# Interpretable Feature Mining for AI Product Design


Sophia Sun[1], Vivian Foutz[2], and Icey Xue[2]

[1]Carnegie Mellon University, [2]Western Albemarle High School, [3]University of Science and Technology of China


## Abstract


Generative AI can now produce high-quality visual content, but its role in product design remains unclear. We present a scalable, interpretable framework for AI-driven product design and evaluate it using 4,400 ebook covers. For each cover, we generate four AI designs from different metadata prompts and compare them to human-designed covers using GPT-4o and human evaluators across six dimensions.

Across all dimensions, AI-generated designs consistently outperform human ones. Through feature mining, we identify 22 interpretable attributes that predict evaluative outcomes and explain prompt-specific advantages. Our findings establish Generative AI as an effective, controllable design tool and offer guidance for large-scale product design.


## Introduction

Visual product design plays a central role in shaping how consumers interpret and evaluate digital products. As Generative AI increasingly demonstrates the ability to create high-quality visuals, it is important to understand how they perform on design tasks that require genre alignment, clarity, and value perception (going beyond aesthetics alone). Despite growing use of AI in creative workflows, its performance in real product design settings has not been examined at scale.

To advance this understanding, we develop a large, controlled evaluation using 4,400 human-designed ebook covers and 17,600 AI-generated alternatives produced with four different metadata prompts. Designs are evaluated by both human and GPT-4o and analyzed using 22 interpretable visual features capturing color, geometry, layout, and typography. We center our study on three core questions:

1. Can GenAI match or exceed human designers across multiple perceptual dimensions?
2. How does prompt structure influence large-scale design outcomes?
3. Which measurable visual attributes explain differences between human and AI-generated designs?

This framework provides a quantitative, interpretable foundation for understanding GenAI's capabilities in product design.

Human Design | Title+Caption | Title+Keywords | Title+Synopsis | Title+All

**Figure 1.** Sample book cover with associated four generated AI results.

## Methods and Materials

We analyze 4,400 human-designed ebook covers spanning 20 genres, paired with each book's metadata (title, keywords, synopsis). Using GPT-4 Vision, we generate four AI cover variants per title based on distinct prompting strategies: Title + Caption, Title + Keywords, Title + Synopsis, and Title + All.

Designs are evaluated holistically (liking) and across five value dimensions (functional, emotional, social, price perception, and behavioral intention) using GPT-4o and 189 human raters. All measures show high reliability (Cronbach's $\alpha$ = .81-.98). To interpret design outcomes, we extract 22 interpretable visual features (color, brightness, geometry, layout, typography) and link them to evaluation scores using PCA and regression analysis, allowing for a quantitative explanation of differences in design.

## Results

Across both evaluation methods, AI-generated covers outperform human designs on all dimensions, with the largest gaps in functional clarity and emotional appeal. Prompt structure produces clear differences:

- Title + Keywords has the strongest emotional, social, and motivational responses, suggesting clear genre genre and affective alignment
- Title + Synopsis achieves the highest functional clarity and perceived price value, indicating that richer prompts support more content-relevant imagery
- Title + Caption shows moderate performance across dimensions without a specific strength
- Title + All performs in the mid-range, sometimes improving clarity or cohesion but not consistently outperforming Keywords or Synopsis

Feature mining identifies 22 interpretable visual attributes linked to higher scores. High-scoring AI designs show higher brightness, richer color palettes, stronger centering and symmetry, lower edge density, and cleaner typography, features that appear more in AI designs than in human ones.

**Table 1.** The 22 interpretable visual attributes

| Category | Feature: explanation (method of extraction) |
|---|---|
| Color | (1) color_richness: Count of unique RGB colors after histogram binning (OpenCV).<br>(2) saturation: Mean saturation of RGB → HSV (OpenCV).<br>(3) color_entropy: Shannon entropy of the binned RGB histogram ([27]).<br>(4) color_contrast: Std. dev. of luminance magnitude (Sobel; luminance or per-channel).<br>(5) color_harmony_index: Normalized entropy of HSV-hue histogram, higher = more unified. |
| Texture | (6) edge_density: Proportion of pixels detected as edges by Canny (grayscale → Gaussian blur → Canny; edge pixels=non-zero in output) = edge_pixel_count / total_pixels (OpenCV; [28]). |
| Composition | (7) brightness: Mean pixel value of the grayscale image, 0–255 or 0–1 if normalized (OpenCV).<br>(8) balance_centering: Normalized distance between grayscale intensity centroid and image center (divided by image diagonal; lower = more centered) (OpenCV).<br>(9) symmetry: Horizontal mirror symmetry via SSIM in image and left-right flip (grayscale; higher = more symmetric) (OpenCV + skimage.metrics.structural_similarity). |
| Typography | (10) typography_density: Proportion of image area covered by MSER-detected text regions, higher = denser text (OpenCV; [29]).<br>(11) typography_clarity: Variance of the Laplacian computed on text regions as a sharpness measure, higher = crisper type (OpenCV; [30]).<br>(12) text_contrast: Mean intensity difference between foreground (Otsu-segmented text) and background, higher = stronger contrast (OpenCV; [31]). |
| Text | (13) text_size: Proportion of image area covered by OCR-detected text (union of EasyOCR + PaddleOCR bounding boxes; normalized by image area).<br>(14) text_horizontal_position: Area-weighted mean x-position of text-box centers, normalized to [0,1], 0=left, 1=right.<br>(15) text_vertical_position: Area-weighted mean y-position of text-box centers, normalized to [0,1], 0=top, 1=bottom. |
| Image-Text Alignment | (16) sim_title: Cosine similarity between CLIP text embedding of the title and CLIP image embedding of the cover (L2-normalized).<br>(17) sim_keywords: Cosine similarity between CLIP text embedding of keywords (mean-pooled) and the cover's CLIP image embedding.<br>(18) sim_synopsis: Cosine similarity between CLIP text embedding of the synopsis (truncated to model token limit) and the cover's CLIP image embedding. |
| Shape | (19) shape_complexity: Fraction of detected shapes that are non-standard, not {triangle, rectangle/square, circle/ellipse} (OpenCV contours + polygon approximation / Hu moments).<br>(20) geometric_vs_organic: (rectangles + triangles) / max(circles + irregular + 1), higher = more geometric (counts from OpenCV contour classification). |
| Human Face | (21) face_density: Faces detected (OpenCV Haar/DNN) divided by total geometric shapes ≥1, higher = more faces per shape (contour classification). |
| Emotion | (22) emotion_value: Overall image sentiment (−1 = negative, 1 = positive). (ensemble ResNet + HuggingFace + CLIP on the full image; model logits → softmax probabilities for positive/neutral/negative, then a weighted average yields the final score). |

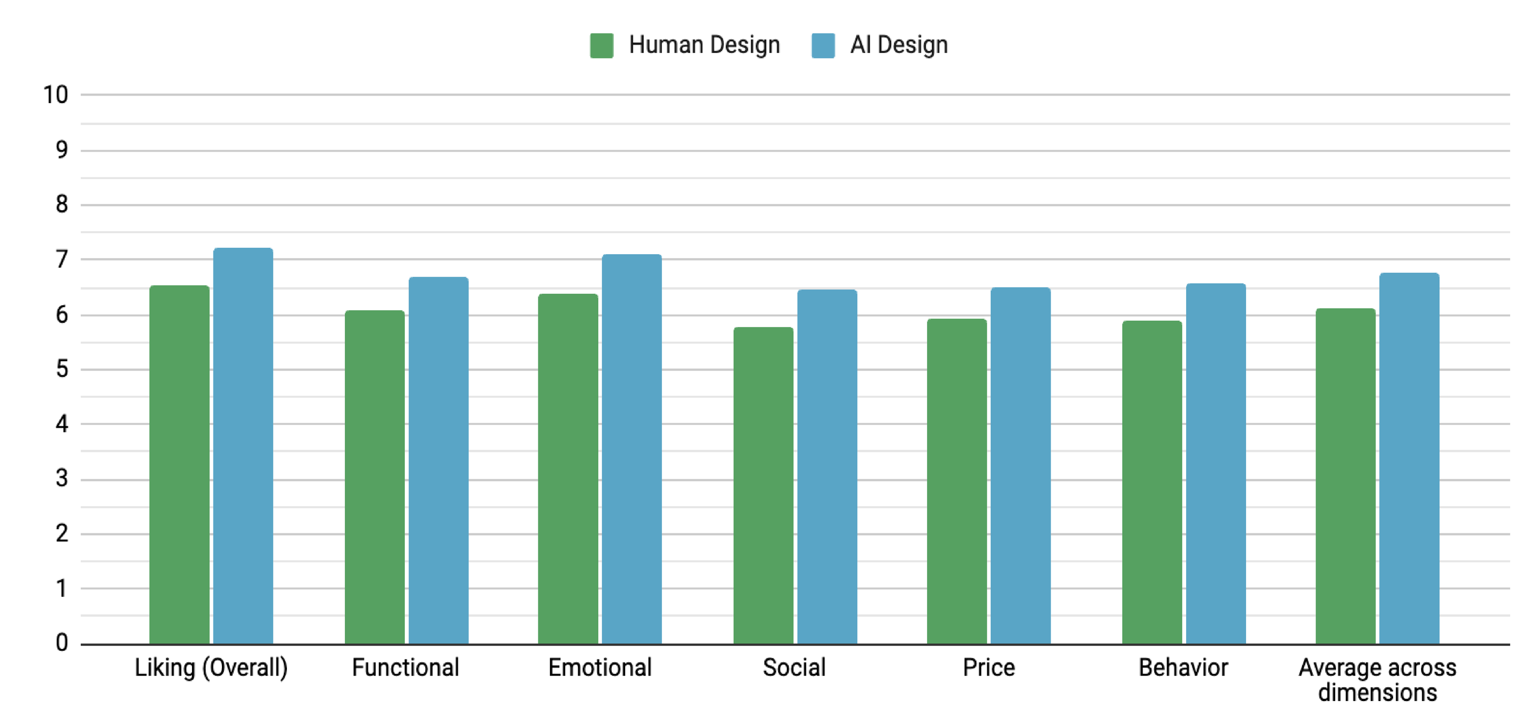
## Discussion

Our findings show that GenAI can reliably produce visual designs that align with perceptual cues linked to clarity, professionalism, and emotional engagement. Metadata-rich prompts, especially keywords and synopses, shape the structure of AI outputs, indicating that prompt formulation is an effective mechanism for targeting different design goals. The consistent advantage of AI across a large and diverse dataset suggests these strengths generalize across genres and styles.

Interpretable feature analysis further explains why AI designs perform well. High-scoring covers show balanced composition, rich but controlled colors, strong symmetry, reduced clutter, and clean typography, attributes closely related to the functional and affective dimensions used in evaluation and expressed more consistently in AI than in human designs.

These insights support the use of AI in systematic, data-driven design workflows and highlight opportunities for hybrid processes where prompt design and feature-level guidance shape large-scale visual creation.

**Chart 1.** AI vs Human designs across dimensions. Bars are mean scores from AI and human raters (0-10).




## Conclusions

AI-generated covers outperform human designs across all six dimensions, showing strong functional clarity, emotional resonance, and perceived value. Prompt structure plays an important role: keyword-based prompts enhance affective appeal, while synopsis-based prompts improve clarity and perceived value. The 22 interpretable features explain these effects by defining specific attributes (brightness, color richness, symmetry, reduced clutter, and clean typography) that predict higher evaluations.

Together, these findings position GenAI as an effective, controllable tool for large-scale product design and highlight the potential for hybrid human-AI workflows, where designers guide prompt strategy and feature-level decisions while AI provides rapid, high-quality visual generation at scale.

## Contact


Sophia Q. Sun
Carnegie Mellon University
sqsun@andrew.cmu.edu
http://linkedin.com/in/sophia-qiyu-sun
(434) 328-0986


## References


1. T. Hagendorff, "Mapping the Ethics of Generative AI: A Comprehensive Scoping Review," 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2402.08323
2. Statista, "Forecast of Digital Media revenue by segment in the World from 2019 to 2027," 2024. [Online]. Available: https://www.statista.com/forecasts/456462/digital-media-revenue-in-the-world-forecast
3. Y. Du, S. Li, A. Torralba, J. B. Tenenbaum, and I. Mordatch, "Improving factuality and reasoning in language models through multiagent debate," 2023
4. Y. Liu, K. J. Li, H. A. Chen, and S. Balachander, "The Effects of Products' Aesthetic Design on Demand and Marketing-Mix Effectiveness: The Role of Segment Prototypicality and Brand Consistency," Journal of Marketing, vol. 81, no. 1, pp. 83–102, 2017
5. N. Mandel and E. J. Johnson, "When Web Pages Influence Choice: Effects of Visual Primes on Experts and Novices," Journal of Consumer Research, vol. 29, no. 2, pp. 235–245, 2002.
6. V. Shankar, "How Artificial Intelligence (AI) is Reshaping Retailing," Journal of Retailing, vol. 94, no. 4, pp. 343–418, 2018.
7. S. Puntoni, R. W. Reczek, M. Giesler, and S. Botti, "Consumers and Artificial Intelligence: An Experiential Perspective," Journal of Marketing, vol. 85, no. 1, pp. 131–151, 2020.
8. J. De Freitas, G. Nave, and S. Puntoni, "Ideation with Generative AI—in Consumer Research and Beyond," Journal of Consumer Research, vol. 52, no. 1, pp. 18–31, 2025.
9. D. Grewal, C. B. Satornino, T. H. Davenport, and A. Guha, "How Generative AI is Shaping the Future of Marketing: A Four-Quadrant Framework," Journal of the Academy of Marketing Science, vol. 53, no. 3, pp. 702–722, 2025.
10. L. I. Labrecque, S. Sohn, B. Seegebarth, and C. Ashley, "Color Me Effective: The Impact of Color Saturation on Perceptions of Potency and Product Efficacy," Journal of Marketing, vol. 89, no. 4, pp. 120–139, 2025.