuting, efficiency, and futuristic appearance."
Keyword extraction	High-frequency keywords extracted from reviews	"Compact EV, urban commuting, maneuverable, efficient, futuristic, LED lighting, practical."
Large language model (LLM)-generated prompt	LLM summarizes consumer feedback into a structured design prompt	"Design a compact electric crossover for urban commuters who value easy driving, efficient daily use, futuristic styling, and practical interior space."
AutoGen-Insight	Probabilistic design vectors-derived prompt with parameterized descriptors	"Professional automotive photography of a compact electric crossover, short overhangs, balanced wheelbase-to-length proportion, moderately sporty stance, sharp character lines, closed front grille with continuous LED light strip, multi-spoke alloy wheels, metallic silver paint with subtle red accents, side profile view, studio lighting, highly detailed, photorealistic."

the CLIP text embedding of the aggregated persona-level review representation and the CLIP image embedding of the generated design:

$$S_{align} = \frac{E_{text}(R_k) \cdot E_{img}(I)}{E_{text}(R_k) E_{img}(I)} \quad (17)$$

where $E_{text}(R_k)$ denotes the text embedding of the aggregated review representation for persona k , and $E_{img}(I)$ denotes the image embedding of the generated design image I . A higher score indicates stronger alignment between generated design outputs and consumer preference signals.

Contrastive language-image pretraining similarity is used here as a scalable proxy for global text-image alignment, not as a complete measure of automotive design quality. To reduce over-reliance on this single proxy, it is complemented by visual plausibility of proportions, multi-view consistency, and blind human evaluation.

5.2.2. Visual proportion plausibility

To evaluate whether the generated designs satisfy basic automotive visual proportion priors, we define a visual proportion plausibility score. The assessed priors include wheelbase-to-length ratio, height-to-width ratio, wheel-to-body proportion, front-overhang ratio, roofline slope, stance width, and vehicle-class consistency. We define a visual proportion plausibility score:

$$S_{pp} = \frac{1}{N} \sum_{i=1}^N I(L_i \leq \hat{P}_i \leq U_i), I(*) = \begin{cases} 1, & \text{if } x \text{ is true} \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

where \hat{P}_i denotes the inferred value of the i -th visual parameter from generated image I , L_i and U_i denote the lower and upper bounds of the corresponding visual-prior interval, N denotes the number of assessed visual parameters, and $I(*)$ is an indicator function.

The parameter values were estimated from generated images using a semi-automatic visual measurement procedure based on vehicle bounding boxes, wheel positions, silhouette height/width, roofline control points, and visible overhang proportions. The resulting values were normalized before comparison with the predefined visual-prior intervals. This metric should be interpreted as a 2D visual proportion plausibility score rather than as evidence of engineering feasibility.

5.2.3. Multi-view consistency

Because the framework generates front, side, and rear views of the same vehicle concept, cross-view consistency is critical. Multi-view consistency is computed using a

combination of global image similarity and attribute-level consistency checks:

$$S_{mvc} = \omega_1 S_{img} + \omega_2 S_{attr}, \omega_1 + \omega_2 = 1 \quad (19)$$

where S_{img} denotes global visual similarity across views, S_{attr} denotes attribute-level consistency, and ω_1 and ω_2 are weighting coefficients. In this study, both terms were weighted equally unless otherwise specified.

Global visual similarity is computed using pairwise CLIP image-embedding similarity:

$$S_{img} = \frac{1}{|V|} \sum_{(p,q) \in V} \frac{E_{img}(I_p) \cdot E_{img}(I_q)}{\|E_{img}(I_p)\| \|E_{img}(I_q)\|} \quad (20)$$

where V denotes the set of view pairs among the front, side, and rear views, and $E_{img}(\cdot)$ denotes the CLIP image encoder.

Attribute-level consistency is computed using a checklist covering vehicle class, dominant color scheme, lighting signature, wheel style, stance, and major styling cues:

$$S_{attr} = \frac{1}{M} \sum_{j=1}^M I[A_j(I_{front}) + A_j(I_{side}) + A_j(I_{rear})] \quad (21)$$

where A_j denotes the j -th visual attribute. M denotes the number of assessed attributes, and $I[\cdot]$ denotes the indicator function. Dominant color consistency was assessed using color histogram comparison; body-proportion consistency was estimated through semi-automatic measurement; vehicle class and major styling cues were checked using a combination of zero-shot image classification and human verification.

This revised formulation avoids relying solely on CLIP to compare fine-grained automotive attributes. CLIP is used for global image-level similarity, while color, proportion, vehicle class, and styling-cue consistency are assessed through attribute-specific procedures.

5.3. Quantitative evaluation with baselines

Table 7 presents the quantitative comparison between AutoGen-Insight and the baseline methods. Values are reported as mean \pm standard deviation across generated samples.

AutoGen-Insight achieved the highest mean score across all three metrics. Compared with direct prompt, the proposed method improved semantic alignment from 0.42 to 0.78. It also outperformed the LLM-generated prompt baseline, suggesting that structured semantic-to-visual

parameter reasoning provides additional value beyond natural language prompt refinement alone.

The proposed method also achieved higher scores for visual proportion plausibility and multi-view consistency. These results suggest that the PDV and PDV-derived parameterized prompts help guide image generation toward more visually coherent automotive concepts. However, these scores should be interpreted as evaluation proxies for early-stage concept images rather than as evidence of production-level design feasibility.

To assess statistical significance, we conducted a one-way analysis of variance followed by Tukey's post hoc tests, with method type as the independent variable and metric score as the dependent variable. Statistical testing was used to support metric-level differences rather than as definitive proof of design superiority. Overall, AutoGen-Insight showed higher mean scores than all baseline methods across semantic alignment, visual proportion plausibility, and multi-view consistency.

5.4. Focused ablation study

To examine whether the main intermediate representations contributed measurable value, we conducted a focused ablation study on three core components: persona-level preference aggregation, conflict-aware reasoning, and PDV parameterization. This ablation was designed as a component-level diagnostic rather than an exhaustive decomposition of the full pipeline.

Each variant removed one component while keeping the diffusion backbone, prompt template structure, negative prompt, image resolution, sampling steps, guidance scale, and random seed policy unchanged. The same four representative personas were used. For each variant, five concepts were generated per persona, resulting in 20 generated concepts per ablation condition. The results are summarized in Table 8. Values are reported as mean \pm standard deviation across generated samples.

The ablation results show a consistent decline when each major component is removed. Removing persona-level aggregation mainly reduced semantic alignment, suggesting that persona-specific profiles provide more targeted guidance than corpus-level averaging. Removing conflict-aware reasoning primarily affected the plausibility of visual proportions, indicating that parameter-level trade-off aggregation helps handle competing requirements such as sportiness, comfort, compactness, and spatial utility. Removing PDV parameterization produced the largest overall decline, especially in visual proportion plausibility and multi-view consistency, suggesting that explicit parameter-derived descriptors are important for controllable generation.

These results indicate that persona aggregation, conflict-aware reasoning, and PDV parameterization contribute to different evaluation dimensions: preference alignment, conflict-sensitive proportion control, and cross-view visual coherence. Because this was a focused, small-scale ablation, the findings should be interpreted as component-level supporting evidence rather than a complete causal analysis of the full pipeline.

5.5. Expert and user evaluation

To complement the quantitative metrics, we conducted a blind evaluation with 25 participants, including 10 professional designers and 15 potential EV consumers. Each participant evaluated anonymized multi-view outputs across the four methods and four representative personas using a five-point Likert scale. The order of samples was randomized to reduce ordering bias. This produced 400 method-persona ratings for each evaluation criterion.

Professional participants were required to have training or experience in automotive, transportation, industrial design, or product design. Consumer participants were screened for EV interest, EV ownership experience, or purchase intention. Method labels were hidden during evaluation, and participants were asked to focus on perceived persona alignment, visual concept plausibility, and cross-view consistency.

The evaluation focused on three criteria:

- (i) Market alignment: whether the design reflects the target person's preferences.
- (ii) Visual design plausibility: whether the design appears plausible as an early-stage automotive concept.
- (iii) Cross-view consistency: whether the front, side, and rear views appear to represent the same vehicle.

Table 9 summarizes the mean participant ratings across the three evaluation criteria. Values are reported as mean \pm standard deviation across participant ratings.

AutoGen-Insight received the highest mean ratings across all three criteria. Designers noted that the generated outputs better reflected realistic product positioning and provided more interpretable trade-offs between competing requirements. Potential consumers also rated the proposed method higher in market alignment, suggesting that the generated concepts more effectively captured persona-level preference signals.

Inter-rater reliability was assessed using Cronbach's alpha (α), yielding ($\alpha = 0.82$), indicating good internal consistency among evaluators. Pairwise comparisons showed that AutoGen-Insight received higher ratings than the strongest baseline across the three criteria. These human evaluation results should be interpreted

Table 7. Quantitative comparison of generated design outputs

Method	Semantic alignment (S_{align})	Visual proportion plausibility (S_{vpp})	Multi-view consistency (S_{mvc})
Baseline A-direct prompt	0.42 ± 0.08	0.61 ± 0.09	0.58 ± 0.10
Baseline B-keyword extraction	0.51 ± 0.07	0.66 ± 0.08	0.62 ± 0.09
Baseline C-large language model-generated prompt	0.59 ± 0.06	0.71 ± 0.07	0.67 ± 0.08
AutoGen-Insight	0.78 ± 0.04	0.87 ± 0.05	0.84 ± 0.06

Table 8. Focused ablation variants and results

Variant	Removed component	S_{align}	S_{vpp}	S_{mvc}
Full AutoGen-Insight	None	0.78 ± 0.04	0.87 ± 0.05	0.84 ± 0.06
Without persona aggregation	Corpus-level preference aggregation instead of persona-specific profiles	0.69 ± 0.06	0.82 ± 0.06	0.80 ± 0.07
Without conflict-aware reasoning	Direct semantic-to-parameter mapping without knowledge graph traversal or trade-off aggregation	0.74 ± 0.05	0.78 ± 0.07	0.79 ± 0.07
Without probabilistic design vector (PDV) parameterization	Natural-language persona prompts without explicit PDV-derived descriptors	0.63 ± 0.07	0.72 ± 0.08	0.68 ± 0.09

Table 9. Expert and user evaluation results

Method	Market alignment	Visual design plausibility	Cross-view consistency
Baseline A-direct prompt	3.1 ± 0.7	2.8 ± 0.8	2.9 ± 0.8
Baseline B-keyword extraction	3.3 ± 0.6	3.0 ± 0.7	3.1 ± 0.7
Baseline C-large language model-generated prompt	3.7 ± 0.5	3.4 ± 0.6	3.5 ± 0.6
AutoGen-Insight	4.6 ± 0.4	4.3 ± 0.5	4.5 ± 0.4

as complementary perceptual evidence rather than as a substitute for large-scale consumer validation. Given the limited participant sample size, the human evaluation was not used as a definitive inferential test of market preference. Effect-size estimation, broader demographic sampling, and correction for multiple comparisons in larger repeated-measures studies are left for future work.

5.6. Qualitative analysis of conflict resolution

To further examine how the proposed framework handles conflicting design requirements, we analyzed a representative case from the urban commuter persona. Consumer feedback for this persona showed a simultaneous preference for a sporty appearance and ease of entry/exit. These two requirements impose partially conflicting visual parameter tendencies: a sporty appearance typically favors a lower, wider stance, whereas easy entry and exit often require a higher roofline and less aggressive cabin profile.

The direct prompt baseline generated a low-slung sports-car-like concept, emphasizing sportiness while weakening accessibility. The keyword extraction baseline produced a taller, more practical vehicle, but the sporty character was weakened. The LLM-generated prompt baseline produced a visually appealing crossover, but its proportions and styling cues were less consistent across views.

In contrast, AutoGen-Insight generated a compact crossover-coupe-like concept with a moderately sloping roofline, short overhangs, balanced wheel-to-body proportion, and a stable lighting signature across views. This result shows that the conflict-aware parameter aggregation step can identify a compromise direction that partially satisfies both requirements rather than collapsing them into an averaged or one-sided design.

Figure 4 illustrates the comparative generation results for the urban commuter persona across the three baselines and the proposed method. The schematic visual descriptors shown in Figure 4 should be interpreted as explanatory previews of PDV-derived proportional tendencies. They were not used as additional image-conditioning inputs in the main diffusion generation experiment.

6. Discussion

6.1. Findings and contributions

The results suggest that AutoGen-Insight improves semantic alignment, visual proportion plausibility, and multi-view consistency compared with direct prompting, keyword-based prompting, and LLM-generated prompting baselines. These improvements indicate that consumer preferences are more effectively preserved when they are

transformed into structured intermediate representations rather than directly compressed into textual prompts.

Compared with the previous consumer-insight modeling study, the present work shifts the focus from insight extraction and retrieval to generative design transformation. The contribution of this paper, therefore, lies not in rebuilding the review-mining pipeline but in demonstrating how retained consumer review data can be reorganized into persona-level semantic preferences and further translated into probabilistic visual parameter representations for prompt-based concept generation.

The proposed semantic-to-visual parameter mapping provides an interpretable bridge between natural language consumer feedback and image-based generative models. By modeling visual design parameters as probability distributions, the framework captures both central tendencies and uncertainty in user preferences. This probabilistic formulation is particularly important because consumer expectations are rarely homogeneous or deterministic. Different user groups may express partially overlapping or even conflicting preferences, which cannot be adequately represented by a single fixed prompt.

The conflict-aware reasoning mechanism further extends parameter-level trade-off aggregation to perceptual design conflicts. Rather than treating competing requirements as noise or averaging them into generic solutions, the system models them as visual parameter trade-offs. This allows the framework to generate compromise solutions that remain aligned with consumer expectations while satisfying basic visual proportion priors.

The focused ablation study further clarifies the role of the major intermediate transformations. Persona-level aggregation primarily contributes to semantic alignment, conflict-aware reasoning supports trade-off handling among competing preference signals, and PDV-based parameterized prompting improves visual proportion plausibility and multi-view coherence compared with natural language prompting alone. These results help address the concern that the proposed workflow may be over-engineered relative to the evidence by showing that the main components contribute to different dimensions of generation quality.

6.2. Practical implications

From an industrial perspective, AutoGen-Insight provides a structured workflow for integrating large-scale consumer feedback into early-stage automotive design. By converting unstructured review data into persona-level preferences and probabilistic visual parameters, the framework can help design teams move more efficiently from market insights to visual concept exploration.



Figure 4. Comparative generation results for the urban commuter persona
Abbreviations: ISO: Isometric; LLM: Large language model.

The framework is especially useful in the fuzzy front end of design, where teams need to compare multiple design directions before committing to detailed modeling. For example, different personas may emphasize compactness, space utility, technology, comfort, or sportiness. AutoGen-Insight makes these differences visible and translates them into inspectable visual design directions, supporting more evidence-based concept exploration.

The inspection interface also improves practical applicability by allowing designers to review persona profiles, reasoning chains, PDVs, generated prompts, PDV-derived visual descriptors, and multi-view outputs. In this sense, the system does not replace professional designers. Instead, it provides interpretable design suggestions and visual candidates that can support faster iteration and more informed decision-making.

However, the framework's practical role should be carefully positioned. It is most suitable for early-stage design exploration, translating consumer insights, and comparing visual concepts. It is not yet a substitute for professional sketching, computer-aided design (CAD) modeling, aerodynamic analysis, ergonomic validation, or engineering feasibility assessment.

6.3. Limitations and future work

Despite promising results, several limitations remain.

First, although the empirical dataset is derived from a large-scale EV review corpus, the current evaluation focuses on a specific consumer review platform and a limited set of representative personas. Future work should examine whether the proposed mapping generalizes across different markets, cultural contexts, vehicle segments, and data sources.

Second, the current system operates primarily in the image domain. Although the PDV introduces visual design parameter priors, the generated outputs are not yet directly connected to CAD-compatible engineering models. Future work should explore integration with 3D generative modeling and parametric CAD systems to support downstream engineering workflows.

Third, the evaluation uses CLIP-based similarity and rule-based metrics for visual proportion plausibility. While these metrics provide a scalable quantitative assessment, they cannot fully capture professional design quality, manufacturability, brand identity, or long-term market acceptance. Future studies should combine computational metrics with larger-scale expert evaluation, consumer preference testing, and real-world design validation.

Finally, the current implementation relies on PDV-based parameterized prompt construction rather than image-level structural conditioning. Although this approach improves traceability and keeps the comparison with prompt-based baselines under control, it does not fully solve the problem of geometric consistency across views. Future work may explore PDV-derived wireframes, ControlNet-style structural guidance, or 3D-aware generative models to further improve spatial coherence and downstream design applicability.

7. Conclusion

This paper presented AutoGen-Insight, a consumer-driven generative design framework that reorganizes retained EV consumer review data into persona-level semantic preferences and transforms them into probabilistic visual parameter representations for generating parameterized EV concepts.

By introducing a probabilistic semantic-to-visual parameter mapping, a conflict-aware reasoning mechanism, and a PDV-derived prompt construction and shared multi-view generation setting, the proposed framework provides a traceable bridge between large-scale consumer feedback and early-stage automotive design exploration. Experimental results show that AutoGen-Insight improves semantic alignment, visual proportion plausibility, and cross-view consistency compared with direct prompting, keyword-based prompting, and LLM-generated prompting baselines.

The findings suggest that probabilistic intermediate representations can enhance the controllability and interpretability of generative design workflows. More broadly, this study demonstrates the potential of combining consumer insight mining, knowledge-based reasoning, and diffusion-based generation to support data-driven and human-centered early-stage product design.

Acknowledgments

This work was supported by the Xie Youbai Design Science Research Foundation (No. XYB-DS-202601). The authors would also like to express their sincere gratitude to Qingyang Jin and Wenyu Yuan for providing the foundational data and preliminary work for this study.

Funding

This work was supported by the Shanghai Planning Office of Philosophy and Social Science under Grant 2022BSH001, and the Ministry of Education in China Humanities and Social Sciences Project under Grant 23YJA760006.

Conflict of interest

Danni Chang is an Editorial Board Member of this journal, but was not in any way involved in the editorial and peer-review process conducted for this paper, directly or indirectly. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions

Conceptualization: Luyao Wang, Danni Chang

Data curation: Luyao Wang

Formal analysis: Luyao Wang

Funding acquisition: Danni Chang

Methodology: Luyao Wang, Chun-Hsien Chen

Project administration: Danni Chang

Software: Luyao Wang

Supervision: Chun-Hsien Chen, Danni Chang

Validation: Chun-Hsien Chen

Visualization: Luyao Wang

Writing—original draft: Luyao Wang

Writing—review & editing: Chun-Hsien Chen, Danni Chang

All authors read and approved the final manuscript.

Ethics approval and consent to participate

This study used previously collected and processed consumer review data and did not involve personally identifiable information. The expert and user evaluation was conducted using anonymized generated design outputs, and participants were informed of the evaluation purpose before providing ratings.

Consent for publication

Not applicable.

Availability of data

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Further disclosure

During the preparation of this manuscript, the authors used AI-assisted tools to support language polishing and editorial refinement. The authors reviewed and edited all AI-assisted outputs and take full responsibility for the content of the published article.

References

1. Bilgram V, Laarmann F. Accelerating innovation with generative AI: AI-augmented digital prototyping and innovation methods. *IEEE Eng Manag Rev.* 2023;51(2):18-25.

- doi: 10.1109/EMR.2023.3272799
2. Channi HK, Kaur A, Kaur S. AI-driven generative design redefines the engineering process. In: *Generative Artificial Intelligence in Finance*. Hoboken, NJ, USA: Wiley; 2025:327-359.
doi: 10.1002/9781394271078.ch17
3. Xiang Y, Wu Y, Song J, Gong Y, Liang P. Generative AI in industrial revolution: a comprehensive research on transformations, challenges, and future directions. *J Knowl Learn Sci Technol*. 2024;3(2):11–20.
doi: 10.60087/jklst.vol.3n2.p20
4. Gupta R, Kyaw AH. Insights informed generative AI for design: incorporating real-world data for text-to-image output. *arXiv*. Preprint posted online 2025.
doi: 10.48550/arXiv.2506.15008
5. Uusitalo S, Salovaara A, Jokela T, Salmimaa M. “Clay to play with”: generative AI tools in UX and industrial design practice. In: *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. New York, NY, USA; Association for Computing Machinery; 2024:1566-1578.
doi: 10.1145/3643834.3661624
6. Demirel HO, Goldstein MH, Li X, Sha Z. Human-centered generative design framework: an early design framework to support concept creation and evaluation. *Int J Hum Comput Interact*. 2024;40(4):933–944.
doi: 10.1080/10447318.2023.2171489
7. Rege A, Kim E, Kim S, Sirkin D, Currano R. Designing generative AI user interfaces for automobiles. In: *Adjunct Proceedings of the 16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. New York, NY, USA; Association for Computing Machinery; 2024:264–267.
doi: 10.1145/3641308.3677393
8. Jin Q, Wang L, Yuan W, Chang D. Mapping consumer voice into engineering insight: a structured language model-driven design support framework for electric vehicles. *J Eng Des*. 2026:1–40.
doi: 10.1080/09544828.2026.2639933
9. Park S, Lin K, Joung J, Kim H. An automated data-driven approach for product design strategies to respond to market disruption following COVID-19. *J Mech Des*. 2025;147(3):031402.
doi: 10.1115/1.4066684
10. Yu Y, Wang B, Zheng S. Data-driven product design and assortment optimization. *Transp Res Part E Logist Transp Rev*. 2024;182:103413.
doi: 10.1016/j.tre.2024.103413
11. Tian D, Li Q, Liu F, Khan J, Abbas MQ, Du Z. VOC data-driven evaluation of vehicle cabin odor: from ANN to CNN-BiLSTM. *Environ Sci Pollut Res Int*. 2024;31(22):32826–32841.
doi: 10.1007/s11356-024-33293-y
12. Wang Z, Liu W, Yang M. Data-driven affective product design using complete three-dimensional surface data. *J Intell Fuzzy Syst*. 2022;42(6):5437–5455.
doi: 10.3233/JIFS-211947
13. Briard T, Jean C, Aoussat A, Véron P. Challenges for data-driven design in early physical product design: a scientific and industrial perspective. *Comput Ind*. 2023;145:103814.
doi: 10.1016/j.compind.2022.103814
14. Liu X, Yang, S. Study on product form design via Kansei engineering and virtual reality. *J Eng Des*. 2022;33(6):412–440.
doi: 10.1080/09544828.2022.2078660
15. Lian W, Wang KC, Li Y, Chen HY, Yang CH, Bahubalendruni MVAR. Affective-blue design methodology for product design based on integral Kansei engineering. *Math Probl Eng*. 2022;2022:5019588.
doi: 10.1155/2022/5019588
16. Jin J, Jia D, Chen K. Mining online reviews with a Kansei-integrated Kano model for innovative product design. *Int J Prod Res*. 2022;60(22):6708–6727.
doi: 10.1080/00207543.2021.1949641
17. Bing Y, Yu L, Li S, Cho Y, Li C. A novel product shape innovation design method based on Kansei Engineering and GAN model with limited sample data. *J Eng Des*. 2026;37(3):981-1006.
doi: 10.1080/09544828.2025.2515553
18. Yang C, Liu F, Ye J. A product form design method integrating Kansei engineering and diffusion model. *Adv Eng Inform*. 2023;57:102058.
doi: 10.1016/j.aei.2023.102058
19. Huang Z, Guo X, Liu Y, Zhao W, Zhang K. A smart conflict resolution model using multi-layer knowledge graph for conceptual design. *Adv Eng Inform*. 2023;55:101887.
doi: 10.1016/j.aei.2023.101887
20. Peng C, Xia F, Naseriparsa M, Osborne F. Knowledge graphs: opportunities and challenges. *Artif Intell Rev*. 2023;56(11):13071–13102.
doi: 10.1007/s10462-023-10465-9
21. Xue B, Zou L. Knowledge graph quality management: a comprehensive survey. *IEEE Trans Knowl Data Eng*. 2023;35(5):4969–4988.
doi: 10.1109/TKDE.2022.3150080
22. Peng H, Zhang P, Tang J, Xu H, Zeng W. Detect-then-resolve:

- enhancing knowledge graph conflict resolution with large language model. *Mathematics*. 2024;12(15):2318.
doi: 10.3390/math12152318
23. Li C, Wu R, Yang W. Optimization and selection of the multi-objective conceptual design scheme for considering product assembly, manufacturing and cost. *SN Appl Sci*. 2022;4:91.
doi: 10.1007/s42452-022-04973-6
24. Marler RT, Arora JS. Survey of multi-objective optimization methods for engineering. *Struct Multidisc Optim*. 2004;26(6):369–395.
doi: 10.1007/s00158-003-0368-6
25. Harkare V, Mangrulkar R, Thorat O, Jain SR. Evolutionary approaches for multi-objective optimization and Pareto-optimal solution selection in data analytics. In: *Applied Multi-objective Optimization*. Springer; 2024:67–94.
doi: 10.1007/978-981-97-0353-1_4
26. Rashed NA, Ali YH, Rashid TA, Salih A. Unraveling the versatility and impact of multi-objective optimization: algorithms, applications, and trends for solving complex real-world problems. *arXiv*. Preprint posted online 2024.
doi: 10.48550/arXiv.2407.08754
27. Dharma IGSS, Setiawan R, Mahardika M, Prabowo AR. Comparative review of multi-objective optimization algorithms for design and safety optimization in electric vehicles. *IEEE Access*. 2024;12:152738–152768.
doi: 10.1109/ACCESS.2024.3475032
28. Zhang X, Tian Y, Wang H, Song Y. Training data selection with gradient orthogonality for efficient domain adaptation. *arXiv*. Preprint posted online 2026.
doi: 10.48550/arXiv.2602.06359
29. Zeng D, Li T, Yang J, *et al*. Expert-inspired multi-agent coordination for multi-objective molecular optimization. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Washington, DC, USA: AAAI Press; 2026;40(41):34575–34583.
doi: 10.1609/aaai.v40i41.40757
30. Lu P, Hsiao SW, Tang J, Wu F. A generative-AI-based design methodology for car frontal forms design. *Adv Eng Inform*. 2024;62:102835.
doi: 10.1016/j.aei.2024.102835
31. Wu Y, Ma L, Yuan X, Li Q. Human-machine hybrid intelligence for the generation of car frontal forms. *Adv Eng Inform*. 2023;55:101906.
doi: 10.1016/j.aei.2023.101906
32. Elrefaie M, Qian J, Wu R, Chen Q, Dai A, Ahmed F. AI agents in engineering design: a multi-agent framework for aesthetic and aerodynamic car design. In: *Proceedings of the ASME 2025 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. New York, NY, USA: American Society of Mechanical Engineers; 2025:V03BT03A048.
doi: 10.1115/DETC2025-169682
33. Fang YM. The role of generative AI in industrial design: enhancing the design process and education. *IET Conf Proc*. 2023;2023(45):135–136.
doi: 10.1049/icp.2024.0303
34. Süner-Pla-Cerdà S, Şen G, Kumbasar E, Şahin B, Ünlü CE. Designer experiences and perspectives on the role of generative AI in industrial design. *AI Soc*. 2026;41:2361–2384.
doi: 10.1007/s00146-025-02504-6
35. Zheng G, Zhou X, Li X, Qi Z, Shan Y, Li X. LayoutDiffusion: controllable diffusion model for layout-to-image generation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE/CVF; 2023:22490–22499.
doi: 10.48550/arXiv.2303.17189
36. Lian L, Li B, Yala A, Darrell T. LLM-grounded diffusion: enhancing prompt understanding of text-to-image diffusion models with large language models. *arXiv*. Preprint posted online 2023.
doi: 10.48550/arXiv.2305.13655
37. Jiang R, Zheng GC, Li T, Yang TR, Wang JD, Li X. A survey of multimodal controllable diffusion models. *J Comput Sci Technol*. 2024;39(3):509–541.
doi: 10.1007/s11390-024-3814-0
38. Hartwig S, Engel D, Sick L, *et al*. A survey on quality metrics for text-to-image generation. *IEEE Trans Vis Comput Graph*. 2025;31(10):9464–9483.
doi: 10.1109/TVCG.2025.3585077
39. Lin Z, Pathak D, Li B, *et al*. Evaluating text-to-visual generation with image-to-text generation. In: *Computer Vision–ECCV 2024*. Springer; 2024.
doi: 10.1007/978-3-031-72673-6_20
40. Tian Y, Liu Y, Wang S, Kwong S. Quality assessment for text-to-image generation: a survey. *IEEE MultiMedia*. 2025;32(2):44–52.
doi: 10.1109/MMUL.2025.3538862
41. Asperti A. Does CLIP perceive art the same way we do? *arXiv*. Preprint posted online 2025.
doi: 10.48550/arXiv.2505.05229
42. Kang R, Song Y, Gkioxari G, Perona P. Is CLIP ideal? No. Can we fix it? Yes! In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE/CVF; 2025:22436–22446.