Proceedings of the Design Society, Volume 5: ICED25
https://doi.org/10.1017/pds.2025.10346

Unlocking product ecosystem insights: analyzing customer sentiment and interoperability through video reviews

Kangcheng Lin and Harrison Kim✉

University of Illinois at Urbana-Champaign, Illinois, USA

✉ hmkim@illinois.edu

ABSTRACT: This paper introduces a novel methodology for analyzing customer preferences within product ecosystems by leveraging video reviews from social media platforms. The approach includes three stages: collecting and preprocessing video reviews, extracting product features using Latent Dirichlet Allocation (LDA), and analyzing sentiment with the VADER package. By utilizing video reviews, this study captures a more detailed and structured understanding of customer experiences compared to traditional textual reviews, offering actionable guidance for product interoperability and user sentiment analysis. The research highlights the importance of understanding the relationships between products and their accessories, providing specific design insights for creating cohesive product ecosystems that resonate with users on both functional and emotional levels.

KEYWORDS: data mining, product ecosystem, topic modeling, sentiment analysis, interoperability

1. Introduction

The notion of a product ecosystem has become pivotal in contemporary product design, emphasizing the interplay of interconnected products that collectively enhance the overall user interaction and satisfaction. Rather than focusing on isolated functionalities, product ecosystems integrate hardware, software, services, and user communities to deliver cohesive and value-driven experiences (Zhou et al., 2011). For example, the Apple iPhone exemplifies this paradigm, combining devices, apps, retail services, and developer networks into a unified system that amplifies its appeal and utility. This approach refects a shift from prioritizing individual product features to fostering synergistic experiences across interconnected products. However, designing such ecosystems requires a nuanced understanding of customer preferences, as successful ecosystems must anticipate and address diverse user needs, contexts, and expectations. Insights into customer behaviors and preferences play a critical role in crafting these ecosystems, ensuring that they resonate with users on both functional and emotional levels.

Understanding customer preferences has always been a cornerstone of decision-making in engineering design. Historically, surveys and questionnaires were the primary tools for gathering customer feedback. However, these traditional methods often proved to be slow, costly, geographically limited, and susceptible to bias. In recent decades, online reviews have emerged as a valuable alternative for capturing customer preferences, receiving significant attention in research (Tuarob and Tucker, 2014; Zhou et al., 2020; Jin et al., 2021). Among the most prominent sources of online reviews are e-commerce platforms such as Amazon and eBay. With the rise of a new generation of internet users, however, e-commerce websites are no longer the primary outlets for customers to share their opinions about products. Social media platforms such as YouTube, Facebook, and TikTok have grown significantly in prominence, offering expansive reach and enabling customers to voice their perspectives through various content formats, including video reviews, unboxing experiences, and product comparisons.

Video reviews offer several key advantages over traditional textual reviews often found on e-commerce websites. First, they are often more structured and comprehensive. Textual reviews can sometimes lack depth, as customers might highlight only a single feature-whether positive or negative-without providing a holistic account of their experience (e.g., “The autofocus is incredibly fast and accurate, making it perfect for capturing fast-moving subjects.”). Second, video reviews provide broader customer representation. For instance, reviews for new iPhones may be underrepresented on Amazon, since these products are typically unavailable there. In contrast, social media platforms are accessible to anyone, regardless of where they purchased their product. Finally, video reviews encourage greater interaction through comment sections, creating opportunities for deeper discussions and providing more nuanced insights into customer opinions.

This paper proposes a streamlined process for analyzing customer reviews in video format to evaluate the interoperability between a product and its accessories within a product ecosystem. While video content includes both visual and auditory information, this study focuses exclusively on the textual information derived from audio transcriptions. Future work could explore multimodal analysis incorporating visual cues to capture additional insights, such as user gestures or product demonstrations. The proposed methodology consists of three key stages: (i) collecting and preprocessing video reviews sourced from social media platforms, (ii) extracting product features of interest to customers using Latent Dirich-let Allocation (LDA) models, and (iii) analyzing the sentiment associated with these features using the Valence Aware Dictionary and Sentiment Reasoner (VADER) package. The main contributions of this paper are listed below:

Leveraging Video Reviews as a Novel Data Source: This study introduces video reviews as a rich and dynamic alternative to traditional textual reviews commonly found on e-commerce platforms. Video reviews offer a more comprehensive representation of customer experiences. Furthermore, their interactive nature-often featuring authentic product demonstrations and fostering discussions in comment sections-enables deeper insights into customer opinions and preferences, which are less accessible through static textual reviews.

Investigating Interoperability in Product Ecosystems: The research addresses the critical issue of interoperability within product ecosystems by examining the relationships between a product and its accessories. Through the analysis of video reviews and their associated comments, this study captures nuanced interactions between users and interconnected products, uncovering user sentiment and functional dependencies. Functional dependencies in the context refer to the extent to which a product’s functionality is contingent upon the presence or performance of its associated accessories, such that a strong functional dependency implies that an accessory significantly enhances or enables the core product’s intended use. Unlike traditional reviews, this approach offers a richer understanding of how interoperability impacts customer satisfaction and purchasing decisions, providing actionable insights for ecosystem-focused product design.

The remainder of this paper is structured as follows. Section 2 reviews related works and key concepts pertinent to the study. Section 3 provides a detailed explanation of the proposed methodology and experimental design. Section 4 presents the data, implementation, and results of a case study demonstrating the application of the methodology. Finally, Section 5 summarizes the findings, discusses future research directions, and concludes the paper.

2. Literature Review

In this section, we will present four main topics related to the paper, namely *Customer Preference Elicitation, Product Ecosystem, Topic Modeling*, and *Sentiment Analysis*.

2.1. Customer Preference Elicitation

In recent years, researchers have increasingly focused on leveraging online reviews to enhance product design. Traditional methods, such as surveys and questionnaires, often face challenges related to scalability, cost, and susceptibility to temporal and geographical biases. In contrast, approaches that utilize online reviews offer a cost-effective and scalable alternative, potentially providing more representative insights (Yang et al., 2019). For example, Chen et al. (2013) introduced analytical discrete choice models to better understand customer preferences and predict purchasing behavior. Tuarob and Tucker (2015) proposed a rule-based approach for feature extraction, employing pre-defined rules and seed features to analyze customer feedback. Beyond traditional e-commerce platforms such as Amazon, customers increasingly express their opinions through diverse channels, including YouTube. Recognizing this shift, Lin and Kim (2023) proposed a four-stage methodology for analyzing video reviews, encompassing feature extraction, sentiment analysis, and feature importance computation, thereby demonstrating their potential as a rich and reliable resource for understanding customer preferences in engineering design.

2.2. Product Ecosystem

Beyond individual products, a holistic approach examines their interactions within an ecosystem, where interoperability and complementary relationships shape user experience and purchasing decisions. A product ecosystem centers on a core product, supported by complementary products and services, to create a seamless experience that outperforms more fragmented alternatives (Zhou et al., 2011). Notable examples include the ecosystems of Apple and Amazon, where customers typically enter by purchasing hardware such as an iPhone or a Kindle. Consequently, researchers have started investigating innovative ways to harness the potential of product ecosystems. Zhou et al. (2011) explored key challenges in product ecosystem design and introduced a conceptual model to identify critical factors and mechanisms that influence user experience. Zhou et al. (2020) proposed a machine-learning approach to analyze customer needs within product ecosystems by examining user-generated reviews. It combined fastText for fIltering, latent Dirichlet allocation (LDA) for topic modeling, rule-based sentiment analysis for sentiment and intensity prediction, and an analytic Kano model to categorize customer needs based on sentiment analysis results.

2.3. Topic Modeling

Extracting insights from user-generated content is key to analyzing customer experiences in product ecosystems. Topic modeling identifies product attributes and themes in large-scale reviews, structuring unstructured feedback effectively. Among the prominent algorithms applied in text analysis are latent semantic analysis (LSA), non-negative matrix factorization (NMF), probabilistic latent semantic analysis (PLSA), and latent Dirichlet allocation (LDA) (Kherwa and Bansal, 2019).

Latent semantic analysis, introduced by Landauer and Dumais in the 1990s (Deerwester et al., 1990), is an algebraic method that uses singular value decomposition. It has been applied extensively in fields such as information retrieval, natural language processing, and modeling human language knowledge (Buckley et al., 1994; Kherwa and Bansal, 2017). NMF and PLSA are both dimensionality reduction techniques: NMF, initially proposed for environmental data (Paatero and Tapper, 1994), has since been adapted to areas like cancer identification using molecular gene expression data (Lee and Seung, 1999). PLSA employs a probabilistic framework and the bag-of-words approach to identify semantic co-occurrence of terms within a corpus (Hofmann, 2013). LDA, introduced by Blei et al. (2003), is a generative statistical model that identifies the distribution of topics within a corpus and associates each topic with specific word clusters. It has been widely adopted in diverse applications, including e-commerce (Zhou et al., 2020; Joung and Kim, 2021). Among these methods, LDA stands out for its flexibility and adaptability across various datasets, making it a preferred choice in many research contexts.

2.4. Sentiment Analysis

Topic modeling identifies key themes in reviews, but it lacks sentiment context. Sentiment analysis complements it by quantifying user sentiment on product features, offering deeper insights into satisfaction and dissatisfaction. As a rapidly evolving field, numerous algorithms have been developed to address various challenges in sentiment extraction (Yan-Yan et al., 2010; Kang et al., 2012; Hutto and Gilbert, 2014). These methods generally fall into two categories: unsupervised (e.g., lexicon-based) (Zagibalov, 2010; Augustyniak et al., 2015) and supervised learning techniques (Gonçalves et al., 2013; Vilares et al., 2017). While supervised methods tend to offer higher accuracy within specific domains, unsupervised methods are advantageous for their lower memory complexity and faster processing times (Mukhtar et al., 2018).

In recent years, sentiment analysis has increasingly been applied to extract customer preferences from online reviews. For example, Jiang et al. (2017) proposed a method using a fuzzy time series model to predict the importance of future product features. Suryadi et al. Suryadi and Kim (2018) employed sentiment analysis along with word embedding and a dependency tree to analyze the relationship between online reviews and sales rank. Bag et al. (2019) developed a framework incorporating the social perception score of a brand and review polarity to predict customer purchase intentions.

3. Methodology

This section outlines the methodology employed in this study, which comprises four key stages: data collection and preprocessing, identification of features of interest, feature sentiment analysis, and quantification of interoperability. While data collection and preprocessing involve some manual work-such as verifying transcriptions and filtering irrelevant terms-subsequent stages are fully automated. Specifically, topic modeling with LDA and sentiment analysis with VADER require no manual intervention. However, qualitative interpretation of results remains a necessary step to ensure meaningful insights for product design.

3.1. Data Collection and Preprocessing

Video reviews of products can be sourced from popular social media platforms, including YouTube, Facebook, and TikTok. The collected data encompass various elements such as video titles, view counts, release dates, durations, comments, and the videos themselves. Each video comprises two primary components: visual and audio. For this study, we focus exclusively on the English-language audio component, which is transcribed into textual data.

Following established methodologies in the literature (Suryadi and Kim, 2018, 2019; Joung and Kim, 2021; Park et al., 2025), the transcribed text undergoes preprocessing, including punctuation removal, emoji stripping, and conversion of all characters to lowercase (Denny and Spirling, 2018). Subsequently, nouns and noun phrases are extracted from the processed text.

Not all extracted nouns and noun phrases are relevant to the analysis. Some may pertain to unrelated concepts (e.g., YouTube channel or subscription), while others may be overly generic, offering no specific insight into product attributes (e.g., Sony or Camera). To refine the dataset, the extracted terms are cross-referenced with relevant product manuals (Suryadi and Kim, 2018, 2019; Park et al., 2025), which can be obtained from manufacturers' official websites or e-commerce platforms. Words not found in the product manuals are excluded, while those that match are retained as product-specific keywords.

3.2. Identification of Features of Interests

In this stage, product features of interest to customers are identified from the product keyword list generated in the previous stage. To achieve this, Latent Dirichlet Allocation (LDA)-a probabilistic topic modeling technique designed to uncover hidden topics within large textual datasets-is employed. LDA operates using a generative statistical model that categorizes all product reviews into a set of common topics (Blei et al., 2003). Each review is represented as a probabilistic mixture of topics, and each topic is described by a probabilistic set of keywords. For example, a camera review may be 70% about 'image quality' and 30% about 'battery life.' Each topic, in turn, is described by a distribution of keywords. The number of topics is determined using a topic coherence metric, and the LDA output is a topic-keyword matrix. Topics are labeled based on their associated keywords and representative reviews, with these labels corresponding to product features of customer interest (Jeong et al., 2019; Zhou et al., 2020; Park et al., 2025).

Once the product features are identified, their associated keywords are expanded by incorporating synonyms. This study employs word embedding techniques for synonym extraction (Mikolov et al., 2013). Initially, feature-relevant keywords are selected from the top 30 nouns within each topic. Subsequently, the top 20 most similar words for each selected keyword are identified based on word vector similarity. The union of these expanded sets forms the comprehensive list of feature-related keywords.

3.3. Feature Sentiment

This study employs an unsupervised approach to sentiment analysis, which involves identifying target words and assigning sentiment indices to them. The feature keywords identified in Section 3.2 are used as the target words. An unsupervised method is preferred due to its efficiency and speed, as it eliminates the need for labeled training data. Specifically, VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto and Gilbert, 2014), a lexicon- and rule-based sentiment analysis model, is utilized to assess customer sentiment toward product features. VADER's reliance on predefined lexicons makes it broadly applicable across various products and domains without requiring manual annotation. VADER calculates both the polarity and intensity of a given sentence. To address the possibility of repeated references to the same feature within a review, this study computes the average sentiment score for each feature across all relevant sentences. For instance, if a reviewer mentions "viewfinder" multiple times in a video, the sentiment score for "viewfinder" is calculated as the average of the scores derived from all corresponding sentences. Sentiment scores are computed for every product feature mentioned in the review, while features not referenced are assigned no sentiment score.

3.4. Quantification of Interoperability

In this paper, interoperability is defined as the ability of a product to seamlessly integrate, communicate, or work with other products, systems, or components, whether they are from the same manufacturer or different brands. Interoperability is a cornerstone of product ecosystems as it enhances user experience, broadens functionality, and fosters customer loyalty by enabling diverse products to complement one another effectively.

The distinctive nature of video reviews and their associated comments provides an opportunity to uncover customer opinions regarding products and their accessories. However, not all comments carry the same weight; comments with a higher number of likes (a voting mechanism prevalent on social media platforms) are considered more representative of community consensus. To address this, we introduce a weighting mechanism, as expressed in Equation 1, which incorporates the number of likes a comment has received and adjusts for other influential factors, including the total views of the video and the time elapsed since the comment was posted.

$$w_i = \frac{\log(L_i + 2)}{\log(L_{\max} + 2)} \times \frac{(\frac{L_i+1}{V})^{\alpha}}{(\frac{L_i+1}{V})^{\alpha}_{\max}} \times \frac{[e^{-\lambda T}(L_i + 1)]^{\beta}}{[e^{-\lambda T}(L_i + 1)]^{\beta}_{\max}} \tag{1}$$

In Equation 1, we have

- L_i = Number of likes the *i-th* comment receives.
- V = Total number of views of the video.
- T = Time (in days) between when the comment was posted and when it was data-mined.
- $L_{\max}$ = Maximum number of likes any comment has received in the dataset.
- $(\frac{L_i+1}{V})_{\max}$ = Maximum value of $\frac{L_i+1}{V}$ in the dataset.
- $(e^{-\lambda T}(L_i + 1))_{\max}$ = Maximum value of $e^{-\lambda T}(L_i + 1)$ in the dataset.
- α,β = Scaling exponents between 0 and 1 to control the influence of each factor.
- λ = A positive decay constant.

Equation 1 comprises three components. The first prioritizes comments with more likes, refecting community consensus while ensuring that a comment with 1 like and a comment with 200 likes will not have weights that are orders of magnitude apart. The second adjusts for video view counts to avoid unfairly penalizing comments on less-viewed videos. The third accounts for the time a comment has been posted, as longer durations increase the likelihood of receiving likes. Building upon this weighting framework, we quantify the interoperability between two products, A and B, using Equation 2.

$$I(A,B) = \rho(S_{\bar{a}}, S_{\bar{b}}) + \frac{1}{n_{A,B}}\sum_{i=1}^{n_{A,B}} w_i(S_{i,\text{general}} - \frac{1}{|\Omega_{A,B}|}\sum_{(a_j,b_j)\in\Omega_{A,B}} |S_{a_j} - S_{b_j}|) \tag{2}$$

In Equation 2, we have

Table 1. Video data and comments by products.

Product	# Videos	# Comments	# Views
Camera	620	110,403	66,427,823
Lens	106	12,275	11,703,001
Flash	134	6,339	8,374,917
Microphone	43	2,227	1,928,012

- $n_{A,B}$ = Total number of comments that mention both A and B.
- $S_{i,\,general}$ = Average sentiment score of *i-th* comment (potentially consisting of multiple sentences) that mentions both A and B.
- $\Omega_{A,B}$ = Set of common product attributes.
- S_{aj}, S_{bj} = Sentiment scores of A and B, respectively, on the common product attribute.
- $S_a = (S_{a_1}, S_{a_2}, \ldots, S_{a_m})^T$, where m = Total number of occurrences that common product attributes of A (common with B) have been mentioned in comments.

In Equation 2, the interoperability score has two components. The first, $\rho(S_{\tilde{a}}, S_{\tilde{b}})$, calculates the Pearson correlation between the sentiment scores of A and B on shared product attributes, with higher correlation indicating greater interoperability. The second adjusts the average sentiment score of comments mentioning both A and B using $\frac{1}{|\Omega_{A,B}|} \sum_{(a_j,b_j)\in \Omega_{A,B}} |S_{a_j} - S_{b_j}|$, where greater disagreement in sentiment scores reduces the weighted aggregation. This accounts for long comments that, despite a positive average sentiment, highlight mismatches between A and B.

4. Case Study

This section presents a case study using video reviews, as well as comments extracted from corresponding comment sections, of mirrorless cameras, camera lenses, camera fashes, and camera microphones from Youtube.

4.1. Data Collection

This study collected two data types: online video reviews and product manuals. YouTube video reviews were sourced using a standardized search pattern *(Brand + Model + 'Reviews'),* yielding 620 mirrorless camera reviews (avg. 10.4 min per video) from Canon, Fujifilm, Nikon, and Sony. Accessories (lenses, fashes, microphones) were gathered similarly *(Brand + Model + Accessory + 'Reviews'),* resulting in 106 lens, 134 fash, and 43 microphone reviews. While accessories were categorized under camera brands, they were not necessarily manufactured by them.
In addition to video reviews, comments were collected, totaling 110,403 for cameras, 12,275 for lenses, 6,339 for fashes, and 2,227 for microphones. Table 1 presents detailed statistics. Audio components were transcribed using Python's *OpenAI Whisper* package.
For product manuals, documents for 27 cameras, 12 lenses, 8 fashes, and 10 microphones were obtained from manufacturers' official sources. Terms found in video reviews but absent in manuals were excluded to ensure product relevance.

4.2. Features of Customer Interests

Table 2 summarizes the LDA-generated topics for mirrorless cameras. The first column lists topic labels, representing product attributes discussed by reviewers. The second column provides the feature-relevant keywords, filtered to exclude terms not present in the product manual documents. The third column indicates the total number of videos referencing each product attribute. The final column shows the percentage of videos (out of the total 620 videos) that mention each product attribute. For simplicity, only the product attributes of camera accessories are displayed in Table 3.

Table 2. Mirrorless camera attributes of customer interest.

Topic label	Keywords	# Videos	Ratio
Autofocus	autofocus, speed, detection, accuracy, tracking, etc.	616	99.4%
Battery Life	battery, life, longevity, charge, capacity, etc.	617	99.5%
Connectivity	connectivity, wi-fi, bluetooth, usb, hdmi, etc.	465	75.0%
Durability	build, shell, material, construction, finish, etc.	559	90.2%
Image Quality	hue, sharpness, clarity, gradient, noise reduction, etc.	605	97.6%
Portability	portability, size, weight, grip, mobility, etc.	609	98.2%
Viewfnder	viewfinder, evf, lcd, screen, tilt, etc.	598	96.5%
Video Performance	video, frame, motion, bitrate, playback, etc.	619	99.8%

4.3. Feature Sentiment

This subsection examines the sentiment scores derived from video reviews and their corresponding comments, uncovering inconsistencies that are often overlooked in traditional customer reviews.

4.3.1. Sentiment Scores from Video Reviews

Figure 1 presents the sentiment scores of mirrorless cameras and their accessories, including lenses, fashes, and microphones, extracted from video reviews. The radar diagrams illustrate sentiment distributions across key product attributes. The results indicate that while mirrorless cameras and lenses exhibit relatively uniform sentiment scores across brands, fashes and microphones display greater variation. This variability may stem from differences in interoperability, build quality, or compatibility, highlighting potential areas for product improvement. For camera fashes, the attributes of power, build quality, and light quality stand out positively for models compatible with Fujiflm cameras. This observation is statistically validated using the rank-sum test (α = 0.01). Conversely, microphones compatible with Fujifilm cameras exhibit significantly lower sentiment scores for build quality.

Table 3. Camera accessory attributes of customer interest.

Lens Topic label	Flash Topic label	Microphone Topic label
Aperture	Build Quality	Audio Quality
Autofocus	Ease of Use	Build Quality
Build Quality	Light Quality	Connectivity
Focal Length	Power	Directionality
Image Quality	Recycling Time	Portability

4.3.2. Sentiment Scores between Video Reviews and Corresponding Comments

A key distinction between video reviews on social media platforms and traditional textual customer reviews on e-commerce websites lies in the interactive nature of video reviews, facilitated by comment sections. Figure 2 compares sentiment scores derived from video reviews and their corresponding comments, revealing differences in sentiment expression between content creators and audiences. The results suggest that video reviews generally exhibit more positive sentiment, while comments tend to be more neutral or critical. This discrepancy indicates that while reviewers may emphasize product strengths, audience feedback often provides a more balanced perspective, helping to validate the robustness of video-based sentiment analysis.

4.4. Interoperability

Out of the 110,403 comments from mirrorless camera reviews, we have extracted 5,791 comments that discuss mirrorless camera and at least one of its accessories. By applying Equations 1 and 2, we calculated the interoperability between cameras and their accessories across various brands. Figure 3 illustrates the interoperability scores between mirrorless cameras and their accessories across different brands. Figure 3(a) quantifies interoperability, while Figure 3(b) visualizes the results using radar diagrams for comparative analysis. Higher interoperability scores indicate stronger integration between cameras and their accessories, leading to a more cohesive user experience. In contrast, lower scores suggest potential compatibility challenges that may impact usability and customer satisfaction. These findings provide actionable insights for manufacturers seeking to enhance product ecosystem coherence and improve cross-brand compatibility.

5. Conclusion

This paper presents a streamlined methodology for assessing the interoperability of a product and its accessories using video reviews. The methodology is applied to cameras and their accessories, illustrating how video reviews and comments can be leveraged to holistically understand customers' preferences. Results reveal that while mirrorless cameras from different brands exhibit similar sentiment patterns (Figure 1), their interoperability varies significantly (Figure 3). The insights derived from this

Figure 1. Sentiment scores of mirrorless camera and its accessories extracted from video reviews

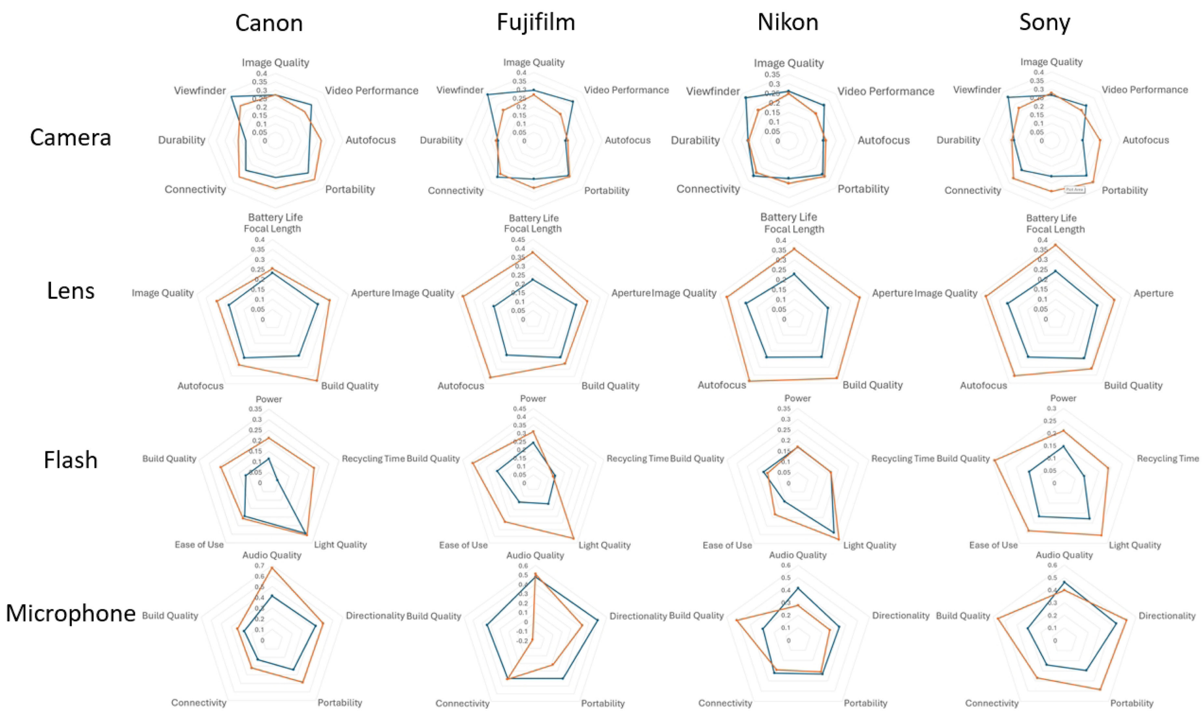


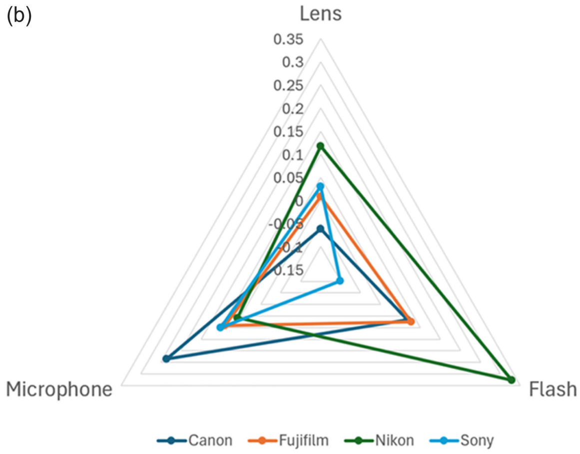
Figure 2. Compare sentiment scores between video reviews and comments

(a)

Camera Brand	Lens	Flash	Microphone
Canon	-0.062	0.067	0.24
Fujifilm	0.0086	0.076	0.091
Nikon	0.12	0.33	0.058
Sony	0.030	-0.10	0.10

Interoperability Scores

(b)



Radar Diagram

Figure 3. (a). Interoperability between camera and its accessories across brands; (b). Corresponding diagrams

study can provide valuable guidance for product design by highlighting key user sentiments and interoperability trends within product ecosystems. By analyzing sentiment distributions across different product features, designers can identify areas of customer satisfaction and dissatisfaction, enabling targeted improvements in future product iterations. Additionally, the interoperability analysis offers critical information on how well a product integrates with its accessories, helping manufacturers refine compatibility and optimize ecosystem coherence. For instance, if certain camera accessories receive consistently negative sentiment regarding connectivity or usability, this signals a need for improved design interventions, such as standardized interfaces or enhanced cross-brand compatibility. While the study of video reviews is still emerging, it has several limitations:

Negative Comments - In this study, we developed a weighting mechanism based on the number of likes each comment receives, a common social media voting system. However, YouTube hiding dislike counts prevents us from identifying controversial comments that may receive an equal number of dislikes and likes. Expanding the study to include other social media platforms may help address this limitation.

Contextual Information - Many discussions of interoperability are context-dependent. For instance, "*For vloggers, the camera offers sharp video, while the microphone ensures noise-free audio, making them a great duo for content creation.*" Incorporating usage context into the methodology could enhance the relevance of interoperability scores for product design.

References

Chen W., Hoyle C. and Wassenaar H. J. (2013). *Decision-based Design: Integrating Consumer Preferences into Engineering Design.* Springer, London.

Tuarob S. and Tucker C. S. (2015). Quantifying product favorability and extracting notable product features using large scale social media data. *Journal of Computing and Information Science in Engineering,* 15(3).

Lin K. and Kim H. (2023). Investigate customer preferences using online video reviews - preliminary results. *Proceedings of the 49th Design Automation Conference (DAC), International Design Engineering Technical Conferences and Computers and Information in Engineering Conference,* Volume 3A, V03AT03A022.

Yang B., Liu Y., Liang Y. and Tang M. (2019). *Exploiting user experience from online customer reviews for product design. International Journal of Information Management,* 46, 173–186.

Zhou F., Xu Q. and Jiao R. J. (2011). Fundamentals of product ecosystem design for user experience. *Research in Engineering Design,* 22, 43–61.

Zhou F., Ayoub L, Xu Q. and Yang J. X. (2020). A machine learning approach to customer needs analysis for product ecosystems. *Journal of Mechanical Design,* 142(1), 011101.

Yan-Yan Z., Qin B. and Liu T. (2010). Integrating intra- and inter-document evidences for improving sentence sentiment classification. *Acta Automatica Sinica,* 36(10), 1417–1425.

Kang H., Yoo S. J. and Han D. (2012). Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications,* 39(5), 6000–6010.

Hutto C. and Gilbert E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216–225.

Zagibalov T. (2010). Unsupervised and Knowledge-Poor Approaches to Sentiment Analysis. *Ph.D. thesis, University of Sussex.*
Augustyniak,Ł., Szymański,P., Kajdanowicz T. and Tuligłowicz,W. (2015). Comprehensive study on lexicon-based ensemble classification sentiment analysis. *Entropy,* 18(1), 4.
Gonçalves,P., Araújo,M., Benevenuto F. and Cha M. (2013). Comparing and combining sentiment analysis methods. *Proceedings of the First ACM Conference on Online Social Networks*, 27–38.
Vilares D., Alonso M. A. and Gómez-Rodríguez,C. (2017). Supervised sentiment analysis in multilingual environments. *Information Processing & Management*, 53(3), 595–607.
Mukhtar N., Khan M. A. and Chiragh N. (2018). Lexicon-based approach outperforms supervised machine learning approach for Urdu sentiment analysis in multiple domains. *Telematics and Informatics,* 35(8), 2173–2183.
Jiang H., Kwong C. K. and Yung K. L. (2017). Predicting future importance of product features based on online customer reviews. *Journal of Mechanical Design,* 139(11).
Suryadi D. and Kim H. M. (2018). A systematic methodology based on word embedding for identifying the relation between online customer reviews and sales rank. *ASME Journal of Mechanical Design.*
Bag S., Tiwari M. K. and Chan F. T. S. (2019). Predicting the consumer's purchase intention of durable goods: An attribute-level analysis. *Journal of Business Research,* 94, 408–419.
Joung J. and Kim H. M. (2021). Explainable neural network-based approach to Kano categorisation of product features from online reviews. *International Journal of Production Research,* 1–21.
Kherwa P. and Bansal P. (2019). Topic modeling: A comprehensive review. *EAI Endorsed Transactions on Scalable Information Systems,* 7(24).
Deerwester S., Dumais S. T., Furnas G. W., Landauer T. K. and Harshman R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science,* 41(6), 391–407.
Buckley C., Allan J. and Salton G. (1994). Automatic routing and ad-hoc retrieval using SMART: TREC 2. *NIST Special Publication SP*, 45–45.
Kherwa P. and Bansal P. (2017). Latent semantic analysis: An approach to understand semantic of text. *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC),* 870–874.
Paatero P. and Tapper U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics,* 5(2), 111–126.
Lee D. D. and Seung H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature,* 401(6755), 788–791.
Hofmann T. (2013). Probabilistic latent semantic analysis. *arXiv preprint arXiv:1301.6705.*
Blei D. M., Ng A. Y. and Jordan M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3(Jan),* 993–1022.
Tuarob S. and Tucker C. S. (2014). Discovering Next Generation Product Innovations by Identifying Lead User Preferences Expressed Through Large Scale Social Media Data. *Volume 1B: 34th Computers and Information in Engineering Conference.*
Suryadi D. and Kim H. M. (2019). Data A.-Driven Methodology to Construct Customer Choice Sets Using Online Data and Customer Reviews. *ASME. J. Mech. Des., November.*
Park S., Lin K., Joung J. and Kim H. (2025). An automated data-driven approach for product design strategies to respond to market disruption following COVID-19. *Journal of Mechanical Design,* 147(3).
Denny M. J. and Spirling A. (2018). Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do about It. *Political Analysis,* 26, 168–189.
Jeong B., Yoon J. and Lee,J.-M. (2019). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management,* 48, 280–290.
Mikolov T., Sutskever I., Chen K., Corrado G. S. and Dean J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26.
Jin J., Jia D., and Chen K. (2021). Mining online reviews with a Kansei-integrated Kano model for innovative product design. *International Journal of Production Research,* 60(22), 6708–6727. https://doi.org/10.1080/00207543.2021.1949641.