

Requirement Analysis and Augmented Reality Interface Design for Custom Wood Veneer Cabinet Visualization Based on BERTopic Modeling

Xueyan Wu, and Yushu Chen *

Augmented reality (AR) interfaces for custom wood-veneered cabinets require accurate requirement elicitation that accounts for the anisotropic optical properties of natural veneer. However, existing approaches lack a systematic pipeline linking large-scale user feedback to interface specifications. This paper presents a data-driven requirement mining framework that couples BERTopic semantic topic modeling with user mental model construction and cross-validates the two through a Jaccard-based semantic mapping coefficient (SMC). Applying the framework to a corpus of 3163 multi-source semantic units, the pipeline consolidated 18 initial clusters into four core requirement themes through hierarchical merging (CV coherence improved from 0.421 to 0.573) and translated them into a three-tier priority specification (P0/P1/P2) via a Comprehensive Priority Index. A high-fidelity AR prototype incorporating physically based anisotropic veneer rendering was evaluated with 60 participants across four user groups. The prototype achieved a System Usability Scale score of 82.5 (SD = 4.4) with no significant inter-group differences ($F(3, 56) = 1.24, p = 0.303$), confirming that the requirement-driven design pipeline can yield robust cross-group usability.

DOI: 10.15376/biores.21.3.5785-5807

Keywords: Wood veneer furniture; Anisotropic texture perception; Augmented reality; BERTopic topic modeling; User mental model; Interaction design; Usability evaluation

Contact information: College of Furnishings and Industrial Design, Nanjing Forestry University, Nanjing, China 210037; *Corresponding author: cys@njfu.edu.cn

INTRODUCTION

Augmented reality (AR) systems for customized furniture design face challenges in representing the optical properties of natural materials within interactive environments. Natural wood veneer is among the predominant facing materials for custom cabinets. It is widely used due to its natural appearance, tactile quality, and perceived value. Its anisotropic grain orientation and gloss distribution make it one of the most difficult surfacing materials to reproduce digitally. Quarter-sawn and flat-sawn veneers from the same species differ fundamentally in grain orientation, figure width, and specular highlight distribution. Because of this inherent material variability, the visual appearance of wood veneer under varying lighting angles and ambient color temperatures cannot be faithfully captured by a single static image (Wan *et al.* 2021; Filip *et al.* 2024). As the furniture industry undergoes digital transformation, consumer demand for personalized wood furniture customization continues to grow (Manavis *et al.* 2024), and conventional two-dimensional color swatches and static renderings are increasingly inadequate for communicating the perceived quality of wood veneer surfaces. The AR technology, which overlays virtual content onto real physical environments, offers a practical means of closing

this perceptual mismatch in online wood furniture presentation.

Custom cabinet configuration in AR environments requires precise integration of spatial modeling, material representation, and user interaction mechanisms. The visual complexity of natural wood—arising from species-dependent grain figures, anisotropic specular reflectance, and batch-to-batch color variation quantifiable *via* CIELAB ΔE^*ab metrics—cannot be adequately communicated through two-dimensional swatches or static renderings. The AR technology can markedly improve consumers' ability to evaluate veneer appearance under realistic lighting and spatial conditions (Odstrcil *et al.* 2024). However, designing AR interfaces that faithfully render the anisotropic optical behavior of wood veneer while maintaining low-friction interaction and supporting purchase decisions remains a problem in furniture digital design.

Formalizing user requirements for such systems poses additional challenges. In the process of selecting and configuring custom cabinets, user requirements are multidimensional, involving both explicit functional demands and implicit perceptual needs related to the appearance of wood material. Because wood veneer perception is shaped by grain pattern, gloss, and color temperature, traditional scale-based methods alone cannot adequately reveal the underlying cognitive structure. Adopting a mental model as the cognitive framework allows systematic mapping of motivations, expectations, and perceptual responses throughout the cabinet customization lifecycle (Hassenzahl and Tractinsky 2006; Norman 2013). Structuring user cognition along orthogonal axes can reveal knowledge gaps and design-model mismatches, which motivates the dual-dimensional model adopted in this study. For analyzing large-scale unstructured feedback data, BERTopic topic modeling has been adopted to capture latent requirement structures. Compared with traditional approaches, such as LDA, BERTopic uses contextual embeddings and density-based clustering (Grootendorst 2022), enabling more reliable semantic extraction from heterogeneous and sparse textual data. The approach is well suited for extracting wood-material-specific perceptual needs from corpora comprising both interview transcripts and online platform reviews.

Three relevant research streams—AR home furnishing applications, user experience requirements mining, and wood material perception science—have evolved largely in parallel, with few attempts to integrate wood-specific material properties across domains. The present study addresses this gap. In AR home furnishing applications, Javornik (2016) and Baytar and Chung (2024) demonstrated that immersion and material realism are key factors influencing consumer trust, and Khirbat and Sriram (2025) further showed that consumer attitudes mediate the link between AR exposure and purchase intention in home décor and furnishings, yet none of these addressed the anisotropic optical properties of wood materials in interface design. In user experience requirements mining, text-mining and deep language model approaches have been used to extract attribute-level user needs from product reviews, but these methods have not been coupled with cognitive models for AR interface design. In wood material perception science, the CIELAB $\Delta E^*ab \approx 2$ just-noticeable difference threshold has been established for evaluating wood color variation (Buchelt and Wagenführ 2012), and large-sample empirical studies have shown that grain uniformity, directionality, and surface gloss anisotropy significantly affect consumer preference and material recognition (Manuel *et al.* 2015; Wan *et al.* 2021). However, these perceptual findings have yet to inform the requirements analysis or functional design of AR interfaces for wood products.

Therefore, this study proposes an interactive interface design methodology following a closed-loop pathway: user research—mental model construction—BERTopic

dimensionality reduction and clustering—requirement mapping—design practice. The following three objectives are pursued: (1) to develop a data-driven framework for extracting and structuring user requirements for AR-based cabinet customization systems; (2) to evaluate the semantic alignment between qualitative mental models and quantitative topic modeling results; and (3) to implement and validate an AR interface prototype based on the extracted requirement hierarchy. The contributions of this work, relative to prior requirement analysis and AR furniture studies, are fourfold: first, the application of BERTopic to AR wood-veneered cabinet requirement mining, where heterogeneous text sources are integrated rather than relying on predefined Kano-style categories; second, a Jaccard-based Semantic Mapping Coefficient (SMC) that cross-validates qualitative mental models against quantitative topic clusters with a permutation-test null baseline; third, a physically based anisotropic rendering pipeline tailored to species-specific veneer appearance (walnut, white oak, cherry) within an AR runtime; and fourth, a Comprehensive Priority Index (CPI) that combines empirical demand weights with expert engineering judgment to drive a P0/P1/P2 priority specification for AR interface functions.

EXPERIMENTAL

Research Framework and Data Collection

This study proposes a data-driven methodological framework for modeling user requirements and developing AR interfaces for custom cabinet systems.

A data-driven pipeline was developed, consisting of four stages: user segmentation, mental model construction, BERTopic-based semantic modeling, and requirement integration with AR interface implementation. The complete research framework is shown in Fig. 1. The framework consists of two parallel processing streams: a qualitative pipeline for mental model construction and a quantitative pipeline for BERTopic-based semantic modeling. The quantitative pipeline performs semantic embedding, dimensionality reduction, density-based clustering, and hierarchical topic merging on the multi-source corpus. The two streams converge through semantic cross-validation and priority ranking before entering high-fidelity prototype development and usability evaluation.

In the quantitative data collection phase, Likert five-point scale questionnaires were distributed online and offline to identify respondents' pain points with traditional cabinet display methods and their acceptance of AR interface interaction. Sampling combined purposive and convenience strategies. Inclusion criteria required respondents to (a) be 20 to 55 years old, (b) have purchased or consulted on custom cabinets within the past 24 months, and (c) possess basic smartphone proficiency. Sources included offline home furnishing stores ($n = 45$) and online home renovation communities ($n = 37$), representing users from different purchasing channels. A total of 82 questionnaires were collected; after excluding invalid responses (completion time < 120 s or ≥ 10 consecutive identical answers), 74 valid questionnaires remained (validity rate 90.2%). Data were analyzed in SPSS for reliability and validity. The overall Cronbach's α was 0.835, the KMO value was 0.762, and Bartlett's test of sphericity was significant ($p < 0.05$), confirming satisfactory internal consistency and structural validity.

For targeted purposive sampling, continuous characteristic indicators from the 74 valid samples were subjected to Min-Max normalization (Eq. 1), followed by hierarchical clustering using cosine similarity for distance measurement and Ward's minimum variance method for merging.

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

The optimal cluster number was determined by calculating silhouette coefficients (SC) for $k = 2$ through $k = 6$. The $k = 4$ solution yielded the highest silhouette coefficient (SC = 0.61), and the dendrogram showed the largest inter-cluster distance gap at the four-cluster cut point. Samples were accordingly segmented into four groups: function-driven, spatial modification, tech-oriented, and specialized needs (Fig. 2). These four user profiles provided the stratified sampling framework for subsequent in-depth interviews.

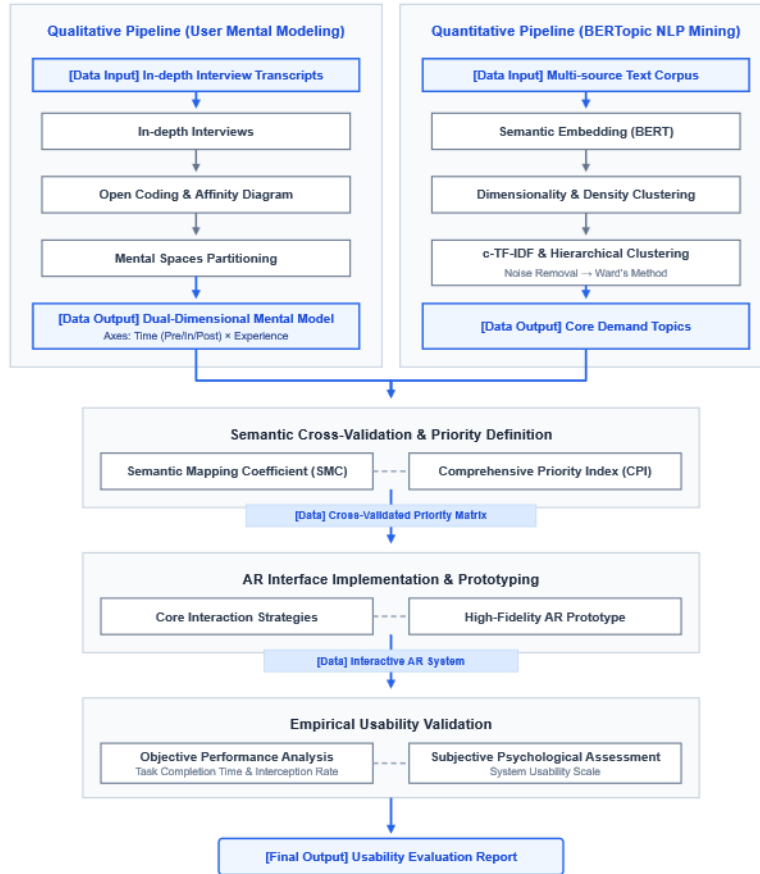


Fig. 1. Proposed data-driven pipeline for AR interface development

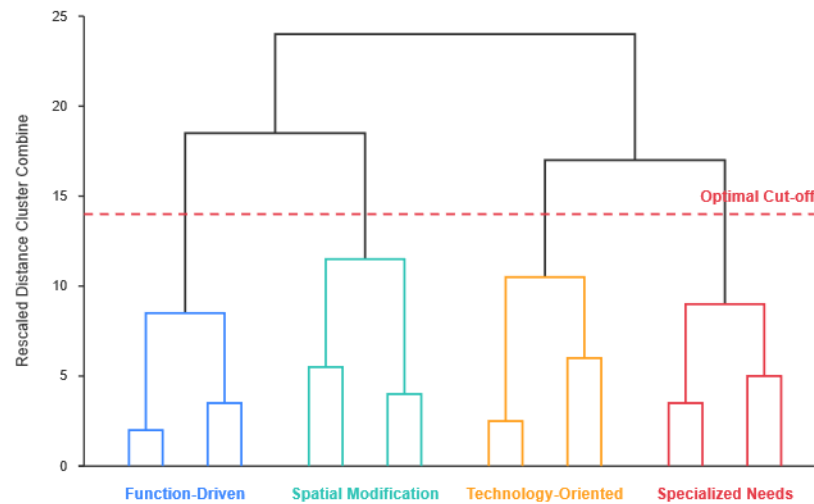


Fig. 2. User group clustering analysis results ($k = 4$)

Mental Model Construction and Key Tasks

Building on the quantitative segmentation, 20 users were selected from the four groups for semi-structured in-depth interviews. The interviews focused on respondents' operational frustrations in real kitchen settings, their visual expectations for AR three-dimensional projection, and their cognitive load during interface customization operations such as dimension adjustment and material replacement. Recordings were transcribed verbatim and analyzed by open coding (Braun and Clarke 2021), focusing on action-oriented phrases and emotional expressions. For example, “the shelf spacing is fixed, tall items cannot fit in” was coded as a functional pain-point task unit, and “designing it myself on my phone feels like I built it with my own hands” was coded as a hedonic task unit. After systematic comparison and deduplication, 39 key interaction task points were extracted. Participant demographic composition (10 male / 10 female; age range 25 to 52, mean 36.8), the three-section interview protocol, open-coding procedure (Cohen's $\kappa = 0.81$; Landis and Koch 1977), saturation assessment (Guest *et al.* 2006), and coded example segments are reported in Appendix A.1.

Using the affinity diagram method, the 39 tasks were consolidated into 15 “task towers” through bottom-up merging based on semantic proximity, without a predefined classification scheme. Based on the final behavioral goals of each task tower, they were further aggregated into six mental spaces: current usage and expectations, visual perception, operational experience, design expression, emotional interaction, and purchase decision. Within the dual-axis framework, the pre-use phase encompassed “current usage and expectations” and “visual perception”; the in-use phase encompassed “operational experience” and “design expression”; the post-use phase encompassed “emotional interaction” and “purchase decision.” The two execution-dimension mental spaces corresponded to users' cognitive processes during pragmatic tasks such as spatial scanning and module configuration; the four evaluation-dimension mental spaces corresponded to hedonic pathways through which users form satisfaction judgments, establish brand trust, and reach purchase decisions. Within the “visual perception” mental space, which encompassed 15 of the 39 original interaction tasks, 8 directly involved wood veneer perceptual attributes (53%), including “real-time preview of wood veneer color differences

under different lighting,” “comparison switching between quarter-sawn and flat-sawn grain effects,” and “dynamic simulation of veneer gloss variation with viewing angle.” This distribution indicated that the anisotropic perceptual properties of wood materials were the dominant driver of user visual requirements, providing qualitative support for the high weight of wood-related semantics observed in the subsequent BERTopic analysis.

The six mental spaces were organized along two orthogonal axes into a dual-dimensional user mental model (Fig. 3): the temporal dimension (pre-use, in-use, post-use) and the experiential dimension (execution vs. evaluation).

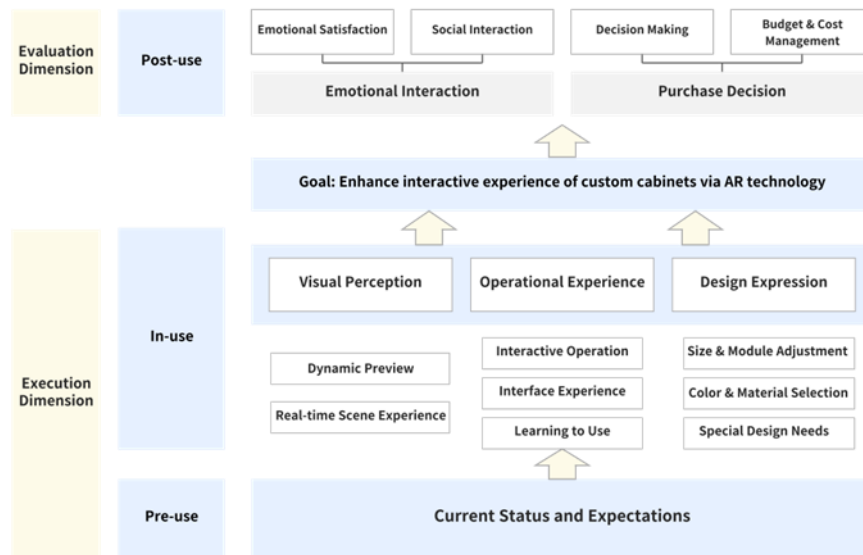


Fig. 3. Dual-dimensional user mental model

BERTopic-Based Topic Mining

Compared with traditional topic modeling approaches, such as LDA, BERTopic enables more effective semantic clustering by leveraging contextual embeddings and density-based clustering (Grootendorst 2022). To supplement the qualitative mental model with large-scale objective evidence, BERTopic topic mining was applied to a multi-source corpus. The internal corpus comprised interview transcripts and open-ended questionnaire responses, totaling approximately 1,200 semantic units. To mitigate feature space sparsity, an external corpus was constructed *via* Python web scraping, collecting user reviews from online home furnishing communities and application stores for “3D design software” and “AR measurement applications.” External data were collected from August to November 2025. After rule-based cleaning, deduplication, and semantic segmentation, 1,963 valid external texts were retained. The combined dataset totaled approximately 3,163 semantic units.

The external corpus drew on reviews of general 3D design and AR applications rather than dedicated wood-veneered cabinet AR feedback. This surrogate source was selected for two reasons: first, no mature AR application specifically serving wood-veneered cabinets exists on the market, making directly comparable large-scale feedback unavailable; second, 3D design software and AR measurement applications share core interaction patterns with cabinet AR interfaces (spatial scanning, 3D model manipulation, material switching, and dimension adjustment). The internal interview corpus (1,200 units) anchored domain-specific semantics, while the external corpus expanded feature space

coverage for general interaction dimensions. This domain substitution may have led to underestimation of certain wood-material-specific perceptual needs; this limitation is addressed later in the Discussion. To verify that the surrogate corpus retained domain-specific perceptual signal, a lexical audit was performed after preprocessing using a curated 87-term dictionary covering grain and texture, color and lighting, surface finish, and material realism. Approximately 47.3% of external semantic units contained at least one wood- or material-related term (*versus* 71.6% in the internal interview corpus), confirming substantial material-perception content in the external sources. The category-level breakdown is reported in Appendix A.3.

Text preprocessing included domain expression standardization, tokenization using the Jieba segmentation engine with a custom domain dictionary, part-of-speech filtering, and stopword removal based on the HIT Chinese stopword list. The cleaned text was encoded into 768-dimensional semantic vectors using the BERT-base-Chinese pre-trained language model (Devlin *et al.* 2019). UMAP was applied for dimensionality reduction and HDBSCAN for density clustering. Representative topic keywords were extracted using c-TF-IDF weighting (Eq. 2),

$$c - \text{TF} - \text{IDF}(\omega, c) = \text{TF}(\omega, c) \times \log\left(1 + \frac{N}{\text{DF}(\omega)}\right) \quad (2)$$

where $\text{TF}(\omega, c)$ is the frequency of word ω in topic cluster c , $\text{DF}(\omega)$ is the number of topic clusters containing word ω , and N is the total number of topic clusters.

The BERTopic parameter configuration was as follows: semantic embedding used BERT-base-Chinese (v1.0) with 768-dimensional output. The UMAP parameters were as follows: `n_neighbors = 15`, `n_components = 5`, `min_dist = 0.0`, `metric = "cosine"`. The HDBSCAN parameters were as follows: `min_cluster_size = 10`, `min_samples = 5`, `metric = "euclidean"`, `cluster_selection_method = "eom"`. The c-TF-IDF retained the top 20 representative tokens per topic cluster. Experiments ran on Python 3.9 with BERTopic 0.15.0, sentence-transformers 2.2.2, UMAP-learn 0.5.3, and HDBSCAN 0.8.33. All code will be released *via* GitHub upon paper acceptance.

Sensitivity analysis was conducted on the `min_cluster_size` parameter (values 5, 10, 15, and 20) to verify robustness. A value of 10 yielded the highest CV coherence score after hierarchical merging (0.573); smaller values (5) led to over-segmentation and semantic overlap, while larger values (15, 20) absorbed low-frequency domain-specific topics. Parallel comparison of the UMAP `n_neighbors` parameter (values 10, 15, and 20) showed that `n_neighbors = 15` produced the most stable topic structure. A complementary scan over UMAP `n_neighbors` (10, 15, 20) under fixed `min_cluster_size = 10` confirmed that `n_neighbors = 15` produced the most stable cluster structure across 10 random-seed runs. Full sensitivity tables are provided in Appendix A.2.

RESULTS AND DISCUSSION

Clustering and Reconstruction of AR Interface Requirement Topics

Initial density clustering using BERTopic produced 19 topic clusters. The HDBSCAN identified 16.64% of documents as noise, resulting in 18 interpretable topic clusters after filtering. To assess potential redundancy among these topics, pairwise cosine similarity was computed (Eq. 3), and a similarity heatmap was constructed (Fig. 4).

$$\text{sim}(A, B) = \cos(\theta) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (3)$$

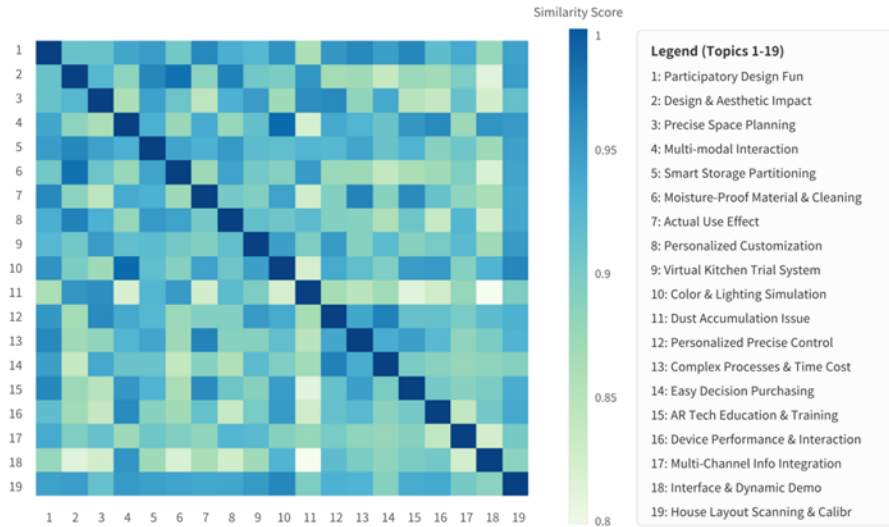


Fig. 4. Cosine similarity heatmap of initial BERTopic sub-topics

Several functionally aligned sub-topics exhibited cosine similarity as high as 0.97, indicating redundancy. Agglomerative hierarchical clustering with Ward's minimum variance criterion (Eq. 4) was therefore applied for secondary structural optimization:

$$D(A, B) = \frac{n_A n_B}{n_A + n_B} \|\vec{\mu}_A - \vec{\mu}_B\|^2 \quad (4)$$

The 18 initial topics were ultimately consolidated into four core requirement themes (Figs. 5 and 6): functional experience and engagement (34.9%), reflecting users' demands for intuitive gestures, such as drag-and-drop placement and single-finger module switching, and emphasizing user empowerment in design; visual perception and material texture (26.5%), focused on high physical fidelity of virtual elements (*e.g.*, cabinet door textures, ambient light reflections); interaction effects and experience (24.7%), emphasizing low latency and smooth feedback during dynamic operations; and spatial adaptation and integration (14.0%), requiring precise spatial collision detection and real-time visual warning at the interface level. These four themes defined the core engineering challenges for AR interaction design.

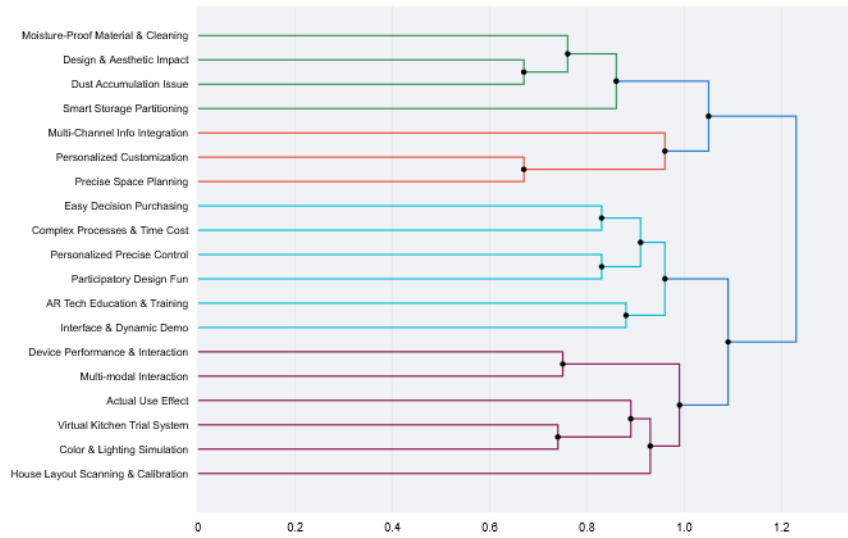


Fig. 5. Distribution of four core user requirement themes

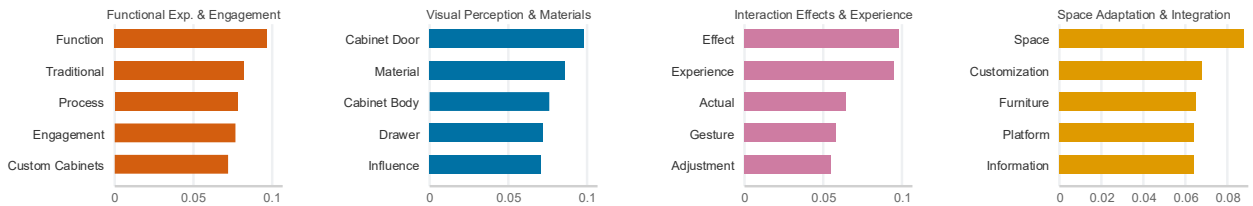


Fig. 6. Hierarchical clustering dendrogram and topic merging results

The average CV coherence score of the 18 initial sub-topics was 0.421 (SD = 0.068). After hierarchical merging, the four core themes achieved an average CV coherence of 0.573 (SD = 0.041), a 36.1% improvement ($\Delta = +0.152$).

Cross-Validation of Requirements and Interface Function Priority Ranking

Triangulation was used to cross-validate the qualitative mental model against the quantitative topic model. A set-intersection-based semantic mapping coefficient (SMC), equivalent to the Jaccard index, was introduced to quantify semantic alignment between the two systems (Eq. 5),

$$SMC(K_i, T_j) = \frac{|K_i \cap \varphi(T_j)|}{|K_i \cup \varphi(T_j)|} \tag{5}$$

where K_i is the c-TF-IDF feature word set of the i -th core theme, and T_j is the interaction task node set of the j -th mental space; $\varphi(\cdot)$ is a semantic generalization function that applies the same Jieba tokenizer and stopword list described above, followed by the TextRank algorithm (damping coefficient $d = 0.85$, co-occurrence window $w = 5$) to extract the top 20 feature tokens by weight.

A permutation test (1,000 random iterations) yielded a random SMC mean of 0.23 (SD = 0.09) with a 99th percentile of 0.47. The strong mapping threshold was accordingly set at $SMC \geq 0.70$ ($p < 0.001$). The bubble matrix (Fig. 7) revealed strong convergence between the two systems at the core requirement level. “Functional experience and engagement” mapped onto “design expression” (SMC = 0.86) and “emotional interaction” (SMC = 0.78). “Visual perception and material texture” formed the strongest single-point

mapping with “visual perception” (SMC = 0.92). “Interaction effects and experience” was tightly coupled with “operational experience” (SMC = 0.88). “Spatial adaptation and integration” mapped strongly onto both “current usage and expectations” (SMC = 0.82) and “purchase decision” (SMC = 0.75). Feature word intersection analysis for the SMC = 0.92 pair showed that approximately 68% of overlapping terms (13 of 19) were wood science terminology (e.g., “grain orientation,” “glossiness,” “color temperature sensitivity,” “anisotropy”), confirming that wood material properties carry substantial weight in the requirement structure.

Potential methodological coupling arising from shared tokenization is examined in the Limitations subsection.

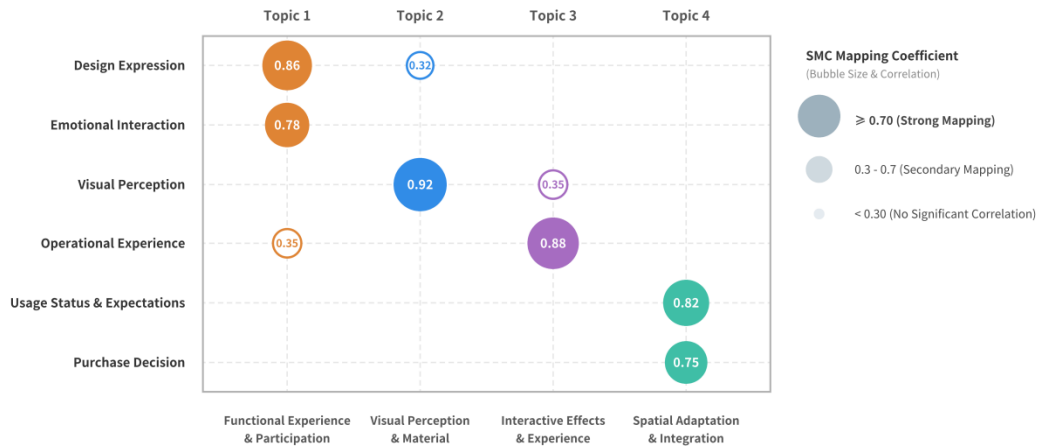


Fig. 7. Semantic mapping coefficient (SMC) bubble matrix between qualitative and quantitative dimensions

Guided by the core requirements and in line with mobile AR design guidelines and Poka-yoke error-proofing principles, a pool of 12 candidate functions was compiled. A Delphi panel of 5 senior UI/UX designers and 3 custom home product managers rated each function on Likert 5-point scales for “technical feasibility” and “user experience benefit.” Min-Max normalization was applied to unify scales (Eq. 6), using theoretical boundary limits ($W_{\min} = 0\%$, $W_{\max} = 40\%$; $S_{\min} = 1.0$, $S_{\max} = 5.0$) to ensure scale comparability and avoid zero-value artifacts in empirical distributions:

$$S_i^* = \frac{S_i - S_{\min}}{S_{\max} - S_{\min}}, \quad W_i^* = \frac{W_i - W_{\min}}{W_{\max} - W_{\min}} \quad (6)$$

The Comprehensive Priority Index (CPI) was defined as Eq. 7,

$$CPI_i = \alpha \cdot S_i^* + \beta \cdot W_i^* \quad (7)$$

where $\alpha = \beta = 0.5$, giving equal weight to raw user demand and expert engineering judgment. Priority rankings are presented in Table 1. P0 functions included “spatial boundary highlight warning” and “PBR-based high-fidelity material rendering,” forming the trust foundation for virtual space reconstruction. P1 functions included “3D gesture mistouch filtering” and “single-finger drag parameter adjustment,” targeting cognitive load reduction. P2 functions included decision-support utilities (scheme comparison, collection, and multi-channel sharing) and hedonic extensions, designated as auxiliary enhancements pending core interaction validation.

Table 1. Priority Classification of AR Custom Cabinet Interface Functional Indicators

| Requirement Type | W_i | S_i | CPI_i | Priority |
|-----------------------|--------|-------|---------|----------|
| Visual Presentation | 26.50% | 4.8 | 0.81 | P0 |
| Functional Engagement | 34.85% | 4.5 | 0.87 | P0 |
| Basic Operation | 24.68% | 4.2 | 0.71 | P1 |
| Decision Support | 13.97% | 4.0 | 0.55 | P2 |
| Emotional Extension | 8.50% | 2.8 | 0.33 | P2 |

Note: W_i = empirically derived demand weight for the i -th requirement (BERTopic topic proportion, or within-Theme-1 sub-node frequency for Emotional Extension), normalized to [0, 1]; S_i = expert engineering composite score (5-point Likert scale, rated by a Delphi panel on technical feasibility, user-experience benefit, and implementation criticality), normalized to [0, 1]. To avoid zero-value artifacts and ensure scale interpretability, Min-Max normalization (Eq. 6) used theoretical boundary limits: $W_{min} = 0\%$, $W_{max} = 40\%$ (upper bound based on plausible topic proportion range); $S_{min} = 1.0$, $S_{max} = 5.0$ (Likert scale endpoints). CPI_i is computed as $\alpha \cdot W_i^* + \beta \cdot S_i^*$ with $\alpha = \beta = 0.5$ (Eq. 7). Priority bands: $CPI \geq 0.80 \rightarrow P0$; $0.60 \leq CPI < 0.80 \rightarrow P1$; $CPI < 0.60 \rightarrow P2$. Requirement categories are derived from the four BERTopic themes (see Fig. 5) combined with Delphi panel assessment; column names reflect interface-level functional categories rather than raw topic labels.

Interactive Interface Design Strategies and High-fidelity Prototype

Explicit mapping for spatial comprehension

Accurate registration of virtual objects within the physical environment is a prerequisite for user confidence in AR cabinet tools. Ecological interface design holds that complex process variables should be made perceptually accessible through direct visual mapping (Vicente and Rasmussen 1992). Accordingly, the spatial construction module translates depth-estimation outputs into perspective reference grids and real-time feature-point overlays. An error-proofing mechanism implements Norman's forcing-function concept (Norman 2013): when a proposed cabinet placement violates an architectural constraint, the boundary turns an alert color, prompting correction before confirmation. For natural veneer cabinetry, where batch-to-batch color variation is unavoidable, rework after installation wastes both material and labor, making pre-placement warnings especially useful.

Progressive disclosure for cognitive load reduction

Although "functional experience and engagement" ranked as the top requirement (P0), showing the full parameter space of cabinet SKUs in a single view would likely exceed working-memory capacity for non-expert users (Sweller 1988). The interface therefore adopts progressive disclosure, staging information in user-initiated layers. Parameters are grouped through hierarchical tab navigation, with subordinate details appearing only on demand. For dimensional input, slider widgets with real-time annotations replace numerical keypad entry, while preset recommendations anchor users to standard module sizes. This layered approach keeps visible complexity proportional to the current task (Nielsen 1994).

Physically based rendering of anisotropic veneer appearance

Faithful digital reproduction of wood veneer poses a domain-specific challenge largely absent from work on engineered panel materials: the anisotropic optical behavior intrinsic to natural wood grain. The practical importance of this issue was confirmed by the high criticality assigned to "wood grain texture and visual quality" ($CPI = 0.81$). A physically based rendering (PBR) pipeline was adopted, consistent with recent

environment-aware AR rendering frameworks (Ferrão *et al.* 2023), in which multi-channel texture assets—albedo, normal, and anisotropic roughness maps—encode the principal optical properties of each species, with anisotropic roughness capturing reflectance asymmetry parallel and perpendicular to the grain. Image-based lighting (IBL) enables continuous evaluation under representative kitchen illuminants (daylight, warm-white incandescent, cool-white LED), addressing the known sensitivity of perceived wood color to correlated color temperature; for highly anisotropic veneers, recent advances in anisotropic specular IBL based on BRDF major-axis sampling (Cocco *et al.* 2024) provide a practical pathway to faithful highlight reproduction within real-time AR budgets. Kinematic simulation of soft-close hinges, undermount drawer slides, and lift-up supports extends prototype fidelity from static appearance to functional behavior.



Fig. 8. AR spatial scanning and collision warning interface



Fig. 9. Parametric customization operation interface



Fig. 10. High-fidelity rendering and material perception interface

Design Evaluation

A single-group usability testing paradigm was adopted to examine the effectiveness of the interaction strategies and prototype. Because this study targeted information architecture and task flow optimization rather than rendering engine performance, the test employed a high-fidelity simulated prototype based on pre-rendered real-scene imagery. This approach served three methodological purposes: (a) eliminating frame-rate fluctuations and SLAM tracking jitter that would confound task completion time measurements; (b) ensuring that all participants experienced identical visual stimuli, thereby controlling for device-dependent rendering variance; and (c) isolating the contributions of interface layout, progressive disclosure logic, and error-proofing mechanisms to usability outcomes. Simulated prototypes have been used in comparable AR usability studies when the evaluation objective is interaction design rather than system-level rendering performance. Therefore, the test validated interface interaction logic; technical rendering fidelity on real AR hardware requires separate assessment in future work.

Experimental design and participants

A total of 60 participants were recruited, 15 per user group. Ages ranged from 25 to 45 (mean = 32.4, SD = 5.2) years old, with a 1:1 male-to-female ratio. Participants completed four progressive tasks in sequence: T1 (spatial mapping), T2 (module invocation and parameter adjustment), T3 (error-proofing and embodied experience), and T4 (scheme output and decision). Task completion time (TCT) and error-proofing interception rate were recorded throughout. Participants completed the SUS questionnaire immediately after testing. All participants signed informed consent forms and could withdraw at any time.

Evaluation results

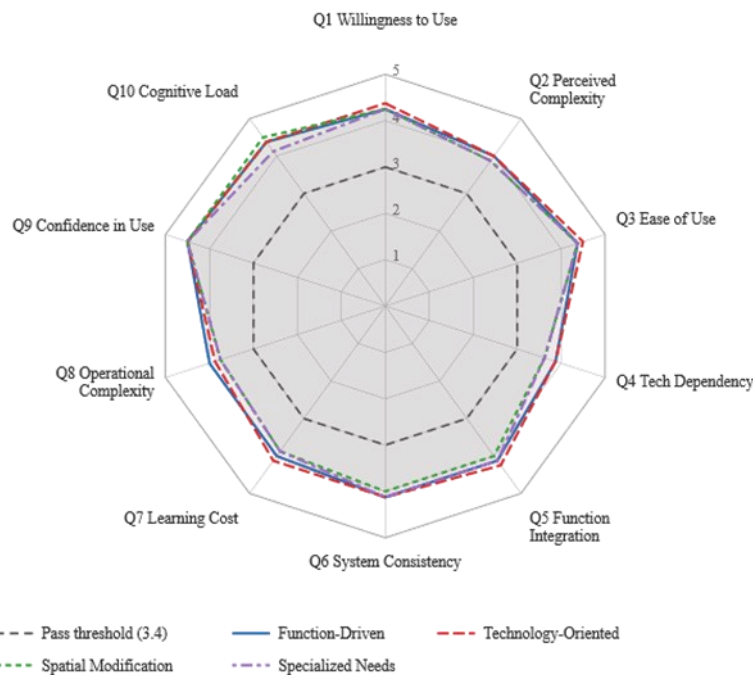
Objective task performance is summarized in Table 2. The 60 participants completed the full task sequence in an average of 215 s. The mean T2 completion time was 68.5 s (SD = 12.4 s), confirming the effectiveness of the cognitive load reduction strategy. In T3, the system achieved a 100% warning trigger and interception rate across all groups, with a mean error recovery time of 3.2 s.

Table 2. Usability Test Results by User Type

| User Typology | N | Mean SUS | SD | TCT (s) | Interception Rate |
|----------------------|----|----------|-----|---------|-------------------|
| Function-Driven | 15 | 83.25 | 4.1 | 218 | 100% |
| Spatial Modification | 15 | 81.50 | 4.8 | 225 | 100% |
| Tech-Oriented | 15 | 84.00 | 3.5 | 198 | 100% |
| Specialized Needs | 15 | 81.25 | 5.0 | 219 | 100% |
| Overall | 60 | 82.50 | 4.4 | 215 | 100% |

Note: SUS = System Usability Scale score (0–100); TCT = Task Completion Time; Interception Rate = percentage of trials in which spatial collision warnings successfully prevented placement errors. Group sizes (N) are equal across the four user typologies (n = 15 per group).

One-way analysis of variance (ANOVA) revealed no significant between-group differences ($F(3, 56) = 1.24, p = 0.303$). The overall SUS score of 82.5 corresponds to "good" on the absolute grading scale of Bangor et al. (2009) and to grade A– on the curved grading scale of Sauro and Lewis (2016), well above the 68-point average usability benchmark. The observed effect size $\eta^2 = 0.062$ was small. Post-hoc power analysis indicated that the sample (n = 60) achieved 0.78 power for detecting a large effect ($f = 0.40$), marginally below the conventional 0.80 threshold; however, because the research hypothesis concerned cross-group equivalence rather than the detection of between-group differences, the non-significant ANOVA result combined with the small effect size already supports the intended conclusion. TOST equivalence testing (equivalence bound $\Delta = 5$ points, selected on the basis that 5 points on the SUS approximates the spacing of one adjective-rating category and lies within the typical 3–5 point standard error of measurement reported for SUS test–retest reliability; Sauro and Lewis 2016; Lewis 2018) further confirmed that all pairwise SUS differences fell within the equivalence interval (TOST $p < 0.05$), providing direct statistical evidence of cross-group usability equivalence independent of power considerations.

**Fig. 11.** SUS sub-dimension score radar chart by user type

Participants rated "operational confidence" (mean 4.6/5.0) and "ease of use" (mean 4.5/5.0) particularly highly across all groups (Fig. 11). Retrospective interviews attributed this to the step-by-step feedforward guidance and perspective grid effects in T1, which reduced unfamiliarity, and the hardware opening/closing animations in T3, which compensated for the absence of tactile feedback and boosted sensory confidence.

Discussion

Interpretation of core themes and the dominance of visual perception

The four core themes identified by BERTopic—functional experience and engagement (34.85%), visual perception and material texture (26.50%), interaction effects (24.68%), and spatial adaptation (13.97%)—reflect a cognitive hierarchy rather than arbitrary topical divisions. Functional engagement ranked highest because it encompasses the foundational pragmatic tasks (module placement, parameter adjustment) that precede any evaluative judgment, consistent with the sequential task-to-evaluation progression in Norman's (2013) action cycle. The unexpectedly high weight of visual perception (26.50%), second only to functional engagement, warrants closer examination. Within this theme, feature-word intersection analysis revealed that approximately 68% of overlapping terms (13 out of 19) were wood-science terminology (grain orientation, glossiness, anisotropy, color temperature sensitivity), confirming that material-specific perceptual properties—not generic 'look good' expectations—drive this requirement cluster. This finding aligns with recent perceptual dimension studies demonstrating that wood is judged along species-distinctive anisotropic axes rather than general color-texture impressions (Filip et al. 2024, 2025), with event-related potential evidence that grain directionality and surface gloss anisotropy produce measurable neural responses distinct from engineered panel materials (Wan et al. 2021), and with empirical observations that color-texture interactions in reconstituted decorative veneer drive distinctive consumer visual responses (Huang et al. 2024). The dominance of these perceptual dimensions in spontaneously generated user feedback also echoes psychological preference findings from heat-treated wood studies (Zhang et al. 2024), which report that surface visual properties exert measurable effects on user evaluation independent of cognitive functional judgments.

Usability performance and cross-group equivalence

The SMC cross-validation confirmed strong convergence between the qualitative mental model and the quantitative topic model, with all core mapping pairs exceeding the $SMC \geq 0.70$ threshold. Interface design strategies derived from CPI were validated by usability testing, yielding an SUS of 82.5 (SD = 4.4)—well above the 68-point usability benchmark (Bangor *et al.* 2009)—with no significant inter-group differences ($F(3, 56) = 1.24, p = 0.303, \eta^2 = 0.062$). The small effect size demonstrated stable usability across user groups. Compared with average SUS scores of 72 to 76 reported for AR retail applications (Javornik 2016), this prototype showed an improvement of approximately 8.6% to 14.6%. Three factors contributed to this gain: the CPI weighting mechanism aligned interface functions with core user demands; feedforward visualization of perspective grids and collision warnings narrowed the "gulf of evaluation" described by Norman; and single-finger slider gestures, replacing complex parameter controls, lowered working memory load—consistent with Sweller's (1988) cognitive load theory.

Anisotropic veneer rendering: from PBR pipeline to species-specific demands

From the wood material digitization perspective, the multi-channel PBR texture design and anisotropic roughness mapping in this prototype provided a viable approach to improving virtual veneer fidelity, addressing the established sensitivity of perceived wood color to light source color temperature and the role of grain directionality in material realism judgments (Filip *et al.* 2024); static renderings, by contrast, cannot respond dynamically to lighting changes and tend to introduce perceptual bias (Wan *et al.* 2021). Different species, however, impose distinct demands on this pipeline. As described in the prototype design section, the three representative species (walnut, white oak, and cherry) each impose distinct demands on roughness-map resolution, subsurface scattering parameterization, and color-temperature modeling. In retrospective interviews, participants rated walnut's large-figure rendering fidelity at 4.4/5.0, whereas white oak's silver-flash reproduction scored 3.6/5.0, a difference confirmed as statistically reliable by a Wilcoxon signed-rank test on the 20 paired ratings ($W = 7.5$, $Z = -3.40$, $p < 0.001$, effect size $r = 0.54$; details in Appendix A.4). The CIELAB $\Delta E^*_{ab} \approx 2$ just-noticeable difference (Buchelt and Wagenführ 2012) provides a quantitative benchmark for future AR color-accuracy validation, and the observed fidelity variation across species supports adopting species-specific texture asset standards calibrated against objective surface characterization metrics such as grain periodicity, specular anisotropy ratio, and CIELAB color coordinates. Recent learning-based methods that infer volumetric annual-ring patterns of solid wood from external surface annotations (Larsson *et al.* 2024) suggest a complementary route toward automated species-specific asset generation when manually authored texture libraries are unavailable. Because the present usability test employed a pre-rendered simulated prototype, the technical fidelity of the PBR textures and IBL lighting on real mobile AR hardware (*e.g.*, frame rate stability, real-time lighting accuracy) remains to be verified.

Comparison with existing requirement mining frameworks

Compared with attribute classification methods, such as the Kano model, widely applied in wood furniture needs analysis (Wang and Chen 2024), the BERTopic pipeline in this study offers complementary strengths. The Kano model classifies predetermined requirement items and provides clear development priorities; BERTopic models unstructured text from multiple sources and can uncover latent requirement structures unanticipated by designers. The two approaches serve different purposes: Kano works best when the requirement space is well defined, BERTopic when requirements must be discovered from large-scale natural language data. In wood furniture digital display research, Chen *et al.* (2023) used Kansei engineering and product semantics to develop a user-centered design framework for custom wood furniture, centering on two-dimensional static color harmony and material preference. Dong *et al.* (2023) applied online reviews and deep learning to evaluate furniture user experience, but their framework did not address the unique influence of wood anisotropic optical properties on user perception. More recently, Zhao *et al.* (2025) integrated KANO theory with an Attention-BiLSTM model to predict user demand evolution from user-generated content, demonstrating that BERTopic-based attribute extraction can be coupled with deep temporal models for dynamic demand forecasting; however, that line of work does not address material-perception fidelity in three-dimensional interactive environments. The present study extended this problem space from two-dimensional static to three-dimensional dynamic AR environments, introducing two material perception dimensions—veneer grain directionality and gloss anisotropy—that are irrelevant in static contexts but critical in interactive three-

dimensional settings. Through BERTopic semantic modeling, the demand weights of these two dimensions were independently confirmed from spontaneous user feedback, providing a user-side methodological supplement to existing literature in BioResources on digital visualization of wood materials.

Limitations and future work

Several limitations should be noted. First, the usability evaluation used a pre-rendered simulated prototype; rendering pipeline fidelity on real AR hardware has not been verified. Future work should pursue system-level validation on real AR devices to confirm that the anisotropic PBR rendering performs under computational constraints. Second, the external corpus (62% of the total) came from general 3D design and AR application reviews, potentially underestimating wood-specific perceptual needs. Although the surrogate corpus retained substantial wood- and material-related terminology (47.3% term coverage; see Appendix A.3), incomplete coverage of cabinet-specific perceptual needs cannot be fully excluded. Collecting user feedback from dedicated wood-veneered cabinet AR applications, once available, would strengthen domain validity. Third, as discussed in the Cross-Validation subsection, the SMC computation shared a single tokenization pipeline; although the permutation test established a random baseline far below the observed values, future work should replicate the analysis with an independent tokenizer (e.g., LTP or THULAC) to quantify any residual pipeline-induced inflation. Fourth, the sample was modest in scale (74 questionnaires, 20 interviews, 60 usability participants) and exclusively drawn from mainland Chinese consumers, limiting cross-cultural generalizability. Fifth, BERTopic is sensitive to corpus scale and cannot track temporal evolution of requirements. Future work should expand the participant pool to include cross-cultural samples and explore dynamic topic modeling to track requirement shifts over time. This study contributes to bridging material perception science and AR interface engineering by integrating anisotropic material properties into user-centered design frameworks.

CONCLUSIONS

A two-stage BERTopic clustering pipeline applied to 3,163 multi-source semantic units identified four core requirement themes—functional experience and engagement (34.85%), visual perception and material texture (26.50%), interaction effects and experience (24.68%), and spatial adaptation and integration (13.97%)—with topic coherence improving by 36.1% after hierarchical merging ($CV = 0.573$). Cross-validation ($SMC \geq 0.70$) confirmed strong alignment between qualitative mental models and quantitative topic structures, highlighting the dominant role of wood material perception, particularly grain directionality and gloss anisotropy, in shaping user requirements. Guided by CPI weighting, visual reproduction requirements were prioritized at P0, and a three-strategy AR interface design framework—explicit spatial mapping, interaction simplification, and anisotropic optical rendering—was developed to address the corresponding material-related challenges. Usability testing of the prototype yielded $SUS = 82.5$ ($SD = 4.4$) with an average task completion time of 215 s and a 100% error interception rate, and showed no significant inter-group differences, indicating robust usability across participants of differing technical backgrounds and customization needs.

The proposed data-driven requirement analysis pipeline and three-tier priority system (P0/P1/P2) thus provide a systematic framework for integrating wood material perception into AR interface design. Because the pipeline is organized around generic stages—corpus construction, topic modeling, cognitive cross-validation, and priority ranking—it can be adapted to other material-aware AR applications beyond cabinetry. Future work will focus on validating the rendering pipeline on real AR hardware and expanding cross-cultural datasets to further improve generalizability.

ACKNOWLEDGMENTS

The authors thank all interview and usability test participants for their time and contributions.

REFERENCES CITED

- Bangor, A., Kortum, P. T., and Miller, J. T. (2009). "Determining what individual SUS scores mean: Adding an adjective rating scale," *J. Usability Stud.* 4(3), 114-123. <https://doi.org/10.5555/2835587.2835589>
- Baytar, F., and Chung, T. (2024). "Augmented reality in the retail environment: A systematic review of consumer experience research," *Int. J. Hum.: Comput. Interact.* 40(16), 4205-4222. <https://doi.org/10.1080/10447318.2023.2212233>
- Braun, V., and Clarke, V. (2021). "One size fits all? What counts as quality practice in (reflexive) thematic analysis?," *Qual. Res. Psychol.* 18(3), 328-352. <https://doi.org/10.1080/14780887.2020.1769238>
- Buchelt, B., and Wagenführ, A. (2012). "Evaluation of colour differences on wood surfaces," *Eur. J. Wood Prod.* 70(1-3), 389-391. <https://doi.org/10.1007/s00107-011-0545-z>
- Chen, Y.-S., Zhu, W.-K., Feng, X.-H., and Wang, Q. (2023). "User-centered design of customized wood furniture based on Kansei engineering and product semantics," *BioResources* 18(1), 1538-1554. <https://doi.org/10.15376/biores.18.1.1538-1554>
- Cocco, G., Zanni, C., and Chermain, X. (2024). "Anisotropic specular image-based lighting based on BRDF major axis sampling," *Comput. Graph. Forum* 43(7), e15233. <https://doi.org/10.1111/cgf.15233>
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). "BERT: Pre-training of deep bidirectional transformers for language understanding," in: *Proceedings of NAACL-HLT 2019*, Vol. 1, Minneapolis, MN, USA, pp. 4171-4186. <https://doi.org/10.18653/v1/N19-1423>
- Dong, M.-Y., Ma, Y.-H., Wang, J., and Li, S. (2023). "Research on furniture user experience evaluation based on online reviews and deep learning," *BioResources* 18(4), 7332-7349. <https://doi.org/10.15376/biores.18.4.7332-7349>
- Ferrão, J., Dias, P., Santos, B., and Oliveira, M. (2023). "Environment-aware rendering and interaction in web-based augmented reality," *J. Imaging* 9(3), Article Number 63. <https://doi.org/10.3390/jimaging9030063>
- Filip, J., Lukavský, J., Děchtěrenko, F., Schmidt, F., and Fleming, R. (2024). "Perceptual dimensions of wood materials," *J. Vision* 24(5), Article Number 12. <https://doi.org/10.1167/jov.24.5.12>

- Filip, J., Děchtěrenko, F., Lukavský, J., Fleming, R. W., and Schmidt, F. (2025). "Comprehensive perceptual analysis and rating of material properties from video data," in: *Pattern Recognition. ICPR 2024 International Workshops and Challenges, Lecture Notes in Computer Science*, vol. 15617, Springer, Cham, pp. 174-188. https://doi.org/10.1007/978-3-031-88217-3_13
- Grootendorst, M. (2022). "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," *arXiv preprint arXiv:2203.05794*. <https://doi.org/10.48550/arXiv.2203.05794>
- Guest, G., Bunce, A., and Johnson, L. (2006). "How many interviews are enough? An experiment with data saturation and variability," *Field Methods* 18(1), 59-82. <https://doi.org/10.1177/1525822X05279903>
- Hassenzahl, M., and Tractinsky, N. (2006). "User experience - A research agenda," *Behav. Inform. Technol.* 25(2), 91-97. <https://doi.org/10.1080/01449290500330331>
- Huang, T., Zhou, C., Wang, X., and Kaner, J. (2024). "A study of visual perception based on colour and texture of reconstituted decorative veneer," *Coatings* 14(1), article 57. <https://doi.org/10.3390/coatings14010057>
- Javornik, A. (2016). "Augmented reality: Research agenda for studying the impact of its media characteristics on consumer behaviour," *J. Retail. Consum. Serv.* 30, 252-261. <https://doi.org/10.1016/j.jretconser.2016.02.004>
- Khirbat, S., and Sriram, M. (2025). "Augmented reality in retail: Exploring the mediating role of attitude on purchase intention in home décor and furnishings," in: *Proceedings of the 2nd International Conference on Sustainable Business Practices and Innovative Models (ICSBPIM-2025)*, Atlantis Press, pp. 1005-1017. https://doi.org/10.2991/978-94-6463-872-1_61
- Landis, J. R., and Koch, G. G. (1977). "The measurement of observer agreement for categorical data," *Biometrics* 33(1), 159-174. <https://doi.org/10.2307/2529310>
- Larsson, M., Ijiri, T., Shen, I.-C., Yoshida, H., Shamir, A., and Igarashi, T. (2024). "Learned inference of annual ring pattern of solid wood," *Comput. Graph. Forum* 43(6), e15074. <https://doi.org/10.1111/cgf.15074>
- Lewis, J. R. (2018). "The System Usability Scale: Past, present, and future," *Int. J. Hum.-Comput. Interact.* 34(7), 577-590. <https://doi.org/10.1080/10447318.2018.1455307>
- Manavis, A., Minaoglou, P., Efkolidis, N., and Kyratsis, P. (2024). "Digital customization for product design and manufacturing: A case study within the furniture industry," *Electronics* 13(13), Article Number 2483. <https://doi.org/10.3390/electronics13132483>
- Manuel, A., Leonhart, R., Broman, O., and Becker, G. (2015). "Consumers' perceptions and preference profiles for wood surfaces tested with pairwise comparison in Germany," *Ann. For. Sci.* 72(6), 741-751. <https://doi.org/10.1007/s13595-014-0427-x>
- Nielsen, J. (1994). *Usability Engineering*, Academic Press, Boston, MA, USA.
- Norman, D. A. (2013). *The Design of Everyday Things: Revised and Expanded Edition*, Basic Books, New York, NY, USA.
- Odstrcil, L., Valent, P., Kaputa, V., and Fabrika, M. (2024). "Digitization and virtualization of wood products for its commercial use," *Forests* 15(12), Article Number 2263. <https://doi.org/10.3390/f15122263>
- Sauro, J., and Lewis, J. R. (2016). *Quantifying the User Experience: Practical Statistics for User Research*, 2nd ed., Morgan Kaufmann, Cambridge, MA, USA.
- Sweller, J. (1988). "Cognitive load during problem solving: Effects on learning," *Cognitive Sci.* 12(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4

- Vicente, K. J., and Rasmussen, J. (1992). "Ecological interface design: Theoretical foundations," *IEEE Trans. Syst. Man Cybern.* 22(4), 589-606.
<https://doi.org/10.1109/21.156574>
- Wan, Q., Li, X., Zhang, Y., Song, S., and Ke, Q. (2021). "Visual perception of different wood surfaces: An event-related potentials study," *Ann. For. Sci.* 78, Article Number 25. <https://doi.org/10.1007/s13595-021-01026-7>
- Wang, Q., and Chen, Y. (2024). "Applying a Kano-FAST integration approach to design requirements for auditorium chairs," *BioResources* 19(3), 5825-5838.
<https://doi.org/10.15376/biores.19.3.5825-5838>
- Zhang, Y., Guo, Y., Wei, P., He, Z., Yi, S., and Zhao, G. (2024). "Effect of changes in surface visual properties of heat-treated wood on the psychological preference," *BioResources* 19(3), 4652-4669. <https://doi.org/10.15376/biores.19.3.4652-4669>
- Zhao, J., Huang, Y., Feng, J., Xie, W., and Jain, K. (2025). "Fusion of KANO theory and Attention-BiLSTM models for user demand analysis and trend prediction," *Inform. Fusion* 122, article 103210. <https://doi.org/10.1016/j.inffus.2025.103210>

Article submitted: April 3, 2026; Peer review completed: April 24, 2026; Revisions accepted: May 4, 2026; Published: May 6, 2026.
DOI: 10.15376/biores.21.3.5785-5807

APPENDIX

Supplementary information supporting the requirement analysis methodology, parameter sensitivity, external corpus composition, and statistical tests reported in the main text.

A.1. Semi-Structured Interview Procedure and Coding Details

Participant composition

Building on the per-group sample summary in the main text (5 per group; 10M/10F; ages 25–52, mean 36.8), the per-group demographic breakdown was as follows: function-driven (n = 5; ages 28–45; 2M/3F; 3 offline / 2 online buyers); spatial-modification (n = 5; ages 25–48; 3M/2F; 2/3); tech-oriented (n = 5; ages 26–40; 3M/2F; 1/4); specialized-needs (n = 5; ages 30–52; 2M/3F; 3/2). Each interview lasted 45–60 min and was conducted at offline home-furnishing stores or via video conferencing. All participants provided informed consent and received a small honorarium.

Interview protocol

The interview guide comprised three sections: (i) background and current practices (10 min) on purchasing experience, pain points with 2D swatches and static renderings, and prior AR exposure; (ii) task-based probes (25–35 min) using a mock-up cabinet configuration scenario with a think-aloud protocol focusing on spatial mapping, material selection, and parameter adjustment; (iii) visual-expectation elicitation (10–15 min) on ideal AR representation of wood veneer under different lighting conditions and grain patterns, supplemented by visual prompts of quarter-sawn and flat-sawn wood samples.

Coding reliability

The training procedure preceding the inter-coder agreement reported in the main text (Cohen's $\kappa = 0.81$, substantial agreement per Landis and Koch 1977) consisted of an initial calibration phase on two transcripts not included in the final analysis, after which the two coders coded all 20 transcripts in parallel. Disagreements were resolved through consensus discussion, and the unified code set was used for subsequent affinity-diagram construction.

Representative coded segments

Examples of open coding (translated from Chinese): "The shelf spacing is fixed, tall items cannot fit in" → fixed_module_constraint (functional pain point); "Designing it myself on my phone feels like I built it with my own hands" → agency_satisfaction (hedonic task unit); "When I tilt the phone, the grain should look different, like how walnut actually behaves" → anisotropic_expectation (visual perception); "I'm afraid the cabinet won't fit after it arrives—I want to see it blocked by the wall here" → spatial_collision_anxiety (pre-placement safety); "The wood color under my kitchen light looks warmer than on the showroom screen" → color_temperature_shift (visual perception).

Saturation assessment

The cumulative new-code curve (Guest et al. 2006) underlying the saturation summary in the main text plateaued at the 16th interview; the final four interviews served as confirmatory rather than generative.

A.2. BERTopic Parameter Sensitivity Analysis

To verify the robustness of the clustering configuration reported in the main text, a systematic sensitivity analysis was conducted on the two most influential parameters: HDBSCAN's `min_cluster_size` and UMAP's `n_neighbors`. All other parameters were held constant at main-text values.

Table A1. Sensitivity of Topic Coherence to `min_cluster_size`

| <code>min_cluster_size</code> | Initial Topics | Noise (%) | Avg CV (pre-merge) | Final Topics | Avg CV (post-merge) | Notes |
|-------------------------------|----------------|-----------|--------------------|--------------|---------------------|----------------|
| 5 | 31 | 12.4 | 0.385 | 6 | 0.512 | |
| 10 | 19 | 16.6 | 0.421 | 4 | 0.573 | Adopted |
| 15 | 14 | 21.3 | 0.398 | 4 | 0.541 | |
| 20 | 11 | 26.8 | 0.376 | 3 | 0.488 | |

Table A2. Sensitivity to UMAP `n_neighbors` (`min_cluster_size` = 10 fixed)

| <code>n_neighbors</code> | Initial Topics | Avg CV (pre-merge) | Cluster Stability | Notes |
|--------------------------|----------------|--------------------|-----------------------------------------------|----------------|
| 10 | 22 | 0.408 | Moderate (2 unstable clusters across 10 runs) | |
| 15 | 19 | 0.421 | High (all clusters reproducible) | Adopted |
| 20 | 17 | 0.405 | High but over-generalized cluster boundaries | |

The configuration `min_cluster_size` = 10, `n_neighbors` = 15 yielded the highest post-merge coherence (CV = 0.573) and the most stable cluster structure across repeated runs. Smaller `min_cluster_size` values produced over-segmentation with semantically redundant clusters, while larger values merged low-frequency but domain-specific topics (e.g., species-specific grain terminology) into generic categories.

A.3. External Corpus Composition and Domain Term Coverage

The external corpus (n = 1,963 semantic units) was collected from online home-renovation communities and application-store reviews between August and November 2025. To assess the extent to which this surrogate corpus captured domain-specific material-perception language, a lexical audit was performed using a curated dictionary of 87 wood- and material-related terms.

Table A3. Wood- and Material-related Term Coverage (% of external semantic units containing ≥ 1 term)

| Term Category (examples) | Coverage (% of units containing ≥ 1 term) |
|--------------------------------------------------------------------------------|------------------------------------------------|
| Grain and texture (e.g., grain, texture, wood pattern, cathedral grain) | 28.4% |
| Color and lighting (color difference, hue, warm/cool light, color temperature) | 19.7% |
| Surface finish (gloss, matte, sheen, finish) | 14.2% |
| Material realism (realistic, fidelity, natural, simulation) | 22.1% |
| Any category, deduplicated | 47.3% |

Approximately 47% of external semantic units contained at least one wood- or material-related term after preprocessing. While this is lower than the 71.6% coverage observed in the internal interview corpus ($n = 1,200$), it confirms that the external corpus carries substantial material-perception signal and is not limited to generic 3D/AR interaction feedback. This mitigates, though does not fully eliminate, the concern that the surrogate corpus may underestimate domain-specific perceptual needs.

A.4. Statistical Supplement: Species-Specific Rendering Fidelity

To evaluate whether the difference in retrospective fidelity ratings between walnut ($M = 4.4$, $SD \approx 0.50$) and white oak silver-flash reproduction ($M = 3.6$, $SD \approx 0.68$) among the 20 interviewees was statistically reliable, a Wilcoxon signed-rank test was performed on the paired ratings. Result: $W = 7.5$, $Z = -3.40$, $p < 0.001$ (exact $p = 0.00067$), effect size $r = 0.54$. The result indicates that participants rated walnut rendering fidelity significantly higher than white oak silver-flash reproduction, supporting the argument in the main text for species-specific texture asset standards.