ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

A Framework That Is AI-Assisted and Driven by User Behavior for Optimizing Form and Interaction of Smart Speakers

Wei Zhou[1] *and Yajun Liu*[1,*]

[1] School of Art and Design, Guangdong University of Science & Technology, Dongyuan Avenue, Dongguan, 523083, Guangdong, China

SUMMARY: *Along with the rapid expansion of smart sound boxes in family places, present products have on the whole converged toward unified forms and voice-guided interaction patterns, while have not sufficiently taken into account different user behavior customs. For the purpose of filling this empty position, this research has proposed a user-behavior-driven artificial intelligence auxiliary design framework which connects behavior expression, model selection, and multi-object design search for intelligent loudspeaker interaction and form optimization. Based on literature, one composite dataset which has 48 typical user character documents is used for constructing a controlled action space, and it does not replace real experience observation. By means of K-means and Gaussian Mixture Models, the analysis of user profiles is conducted by us for the extraction of interpretable archetypes. After that step, one mapping that behavior turns into design is given for four design variables: interaction strategy, feedback richness, form factor, and transparency/control. NSGA-II hence is utilized by us to carry out exploration of Pareto-optimal trade-off relations among these objectives. The optimization outcome displays directional gathering across the artificial design space: interaction comfort increases from 0.571 to 0.680, feedback-richness matching from 0.595 to 0.750, transparency matching from 0.586 to 0.645, meanwhile the form-punishment goal reduces from 0.208 to 0.000. These findings show that the variables at interaction level control design effect, whereas compact and visually not noticeable shapes still keep structural advantages. Therefore, this study should be regarded as a framework-validation work which makes the translation from behavior to design be clear and repeatable inside a controlled artificial space. Its contribution is in the providing of a systematic working flow for behavior-informed AI-aided design that can be expanded to wider smart-home and human-AI interaction contexts.*

KEYWORDS: *Smart speaker design, User behavior modeling, AI-assisted design, Multi-objective optimization, Human-AI interaction*

1 Introduction

Along with the fast popularization of custom-made indoor design and individualized living spaces, user anticipations for intelligent household apparatuses have more and more moved toward adaptive, context-conscious, and behavior-reactive interactions. Along with families keep on putting Internet-of-Things (IoT) technologies into daily use, intelligent sound boxes have already developed from simple devices that can play sound into core centers that let artificial intelligence do interaction inside houses. The recent forecast of market can further

*18326618661@163.com
https://doi.org/10.65102/is2026322 Published: 30 April, 2026

explain that smart home technologies are becoming main stream on the whole world scope. The whole world smart home market is forecasted to produce 174.0 billion US dollars of income in 2025 and keep a compound annual growth rate (CAGR) of 9.55% from 2025 to 2029, finally getting an expected market size of 250.6 billion US dollars. The family permeation ratio is expected to ascend from 77.6% in the year 2025 to 92.5% in the year 2029, which makes clear that the linked devices in domestic environments are getting more and more extensive. According to estimation, on average every installed smart home can bring about 62.46 dollar of income, therefore among all countries the United States holds the biggest market share, its 2025 estimated value is 43.0 billion dollar. [1].

In this quickly growing ecological system, smart sound boxes have already become core interactive nodes in modern intelligent houses [2]. In comparison with intelligent hand-held phones or traditional family control apparatuses, intelligent sound boxes allow for permanent open, hand-does-not-need interaction, thus making them locate themselves as natural key points for cooperating intelligent lamplight, electric equipment, amusement systems, and family services. Big platform suppliers have built up ripe product ecological systems through wide equipment matching and cloud-based artificial intelligence basic structures. Representative goods for example Amazon Echo, Baidu Xiaodu, Tmall Genie, and Xiaomi XiaoAi (Figures 1-4) show a high level of gathering together on both interaction method and physical shape, usually using small, round-barrel or disk-like shapes with voice-leading interaction added by very few vision hints.

Although differences exist in brand construction and ecological system combination, these main present smart speaker products have shared design approaches: interactive logic is for the most part arranged beforehand at the platform level, physical form emphasizes neutral character and the feature of not being prominent in space, and personal customization is mainly realized through software arrangement but not design alteration. Although these methods already have verification of their effectiveness for large-scale deployment, they also bring out a key limitation: user behavior in most cases is treated as an input for system optimization or recommendation, and not as a source of structured design knowledge that can actively supply guidance for interaction strategies, feedback configurations, and shape-giving decisions.

For the purpose of providing firm verification for the assertion that nowadays main-stream smart loudspeakers have reached a high degree of formal and interactive integration, Figures 1-4 display four representative goods from large business ecologies. We here use these cases not for comparing brands, but for finding repeated design trends in form language, interaction method, and visible control provided functions.

Figure 1: Amazon Echo Dot Max [3].

Figure 2: Xiaodu smart speaker Mate [4].

Figure 3: Tmall Genie Sugar Q [5].

Figure 4: XiaoAi smart speaker Pro [6].

Putting all together, the four examples display that a limited design word stock is shared by all platforms. Figure 1 and Figure 2 are examples of compact, voice-priority devices that have few physical structures and low visual importance; Figure 3 gives a softer and more decorative showing, however its interactive logic still takes far-field voice input as center and has limited feedback clues; Figure 4 has increased stronger hardware integration and more explicit light-related feedback, but it still keeps the same basic voice-led research framework. This comparison gives support to the core problem description of this research: current products have differences in branding and ecosystem connection, nevertheless they offer limited behavior-related particular differentiation on the design level.

In the same time, AI-aided design study has developed from rule-based expert systems and parameterized automatic methods toward data-guided, learning-relied, and mixed-start design surroundings [7-9]. This change has importance for intelligent product research and development, because design problems more and more include different kinds of behavior tracks, environment uncertainty, and many mutual conflict goals, which therefore cannot be processed well only by means of fixed experience rules.

Three branches of recent research are especially related here: mixed-start human-AI together-creation [10-13], multi-mode interactive design, and behavior-conscious experience design for intelligent products. Put together, these research works show that AI possesses the biggest usefulness in design not at the time when it takes the place of design judgement, but at the time when it assists in transforming scattered information about users into structures that can be understood, which are able to give support to later design choices.Automation-oriented tools toward adaptive, human-centered AI systems capable of translating raw behavioral data into actionable design insights.

Recent empirical studies further demonstrate the relevance of this problem. Conversational smart products in domestic environments have been examined from the perspectives of usability [14], trust formation and continued use [15], adoption of smart home speakers [16], and the broader integration of AI into UX design processes [17]. These research works give precious proofs concerning perception, acceptance, and system utilization, but they mostly still stay on the level of evaluation, adoption model construction, or interaction performance.

Therefore, there still exists a gap between the behavior analytics and the design practice in the development of smart speakers. Current products have shown mature shapes and interactive modes, nevertheless, the design procedure itself frequently has no clear mechanism that maps user behavior rules to design parameters like interaction method, feedback abundance, physical existence and clearness. This gap restricts the capability of smart sound boxes to carry out adaptation not merely on the dimension of system reaction, but also on the dimension of design logic.

As a response, this research puts forward a user-behavior-pushed AI-supported design method for smart speaker interaction and shape giving. This research ought to be explained as a work for verifying framework, not an empirical research at population level: its goal is to give formalization to a variable space from behavior to design, prove a repeatable working flow of clustering and optimization, and show that different behavior tendencies cause different trade-offs among interaction, feedback, form and transparency. The obtained results hence are explained as design hints in a limited artificial behavior area, not as direct experience-based broad statements for all smart-speaker users.

2 Methodology

2.1 Research Framework Overview

This research puts forward an AI-aided design frame that is driven by user behaviors, and therefore positions it as an exercise for verifying design methods. Instead of regarding AI as an independent producer of ultimate shapes, this framework employs AI as an explainable middle agent that connects behavior description, model picking-out, and design thinking.

After we have already built up the product-level convergence problem, the next work is to make clear how this current study changes different kinds of user behaviors into clear design choices. Therefore, Figure 5 is the summary of the complete analytical logic of the framework we have proposed.

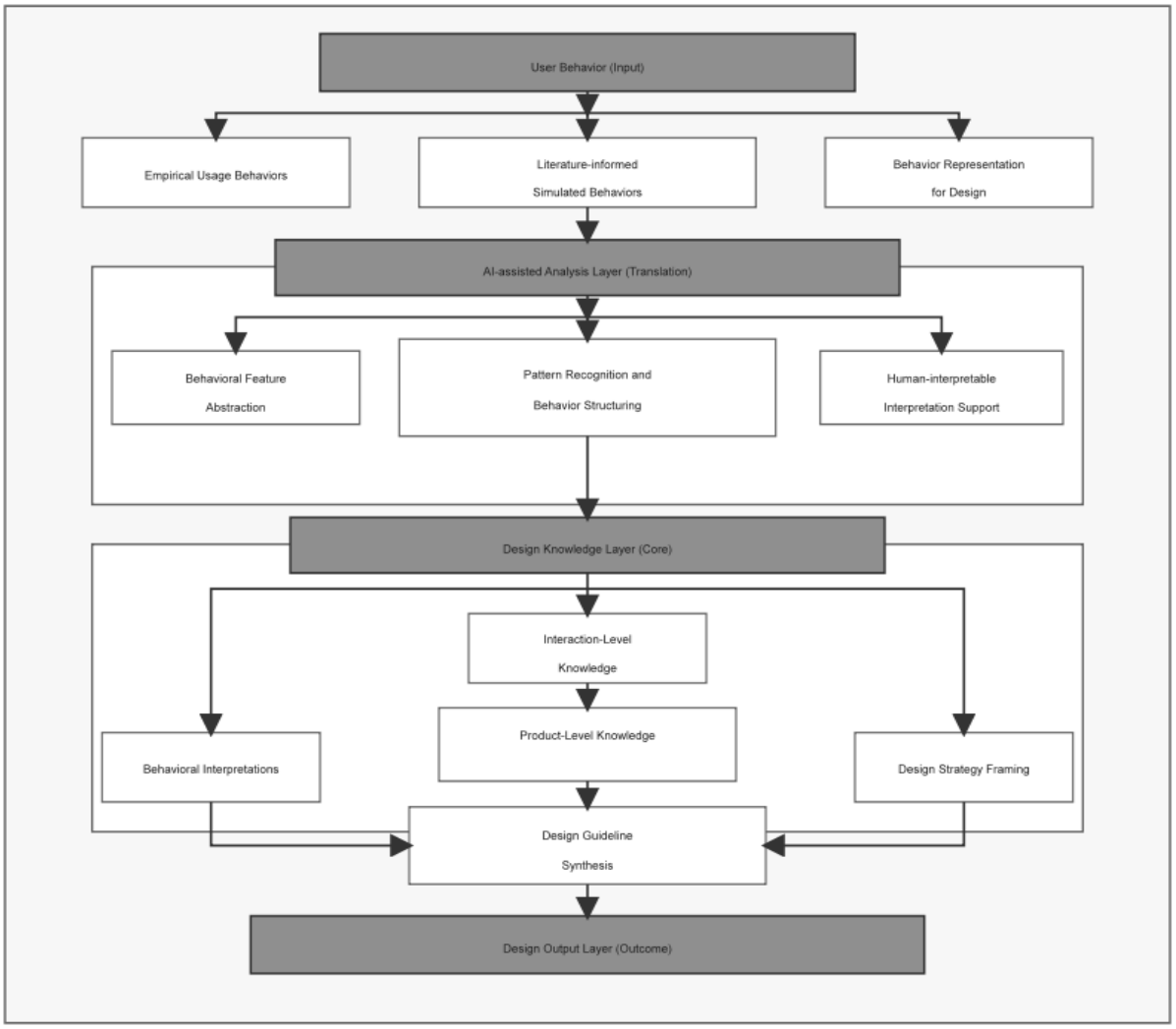

Figure 5: Research framework of the behavior-informed AI-assisted design pipeline.

From Figure 5, it can be seen that the contribution which this study makes is located in staged translation, not end-to-end prediction. The first layer builds a regulated action space out of variables which are got from literature; the second layer, by means of clustering, abstracts this space into interpretable archetypes; and the third layer carries out the mapping from these archetypes to design variables and Pareto-optimal selection schemes.

The first layer puts emphasis on the obtaining of behaviors. The utilization of smart speakers is expressed by a small but understandable group of variables which capture engagement strength, time arrangement, usage direction, interaction complication, activeness, and trust/control tendency inside family environments.

The second layer and the third layer proceed from abstractness to actualization. No-supervised learning is utilized to recognize regions with similar behaviors inside the synthesized data collection; after that, behavior-informed goal functions change these regions into design standards for interaction strategy, feedback abundance, shape parameter, and transparency/control. In this way, AI supports industrial design reasoning by making the behavior-to-design pathway explicit and inspectable.

2.2 Data Collection and Dataset Description

Foregoing researches about intelligent sound boxes and pronunciation helpers have already investigated user experience, using action, and accepting through comment digging, talking with people, question paper collecting, and machine-learning-based shape drawing [18-23]. In all these researches, several behavior aspects appear again with high uniformity: interaction times, time rule, use emphasis, interaction complication, initiation, and worry connected to trust. For the achieving of design research, however, the existing empirical data sets are hard to be directly used. They are commonly broken into pieces, specific to each method, and are rarely organized to link behavior features with the corresponding design decision variables. This therefore causes difficulty for the studying of how different user behaviors can be converted into adaptive interaction plans, feedback setups, shape elements, and transparent mechanisms inside one single analysis working procedure.

To bridge this gap, the present study adopts a literature-informed synthetic dataset as a controlled abstraction layer [24, 25]. The synthetic dataset is not introduced as a substitute for empirical observation; rather, it operationalizes recurring patterns reported in prior work so that behavior-to-design relationships can be explored in a transparent and comparable manner.

The six input variables which are defined in Table 1 give 972 theoretical behavior combinations (3×3×4×3×3×3). For getting a design space that can be handled and still has diversity, 48 representative character outlines were built by carrying out restricted layered sampling on this combinatorial space, which retains both extreme and middle behavior structures meanwhile removing combinations that are not reasonable in concept when considering repeated use patterns which existing smart-speaker research has recorded.

The dataset that we get after processing contains six input behavior characteristics and four output design characteristics. For the purpose of making the synthetic behavior space operable, six input variables have been chosen to represent the user behavior dimensions which appear most repeatedly in previous smart-speaker research works. In Table 1, the definitions of these variables and their interpretable levels are given by us.

Table 1: Input feature structure of the synthetic smart speaker user dataset

Variable	Levels	Represented Factors
Interaction frequency	Low / Medium / High	Overall intensity of user–device interaction, reflecting engagement level and dependence on the smart speaker.
Temporal usage pattern	Morning-dominant / Evening-dominant / Evenly distributed	Dominant time-of-day usage pattern, indicating daily routines and context of use.
Primary usage scenario	Information / Control / Entertainment / Mixed	Main functional purpose across typical smart speaker applications.
Interaction style complexity	Simple / Moderate / Complex	Degree of linguistic and command complexity, representing cognitive load and user proficiency.
User initiative level	Passive / Balanced / Proactive	Extent to which users actively explore functions rather than only responding to prompts.
Trust and control orientation	Low / Medium / High	Preference for transparency, privacy control, and system autonomy.

What Table 1 displays is that the input space has combined together engagement intensity, temporal routine, functional focus, interaction complexity, initiative, and trust/control preference. This disposition has analytical value, hence it not only holds what behaviors users conduct with intelligent sound boxes, but also holds how actively, how frequently, and under what understanding restrictions they conduct these actions. Therefore, this table lays the behavioral foundation for subsequent archetype picking-out, hence it does not act as a general descriptive item list.

Behavior analysis alone is not sufficient for design reasoning unless it can be translated into actionable design levers. Table 2 therefore defines the four output variables that constitute the design space explored in this study.

Table 2: Output feature structure of the synthetic smart speaker user dataset

Variable	Levels	Represented Factors
Adaptive interaction strategy	Guided / Adaptive / Exploratory	System guidance and autonomy during interaction.
Multimodal feedback configuration	Voice only / Voice + light / Voice + visual / Multimodal rich	Combination of feedback channels supporting clarity, accessibility, and user perception.
Product form factor	Compact invisible / Neutral object / Expressive central	Physical and symbolic presence of the device in the environment.
Transparency and control mechanism	Minimal / Moderate / Explicit	Degree of explicability, user control, and feedback regarding system behavior and data use.

Table 2 limits the output space to variables that can be directly interpreted and manipulated at the design level: interaction strategy, feedback richness, physical presence, and transparency/control. This limiting rule is on method grounds needed because it lets the optimization outcomes be read as design choices instead of abstract digital conditions. In this research, the synthetic dataset hence ought to be regarded as a hypothesis-producing and design-investigating instrument, which is used to complement but not substitute the follow-up empirical verification.

2.3 AI-Assisted Design Optimization Framework

On the basis of this controlled behavior space, the unsupervised clustering method is first utilized to discover representative behavioral prototypes which serve as middle connecting links between user features and design reactions. For the clustering work, nominal type variables are processed separately from ordered type variables, thus category labels cannot bring artificial distance relations into being; For the calculation of alignment in the optimization stage, the qualitative levels are encoded on limited numerical scales which can still keep the semantic interpretability.

Based on the already found behavioral archetypes, the design of smart speaker is hereby formulated as an AI-assisted multi-objective optimization problem. Each artificial user profile is connected with one candidate design resolution that is defined by four design output variables which are summarized in Table 2: adaptive interaction strategy, multimodal feedback setting, product shape factor, and transparency and control mechanism. Discrete categorical values are transformed into standardized number expressions, therefore enabling evolution search while retaining meaning explanation ability.

To evaluate alternative design candidates, a set of behavior-informed objective functions is defined to quantify the alignment between user behavior characteristics (Table 1) and design outputs. These objectives represent normalized design fitness.

The design problem is formulated as a behavior-informed multi-objective optimization task. Let x1-x6 denote the six user-side variables (interaction frequency, temporal usage pattern, primary usage scenario, interaction style complexity, user initiative level, and trust/control orientation), and let y1-y4 denote the four design-side variables (interaction strategy, feedback richness, form factor, and transparency/control). For the purpose of making this mapping have explicit expression, four middle objective numerical values are given definition in the way below:

$$t1 = (x1 + x4 + x5)/3; t2 = (s(x3) + x4)/2; t3 = (x1 + x5 + x6)/3; t4 = x6$$

where s(x3) maps usage scenario to a normalized feedback demand (information/control = 0.33, entertainment = 0.66, mixed = 1.00).

The four objective functions are then defined as: $f1 = 1 - |y1 - t1|$ for interaction comfort and interaction-strategy alignment; $f2 = 1 - |y2 - t2|$ for feedback-richness alignment; $f3 = y3 \cdot (1 - t3)$ for form efficiency, which is minimized as a penalty on excessive physical expressiveness; and $f4 = 1 - |y4 - t4|$ for transparency and trust alignment.

This formulation lets the translation from behavior to design become clearly expressed. The driving factors of f1 and f2 are mainly engagement degree, initiative spirit, scenario diversity degree, and interaction complexity degree; f3 presses down physically leading solutions when behavior proof does not give reason for such obvious position; and f4 divides users on the basis of trust and control expectations, not merely on the single basis of interaction work load. When taken together, these goals determine a multi-goal design space which is featured by built-in trade-offs, for example simplicity against expressiveness and transparency against cognitive load.

NSGA-II is chosen to be the core search method, because the design problem in itself is multi-objective, and no single weighted optimal result is expected in advance. In each generation there are 50 candidate solutions, and the algorithm makes the population evolve by means of selection, crossover and mutation, till the four normalized objectives reach a stable state. Therefore, the output is a Pareto set of design alternatives which are informed by behavior, not a single solution that people suppose to be universal. After the objective structure is already built up, the optimization flow that is carried out itself must be clearly put forward. Figure 6 gives an explanation of the way that NSGA-II carries out the search in the discrete design space and keeps the non-dominated alternative solutions.

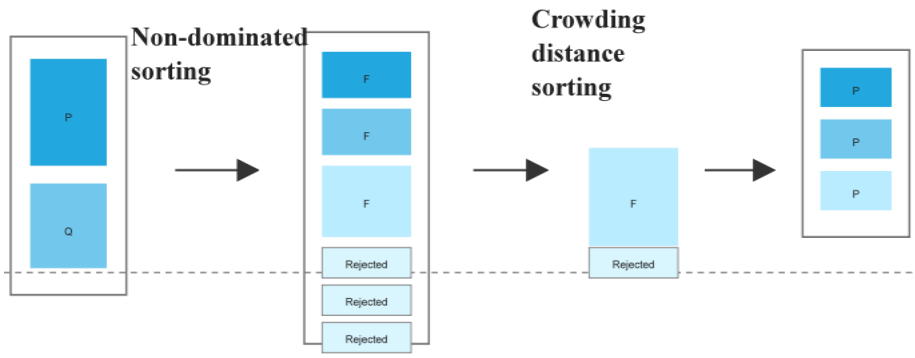

Figure 6: NSGA-II workflow for behavior-informed smart speaker design optimization.

Figure 6 lets us clearly know that this method is in the design for producing a group of trade-off solutions, but not a single optimal answer. The candidate design schemes are given initial values, are checked against the four objectives which are based on behaviors, are arranged in order by non-domination, and are made more various through selection and change operators. From methodology perspective this is suitable, because when we enhance feedback abundance or visibility for one archetype, hence this can raise interaction load or physical conspicuousness for another.

3 Results and Discussion

3.1 Encoding Strategy and Behavior-to-Design Mapping

The synthesized intelligent sound box user data set unites a behavior space and a design space. In the optimization phase, all qualitative grades are coded on normalized scales within [0,1], thus design alignment can be calculated uniformly across different types of variables, hence no interpretability is lost on the level of individual design selections. Here, a difference on method side has importance. The scalar coding contents recorded in Tables 3 and 4 are utilized for behavior-to-design alignment computation, hence clustering is explained via behavior structure instead of via random ordinal arrangement of nominal categories. This dividing method enhances the openness of the working flow and stops the numerical optimizing step from being mixed up with the grouping step.

3.2 Clustering Results and Optimization Setup

To examine whether the synthetic behavior space contains interpretable structure that can support design reasoning, a two-stage analytical procedure was used. First, encoded profiles were compared through clustering in standardized feature space. Second, the resulting archetypal tendencies were translated into a multi-objective design search process.

To evaluate alignment between user behavior and design outputs, the qualitative levels in Table 1 must be transformed into a bounded numerical representation. Table 3 reports the scalar encoding used for the optimization-stage calculation of behavior-design alignment.

Table 3: Input Feature Encoding for Smart Speaker User Behavior Dataset

Feature	Encoded Values (0–1)	Description
Interaction frequency	0 / 0.5 / 1	Low = 0, Medium = 0.5, High = 1.
Temporal usage pattern	0 / 0.5 / 1	Morning-dominant = 0, Evening-dominant = 0.5, Evenly distributed = 1.
Primary usage scenario	0 / 0.33 / 0.66 / 1	Information = 0, Control = 0.33, Entertainment = 0.66, Mixed = 1.
Interaction style complexity	0 / 0.5 / 1	Simple = 0, Moderate = 0.5, Complex = 1.
User initiative level	0 / 0.5 / 1	Passive = 0, Balanced = 0.5, Proactive = 1.
Trust and control orientation	0 / 0.5 / 1	Low = 0, Medium = 0.5, High = 1.

Table 3 makes the behavior space computationally tractable while preserving the interpretability of each level. At the same time, these scalar values should be read with an important methodological qualification: they are used for objective-function calculations, whereas clustering is interpreted from standardized behavior structure rather than from imposed ordinal distances between nominal categories.

Once user inputs are numerically represented, candidate design outputs must be encoded on

the same normalized scale to support multi-objective search. Table 4 provides this output-side encoding scheme.

Table 4: Output Feature Encoding for Smart Speaker Design Variables

Feature	Encoded Values (0–1)	Description
Adaptive interaction strategy	0 / 0.5 / 1	Guided = 0, Adaptive = 0.5, Exploratory = 1.
Multimodal feedback configuration	0 / 0.33 / 0.66 / 1	Voice only = 0, Voice + light = 0.33, Voice + visual = 0.66, Multimodal rich = 1.
Product form factor	0 / 0.5 / 1	Compact invisible = 0, Neutral object = 0.5, Expressive central = 1.
Transparency and control mechanism	0 / 0.5 / 1	Minimal = 0, Moderate = 0.5, Explicit = 1.

As what Table 4 has displayed, every design variable is coded on an understandable low-to-high continuous range instead of being taken as a random mark. This permits every optimized solution to be converted back into design words without any confusion: higher numerical values stand for more exploratory mutual action, richer multi-mode feedback, stronger expressive ability on physical aspect, and more explicit transparency/control. Therefore, this table can play the role of decoding key that is used for people to read the Pareto solutions which appear later. In order to carry out test on whether the synthesized behavior space contains partitions that can be interpreted, the coded user feature files were compared by using both K-means and Gaussian Mixture Models in the standardized feature space. PCA was then only utilized for the purpose of visualization, hence the geometric connection among all profiles could be examined without changing the clustering process itself. The utilization of both these two methods serves different explanatory objectives. K-means emphasizes close and distance-driven separations, hence GMM makes overlapping and boundary indeterminacy more clear. Therefore, this comparison is utilized by us here to sustain archetype explanation, instead of putting forward the claim that smart-speaker users can be divided into inflexible natural categories.

After defining the input and output encodings, the next methodological step is to specify how user behavior is translated into evaluative design criteria. Table 5 summarizes the revised behavior-to-design mapping for the four objective functions used in the optimization.

Table 5: Mapping Between User Behavior Inputs and Design Output Features (f1-f4) with Interpretation

Output Feature	Influencing Inputs	Numerical Interpretation (0–1)
f1: Adaptive interaction strategy / interaction comfort	Interaction frequency, interaction style complexity, user initiative level	Higher f1: more adaptive or exploratory interaction aligned with frequent, complex, proactive users. Lower f1: guided interaction for low-frequency, simple, passive users.
f2: Multimodal feedback configuration / feedback richness	Primary usage scenario, interaction style complexity	Higher f2: richer feedback for diverse or cognitively demanding use. Lower f2: lighter feedback for simple or constrained scenarios.
f3: Product form factor / form efficiency	Interaction frequency, user initiative level, trust/control orientation	Higher f3: stronger penalty for expressive forms when behavior does not justify physical prominence. Lower f3: compact or neutral forms aligned with low-burden use.
f4: Transparency and control mechanism / trust alignment	Trust/control orientation	Higher f4: more explicit transparency and user control for privacy-sensitive users. Lower f4: lighter transparency for lower control demand.

The modified diagram makes clear that the four goals do not have same contributions to design difference. f1 and f2 are mainly pushed forward by engagement degree, initiative spirit, scenario diversity degree, and interaction complexity degree; f3 punishes overmuch physical expression when behavior proof cannot support physical outstand; and f4 is controlled mainly by trust/control direction, hence it becomes a distinguishing yet second-level goal when users have similar interaction modes but have differences in privacy expectation.

To test whether the synthetic behavior space contains interpretable hard partitions, the standardized user profiles were first clustered with K-means and projected into two dimensions for visualization. Figure 7 shows the resulting partition structure in PCA space; PCA is used here only for visualization, not for model fitting.

Because hard dividing can hide the overlapping that exists between neighbor user kinds, a Gaussian Mixture Model was also fitted by us onto these same standardized profiles. Figure 8 gives the GMM outcome on the identical PCA projection for making direct visual contrast with Figure 7. Putting all together, these two visual figures are utilized as proof for both stable cluster core parts and transition areas. Therefore, clustering is regarded by this paper as an abstract tool for design thinking, and is not a statement that smart-speaker users belong to fixed, mutually exclusive experience categories.

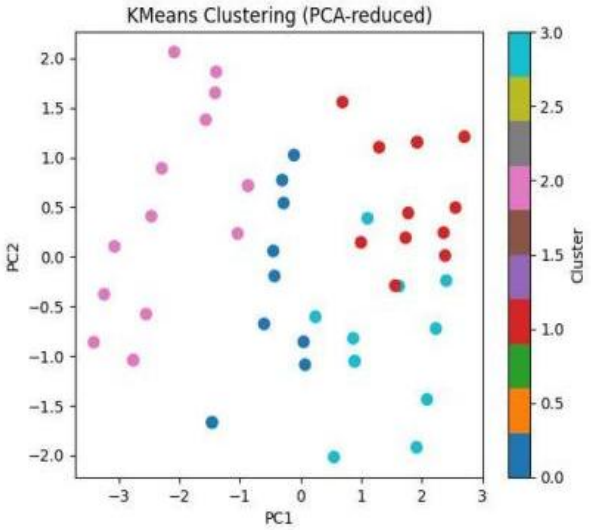

Figure 7: K-means clustering of synthetic smart speaker user profiles in PCA-reduced feature space.

Figure 7 gives us that several groups are tight and can be separated by eyes, hence this shows that the chosen behavior variables have enough distinguishing ability for distance-based clustering. However, the plot cannot be regarded as the self-contained proof of the model's appropriateness; its explanation must be read together with grouping quality standards and the actual understandability of the obtained archetypes. Figure 8 lets boundary ambiguity become more clear, for some profiles are placed in overlap areas not in clearly separated cluster centers. This difference is in analysis very important: K-means gives more clear partitions which are easier to explain as archetypes, therefore GMM can better express transitional users and mixed behavior tendencies. For the current frame-verification goal, the meaning of the comparison lies not so much in choosing a globally better clustering method as in proving that the synthesized behavior space includes both clear modes and gradual changes. This strengthens the view that the transformation from conduct to design is required to include stepped inclinations, hence it does not presume design regulations that suit all circumstances in the same

way. After this behavioral structure has got explanation, the analysis focus changes from segmentation to design balanced choices, that is to say how interaction strategy, feedback, form factor, and transparency change within the Pareto set which is produced by NSGA-II.

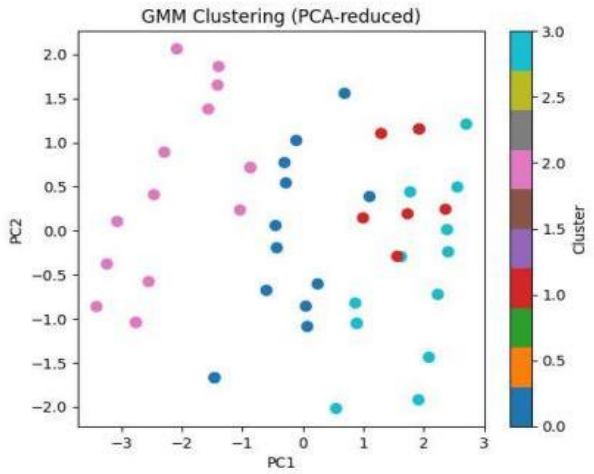

Figure 8: GMM clustering of synthetic smart speaker user profiles in PCA-reduced feature space.

This clustering together with optimization method shows that different behavior feature pictures can be changed into distinct but can be understood smart-speaker design plans inside a man-made space which has control.

3.3 Multi-objective Optimization Results

According to the encoded design variables and the mapping from behavior to design that we have given above, a multi-objective optimization work has been carried out by means of NSGA-II. The aim of this phase is not to obtain one single global optimum, but to investigate how the four targets develop together and what compromises still exist among alternative design schemes.

In every generation, there are 50 candidate solutions for design. By means of repeated selection and change, the algorithm seeks out non-controlled groupings of interacting comfort, feedback density matching, shape effectiveness, and transparency matching, thus permitting both gathering behavior and Pareto variety to be explained on the design layer. The pure visual cluster structure is not capable of proving the design performance. For the purpose of assessing whether the optimization brings about systematic improvement, not random fluctuation, Table 6 gives representative objective statistics from chosen generations together with two Pareto-optimal solutions.

Table 6: Evolution of objective values during NSGA-II optimization

Generation	f1	f2	f3	f4
0 (initial avg)	0.571	0.595	0.208	0.586
3 (mid avg)	0.646	0.721	0.039	0.647
6 (near-conv.)	0.680	0.750	0.000	0.645
Best solution A	0.681	0.750	0.000	0.644
Best solution B	0.660	0.750	0.000	0.672

Table 6, it clearly displays that directional convergence exists in the whole search process. From the initial population to generation 6, f1 increases from 0.571 to 0.680 (+19.1%), f2 from 0.595 to 0.750 (+26.1%), and f4 from 0.586 to 0.645 (+10.1%), while f3 decreases from 0.208 to 0.000.

The maximal increment occurs in feedback-richness matching, interaction comfort follows it, therefore this indicates that the search possesses the priority to advance software-level interaction variables earlier than it optimizes transparency-connected control. The whole reduction of f3 to zero further shows that shapes with physical manifestation are constantly governed by compact or neutral constructions within the current synthetic acting space. The two representative Pareto solutions have revealed a stable design core which has a limited but meaningful trade-off. Solution A obtains a slightly higher interactive comfort degree (f1 = 0.681), meanwhile it maintains the identical peak feedback alignment with Solution B (f2 = 0.750), therefore, therefore Solution B is able to generate higher transparency alignment (f4 = 0.672 when compared to 0.644). After the form inefficiency has been eliminated, the remaining choice for design is hence between the interaction smoothness and the explicit user control, not between the mutually competing hardware-guided options. From the substantial point of view, these rules show that user engaging features-especially interacting frequency, taking initiative, and interacting complexity-are handled most effectively by means of interacting strategy and multimodal returning information but not by means of increasing physical presence. Small-sized and visually not obvious shapes still have structure-related benefits in most areas of the synthetic behavior range. Transparency and control still have importance, but they work as supporting elements instead of being the main pushing forces of optimization. Lightweight privacy indices, as-needed explanations, and simple control abilities are therefore more match with the current design field than those with many instruments or visually strong control systems. Therefore, the results should be interpreted as being the evidence for objective convergence that exists in a controlled synthetic design space. These results support that the suggested work flow has use as a behavior-informed design exploration tool, while wider experience-based conclusion still needs to get confirmation from collected user information.

4 Conclusion

This research has proposed an AI-aided design framework which is pushed by user conduct for smart speaker mutual action and form optimization, with the explicit objective to link user conduct analysis and design decision making. It does not treat behavior only as an input for system adaptation, this framework translates behavior into structured design variables and trade-off relations that can be explained by people.

In the whole proof-of-concept work flow, the different tendencies of users were extracted into typical behavior areas by means of clustering, and then were mapped onto four design goals: interaction comfort, feedback richness, form efficiency, and transparency alignment. The optimizing results point out that interaction-layer variables control design achievement, whereas small and eye-not-unobvious shapes keep stably beneficial inside the present artificial behavior region.

The contribution of the study is therefore methodological as much as substantive. It shows how AI can operate as an intermediary that makes the behavior-to-design translation explicit, reproducible, and open to comparison, rather than as a black-box generator of form proposals. Accordingly, the present study should be read as a proof-of-concept demonstration of a behavior-informed computational design workflow rather than as a population-level empirical generalization about all smart-speaker users.

Although this research puts its focus on smart speakers, the framework is capable of being

transferred to other smart-home devices and interactive products which have the character of continuous user interaction and behavioral feedback. In the future work, researchers should bring in actual user observed data, report complete clustering metrics and profile matrices in supplementary materials, and combine behavior-informed optimization with more abundant generative form synthesis.

About the Author

Wei Zhou Born in Hefei, Anhui Province, China in 1986. Doctor of Philosophy in Art from the Eternal University. Main investigation orientations: intelligent goods, building ornamentation. Yajun Liu was born in Baoji, which is in Shaanxi Province, China, in the year 1980. Master of Arts in Art obtained from Guangzhou University. Main research investigation direction: visual art design.

Funding

Doctoral Research Fund Project

Project Title: Research on Styling and Interactive Mechanism Design of Trendy Toys Integrated with Mortise and Tenon Craftsmanship

Project Number: XJ2026001501

References

[1] Statista. (2025). Statista - smart home - worldwide.

[2] Kowalczuk, P. (2018). Consumer acceptance of smart speakers: A mixed methods approach. Journal of Research in Interactive Marketing, 12(4), 418-431.

[3] Amazon.com. (2025). Amazon Echo Dot Max Alexa speaker with room-filling sound and built-in smart home hub, Glacier White.

[4] Xiaodu / Baidu. (2025). Xiaodu smart speaker Mate.

[5] Tmall Genie / Alibaba Group. (2025). Tmall Genie Q Sugar smart speaker.

[6] Xiaomi / XiaoAi. (2025). XiaoAi smart speaker Pro.

[7] Knapp, D. W., & Parker, A. C. (1986). A design utility manager: The ADAM planning engine. In 23rd ACM/IEEE Design Automation Conference (pp. 48-54).

[8] Kim, S.-G., Yoon, S. M., Yang, M., Choi, J., Akay, H., & Burnell, E. (2019). AI for design: Virtual design assistant. CIRP Annals, 68(1), 141-144.

[9] Zhao, W., & Sun, Y. (2024). The exploration of emotional aspects of artificial intelligence (AI) in artistic design. International Journal of Interdisciplinary Studies in Social Science, 1(1), 58-65.

[10] Lin, Z., Ehsan, U., Agarwal, R., Dani, S., Vashishth, V., & Riedl, M. (2023). Beyond prompts: Exploring the design space of mixed-initiative co-creativity systems. arXiv.

[11] Zhu, J., Liapis, A., Risi, S., Bidarra, R., & Youngblood, G. M. (2018). Explainable AI for designers: A human-centered perspective on mixed-initiative co-creation. In 2018 IEEE Conference on Computational Intelligence and Games (CIG) (pp. 1-8).

[12] Liu, Z. (2025). Human-AI co-creation: A framework for collaborative design in intelligent systems. arXiv.

[13] Du, X., An, P., Leung, J., Li, A., Chapman, L. E., & Zhao, J. (2023). DeepThInk: Designing and probing human-AI co-creation in digital art therapy. International Journal of Human-Computer Studies, 181, 103139.

[14] Al-Hussein, A.-H. (2024). Envisioning a conversational smart chair to support people working from home: A usability-based evaluation.

[15] Hsieh, S. H., & Lee, C. T. (2024). The AI humanness: How perceived personality builds trust and continuous usage intention. Journal of Product & Brand Management, 33(5), 618-632.

[16] Bouallegue, S., Chtioui, J., Nefzi, A., & Chaney, D. (2025). The moderating role of user skills in consumers' adoption of smart home speakers in emerging markets. EuroMed Journal of Business.

[17] Stige, Å., Zamani, E. D., Mikalef, P., & Zhu, Y. (2023). Artificial intelligence (AI) for user experience (UX) design: A systematic literature review and future research agenda. Information Technology and People, 37(6), 2324-2352.

[18] Yoon, S.-H., Park, G.-Y., & Kim, H.-W. (2022). Unraveling the relationship between the dimensions of user experience and user satisfaction: A smart speaker case. Technology in Society, 71, 102067.

[19] Ammari, T., Kaye, J., Tsai, J. Y., & Bentley, F. (2019). Music, search, and IoT. ACM Transactions on Computer-Human Interaction, 26(3), 1-28.

[20] Kim, S. (2020). Exploring how older adults use a smart speaker-based voice assistant in their first interactions: Qualitative study. JMIR mHealth and uHealth, 9(1), 20427.

[21] Romero, J., Ruiz-Equihua, D., Loureiro, S. M. C., & Casaló, L. V. (2021). Smart speaker recommendations: Impact of gender congruence and amount of information on users' engagement and choice. Frontiers in Psychology, 12, 659994.

[22] Ashfaq, M., Yun, J., & Yu, S. (2020). My smart speaker is cool! Perceived coolness, perceived values, and users' attitude toward smart speakers. International Journal of Human-Computer Interaction, 37(6), 560-573.

[23] Choi, Y., & Lee, C. (2024). Profiling the AI speaker user: Machine learning insights into consumer adoption patterns. PLoS ONE, 19(12), 0315540.

[24] Latif, E., Chen, Y., Zhai, X., & Yin, Y. (2024). Human-centered design for AI-based automatically generated assessment reports: A systematic review. arXiv.

[25] Nikolenko, S. I. (2021). Synthetic data for deep learning.

[26] Rogers, Y., Sharp, H., & Preece, J. (2023). Interaction design: Beyond human-computer interaction (6th ed.). Wiley.

[27] Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. Theoretical Issues in Ergonomics Science, 2(4), 352-367.

[28] Löwgren, J., & Stolterman, E. (2004). Thoughtful interaction design.

[29] Varadarajan, R. (2023). Resource advantage theory, resource based theory, and theory of multimarket competition: Does multimarket rivalry restrain firms from leveraging resource advantages? Journal of Business Research, 160, 113713.

[30] Leal, S. S., & De Almeida, P. E. M. (2023). Traffic light optimization using non-dominated sorting genetic algorithm (NSGA2). Scientific Reports, 13(1), 15550.