PUSAN NATIONAL UNIVERSITY

Committee Members:
Professor Jongki Kim
Professor Mou Jian
Professor Taeho Hong

Thesis for the Degree of Master of Arts in Business Administration
Academic Advisor: Professor Taeho Hong

How Visual and Textual Information Shape Review Helpfulness Across Product Types: An Integrated Deep Learning and QCA Approach

2026.05.21

Presented by : Camila Estefanía Painefilo Hermosilla

College of Business
The Graduate School
Pusan National University

Content

I. Introduction – Research Background

Online Reviews and Consumer Decision-Making:

- Online reviews are a **key source of information** in consumer decision-making processes (Chevalier & Mayzlin, 2006).
- Reviews provide diagnostic cues that **shape product evaluation** (Mudambi & Schuff, 2010).

Rise of Multimodal Reviews:

- Online reviews increasingly **combine textual and visual information**, as many platforms enable users to share both text and images.
- User-generated images help consumers better understand **product appearance, usage contexts, and what to expect from the product** (Ma et al., 2018).

Visual Information in Reviews:

- The presence of images in online reviews can **enhance perceived helpfulness** by increasing credibility and reducing uncertainty (Filieri, 2015).
- **Aesthetic image quality** has also been shown to increase review helpfulness (Han et al., 2025).
- However, prior research has often treated visual information through **image presence rather than image quality** (Sun et al., 2019; Chatterjee, 2020; Lee et al., 2021).

Measuring Visual Quality:

- Deep learning enables scalable measurement of visual quality, including MUSIQ for aesthetic quality (Ke et al., 2021), while algorithms such as BRISQUE capture technical image quality (Mittal et al., 2012).

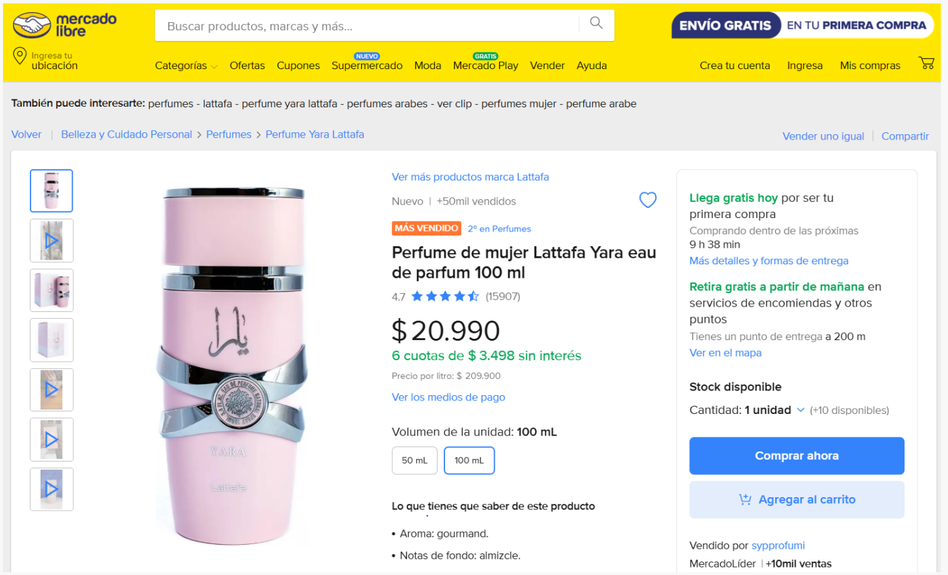
I. Introduction – Research Background

Product Type Differences:

- The influence of review information **varies across product types**, as consumers rely on different cues depending on evaluation uncertainty (Mudambi & Schuff, 2010).
- **Search goods** allow quality evaluation before purchase and rely more on objective information (Nelson, 1974).
- **Experience goods** require post-consumption evaluation and rely more on experiential and visual cues (Jiang & Benbasat, 2007).

Empirical Context: MercadoLibre

- MercadoLibre is a **major e-commerce platform in Chile** and Latin America, with a large and growing user base (Statista, 2025).
- The platform supports multimodal reviews combining textual and visual information since 2023.
- It covers diverse product categories, including both search and experience goods.



[Figure 1] Example of a product page from MercadoLibre.

Research Gap in Review Helpfulness:

1. Limited attention has been given to the role of image quality beyond image presence.
2. There is limited understanding of how textual and visual review characteristics jointly shape review helpfulness.
3. Configurational approaches remain scarce in review helpfulness research, especially for explaining how textual and visual conditions combine into pathways to high helpfulness.

I. Introduction – Research Questions & Objective

Research Questions

RQ1. To what extent do **aesthetic and technical image quality features** contribute to review helpfulness beyond **image presence and textual features**, and how does this contribution vary across **search and experience goods**?

RQ2. What **configurational combinations of visual and textual review conditions** lead to high review helpfulness, and how do these pathways differ across **search and experience goods**?

RQ3. What **complementary insights** emerge from integrating **predictive and configurational analyses** regarding how visual and textual information shape review helpfulness?

Research Objective

This study aims to **predict and explain review helpfulness** by developing an **integrated framework for multimodal e-commerce environments**. Specifically, the study examines the predictive contribution of **aesthetic and technical image quality** alongside textual characteristics, identifies **configurational pathways** leading to high review helpfulness through fuzzy-set Qualitative Comparative Analysis (fsQCA), and integrates both analytical perspectives across **search and experience goods**.

II. Literature Review

2.1 Online Reviews and Review Helpfulness

- Online reviews are a **primary source of market information** and an important form of **eWOM** that shapes consumer evaluations and purchase decisions (Chevalier & Mayzlin, 2006; Pavlou & Dimoka, 2006; Filieri, 2016).
- **Review helpfulness** reflects the **perceived informational value** of a review for decision-making (Mudambi & Schuff, 2010).
- Core textual determinants include **length, detail, diagnostic content, and sentiment,** which increase perceived helpfulness (Hu et al., 2014).
- Reviewer-related cues (identity disclosure, experience, verified purchase) enhance **credibility** and thus helpfulness (Forman et al., 2008; Yao et al., 2025).
- Prior research predominantly uses **linear, additive models**, examining independent effects of each factor (Hu et al., 2014).
- Configurational approaches such as **fsQCA** suggest that **multiple combinations of review attributes** can produce high helpfulness (Kwon & Jahng, 2023).

However, most review helpfulness studies have examined how **separate review characteristics** affect helpfulness, while less is known about how **multiple characteristics work together** to produce **highly helpful reviews**.

II. Literature Review

2.2 Visual Information and Image Quality in Online Reviews

- Online reviews increasingly include **visual information** in addition to text.
- **User-generated images** provide concrete cues about product usage and context, helping consumers visualize product performance and enhancing perceived **diagnosticity** while reducing uncertainty about product attributes (Mudambi & Schuff, 2010; Ma et al., 2018)
- Prior review helpfulness research has largely focused on **textual cues**, while visual studies often treat images in a **binary way (**present or absent) rather than examining variation in **image quality** (Hu et al., 2014; Sun et al., 2019; Chatterjee, 2020; Lee et al., 2021).
- Recent studies move beyond image presence toward the **aesthetic quality** of review photos (Han et al., 2025):
  - Visually appealing images can increase perceived review helpfulness by attracting attention and improving the **processing of visual information**.
- **Technical image quality;** the extent to which an image is free from distortions such as blur, noise, or compression artifacts, has not been examined in review helpfulness research, despite being well-established in computer vision through no-reference algorithms such as BRISQUE (Mittal et al., 2012).

This motivates examining both **aesthetic and technical image quality as complementary dimensions of visual information**, rather than treating images as binary signals.

II. Literature Review

2.3 Multimodal Information in Review Helpfulness Research

- Review helpfulness research has expanded from traditional textual and reviewer-related cues to a broader range of **user-generated content**, including **textual, visual, reviewer-related, and contextual information** (Ma et al., 2018; Li & Zhang, 2022).
- These feature categories reflect different dimensions of how review information is **generated, presented, and interpreted** in online environments (Hong et al., 2017; Hu, 2020).
- Prior research has examined textual features such as **review length, sentiment, readability, and level of detail**, reviewer-related cues such as **credibility and prior activity**, and contextual factors such as **product category and temporal dynamics** (Hu et al., 2014; Salehan & Kim, 2016).
- More recently, **visual features** have gained attention as platforms increasingly support user-generated images in reviews (Li et al., 2022).
- However, within the visual dimension, prior studies have more commonly focused on **simple indicators**, such as **image presence** or **number of images**, while the **quality of visual content** has received less systematic attention (Sun et al., 2019; Chatterjee, 2020; Kübler et al., 2024).

Review helpfulness research has become increasingly multimodal, but the **visual dimension remains more limited in quality-based representation**.

II. Literature Review

[Table 1] Summary of used features in review helpfulness, adapted and extended from Yao (2023).

Factor		Studies
Rating		Forman et al.(2008); Mudambi and Schuff(2010); Yin et al.(2014); Liu and Park(2015); Kuan et al.(2015); Ullah et al.(2015); Huang et al. (2015); Weathers et al.(2015); Yin et al. (2016); Hong et al.(2017); Filieri et al.(2018); Eslami et al.(2018); Wang et al.(2019); Ren and Hong (2019); Fresneda and Gefen(2019); Srivastava and Kalro(2019); Sun et al. (2019); Hlee (2020); Huang et al.(2020); Chatterjee (2020); Zhou et al. (2020); Li and Huang(2020); 야오즈옌 등(2021); Wu et al.(2021); Yang et al.(2021); Lee et al.(2021); Lei et al.(2021); Mauro et al. (2021); Han(2022); Kashyap et al.(2022)
Quadratic Rating		Mudambi and Schuff (2010); Yin et al. (2014); Ullah et al. (2015); Hong et al. (2017); Ren and Hong (2019); Srivastava and Kalro (2019); Zhou et al. (2020); 야오즈옌 등(2021); Yang et al. (2021)
Rating-related Factors	Average Rating	Yin et al. (2014); Hong et al. (2016); Lei et al. (2021)
	Number of Ratings	Yin et al. (2014); Yin et al. (2016); Lei et al. (2021)
	Rating Deviation	Kuan et al. (2015); Yin et al. (2016); Hong et al. (2016); 이시환 등(2020); Lee et al. (2021); Mauro et al. (2021)
	Equivocality	Karimi and Wang (2017); 야오즈옌 등(2020); Zhou et al. (2022)
	Review Extremity	Siering et al. (2018); Filieri et al. (2018); Choi and Leon (2020); Lee and Choe (2020)
	Review Inconsistency	Choi and Leon (2020); Wu et al. (2021)
Length		Mudambi and Schuff (2010); Liu and Park (2015); Kuan et al. (2015); Ullah et al. (2015); Huang et al. (2015); Weathers et al. (2015); Yin et al. (2016); Hong et al. (2017); Karimi and Wang (2017); Siering et al. (2018); Eslami et al. (2018); Fresneda and Gefen (2019); 야오즈옌 등(2020); Hlee (2020); Huang et al. (2020); Chatterjee (2020); 이시환 등(2020); zhou et al. (2020); Li and Huang(2020); Hu and Yang(2020); Lee et al.(2021); Lei et al.(2021); Li et al.(2022); Hang(2022); Kashyap et al.(2022)
Quadratic Length		Li and Huang (2020)

II. Literature Review

[Table 1] Summary of used features in review helpfulness, adapted and extended from Yao (2023).

Factor		Studies
Image	Image Presence	Srivastava & Kalro (2019); Sun et al. (2019); Chatterjee (2020); Fan & Zhang (2020); Lee et al. (2021); Han (2022); Ceylan et al. (2024); Kübler et al. (2024)
	Number of Images	Kübler et al. (2024)
	Image Aesthetic Quality	Han et al. (2025); Yang et al. (2024)
	Photo-Text Congruence	Ceylan et al. (2024); Kübler et al. (2024); Yu et al. (2026)
Sentiment		이현애 등(2017); 황침 등(2018); Eslami et al.(2018); Fresneda and Gefen (2019); Srivastava and Kalro(2019); 야오즈옌 등(2021); Yang et al.(2021); 노민정(2021);Han(2022); Zhou et al.(2022)
Polarity		Siering et al. (2018); Eslami et al. (2018); Lei et al. (2021); Mauro et al. (2021); Namvar and Chua (2022)
Emotion		Ullah et al. (2015); Yin et al. (2016); Hong et al. (2016); Wang et al. (2019); Ren and Hong (2019); Srivastava and Kalro (2019); Huang et al. (2020); Wu et al. (2021); Zhou et al. (2022); Xu et al. (2022)
Readability		Liu and Park(2015); Kuan et al.(2015); Yin et al.(2016); Hong et al. (2017); Siering et al. (2018); Wang et al.(2019); Fresneda and Gefen(2019); Srivastava and Kalro(2019); 야오즈옌 등 (2020); 이시환 등(2020); 야오즈옌 등(2021); Wu et al.(2021); 노민정 (2021)
Content		Qazi et al. (2016); Fresneda and Gefen (2019); Wu et al. (2021); Li et al. (2022); Zhou et al. (2022); Kashyap et al. (2022); Namvar and Chua (2022)
Linguistic Characteristics		Weathers et al.(2015); Siering et al.(2018); Wang and Karimi(2019); Wang et al.(2019); Srivastava and Kalro(2019); Sun et al.(2019); 야오즈옌 등(2020); Zhou et al.(2020); Lei et al.(2021); Mauro et al.(2021); Kashyap et al.(2022)
Verified Purchase		Li and Huang (2020)

II. Literature Review

[Table 1] Summary of used features in review helpfulness, adapted and extended from Yao (2023).

Factor	Studies
Verified Purchase	Li and Huang (2020)
Elapsed Days	Kuan et al.(2015); Yin et al.(2016); Hong et al.(2017); 황침 등(2018); Siering et al.(2018); Wang and Karimi (2019); Wang et al. (2019); Sun et al.(2019); Huang et al.(2020); 이시환 등(2020); Hu and Yang(2020); Zhou et al.(2020); Fan and Zhang(2020); Lee et al.(2021); 노민정(2021); Kashyap et al.(2022)
Exposure Days	Mudambi & Schuff (2010), (Hong et al. 2017), Kuan et al. (2015), Siering et al. (2018)
Two-Sidedness	Filieri et al. (2018); Chen (2016); Kim et al. (2017); Fan et al. (2022)
Topic Diversity	Li, Fu & Zhang (2022); Li et al. (2022)

II. Literature Review

2.4 A Multi-MAP Perspective on Review Helpfulness Research

- As review helpfulness becomes more complex in **multimodal environments**, examining it through a **single analytical lens** may provide only a partial understanding of the phenomenon.
- A **Multi-MAP approach** integrates **Methods, Approaches, and Perspectives** to address distinct but complementary aspects of a research problem (Levallet et al., 2021).
- For this study, the distinction between **variance** and **systems** perspectives is especially relevant (Levallet et al., 2021):
  - A **variance perspective** examines whether specific variables or feature sets contribute to explaining or predicting an outcome.
  - A **systems perspective** examines how multiple conditions operate together as configurations leading to an outcome.
- These perspectives are complementary for review helpfulness research:
  - It is important to assess whether **visual and textual characteristics** provide additional information for identifying helpful reviews.
  - It is also important to explain how these characteristics **combine into pathways** to high helpfulness.

This study adopts a **Multi-MAP design** to examine both the **predictive contribution** and the **configurational role** of visual and textual review information.

II. Literature Review

2.4 A Multi-MAP Perspective on Review Helpfulness Research

[Table 2] Comparison of Variance and Systems Perspectives in a Multi-MAP Approach

Dimension	Variance Perspective	Systems Perspective	Variance–Systems Multi-MAP
Primary focus	Individual variables and their relationships with an outcome	Combinations of interdependent conditions that produce an outcome	Complementary understanding of both individual contributions and configurational patterns
View of relationships	Net effects, often examined as additive or directional relationships	Conjunctural relationships, where conditions operate together	Recognition that individual effects and configurational relationships can offer different but complementary insights
Analytical logic	Effects-of-causes approach	Causes-of-effects approach	Integrates both perspectives to examine the strength and form of relationships
Typical analytical approaches	Regression, econometric models, predictive modeling	QCA, fsQCA, configurational analysis	Sequential or parallel use of variance- and systems-oriented approaches
Main strength	Identifies whether and to what extent individual variables contribute to an outcome	Identifies how multiple conditions combine into sufficient pathways leading to an outcome	Provides a more comprehensive understanding of complex phenomena
Main limitation	Limited ability to explain how multiple conditions jointly form pathways to an outcome	Limited ability to assess the individual predictive contribution of specific variables	Requires careful integration of findings across perspectives

II. Literature Review

2.4 A Multi-MAP Perspective on Review Helpfulness Research

Predictive Modeling and Configurational Analysis as Complementary Perspectives

2.4.1 Predictive Modeling — Variance Perspective

- Predictive modeling evaluates whether specific review characteristics or feature sets provide **additional value for identifying helpful reviews** (Levallet et al., 2021).
- In review helpfulness research, prior studies have incorporated structured, textual, and multimodal inputs to improve prediction performance (Salehan & Kim, 2016; Li & Zhang, 2022).
- However, predictive modeling provides more limited insight into **how multiple review characteristics work together** to produce high helpfulness.

2.4.2 Configurational Analysis — Systems Perspective

- Configurational analysis examines how **combinations of conditions** are jointly sufficient for an outcome, rather than estimating each factor separately (Ragin, 2000, 2008; Oana et al., 2021).
- This perspective captures **conjunction, equifinality, and causal asymmetry**, which are useful for explaining complex review helpfulness patterns (Fiss, 2011; Schneider & Wagemann, 2012).
- Applications of configurational analysis in **review helpfulness research remain limited**, motivating its use in this study (Kwon & Jahng, 2023).

Predictive modeling shows **whether review features add value**, while configurational analysis explains **how those features combine** to produce high helpfulness.

II. Literature Review

5. Product Type differences: Search and Experience Goods

- Product type theory distinguishes **search goods** (quality can be evaluated before purchase) and **experience goods** (quality evaluated after consumption) (Nelson, 1974)
- Consumers rely on different **information cues** depending on product type
- For **search goods**, attribute-based and detailed textual information is especially important (Mudambi & Schuff, 2010)
- For **experience goods**, consumers rely more on experiential narratives and contextual cues, often conveyed through images (Ma et al. 2018)
- The same review feature (e.g., length, emotion, imagery) can have different effects across product types (Mudambi & Schuff, 2010)
- **Visual elements** may play a stronger role for **experience goods**, while **text-based informativeness** may be more critical for **search goods** (Han et al., 2025).
- Prior review helpfulness studies have often examined product type through **separate feature effects** or as a **moderator within linear models** (Hu et al., 2014).

This motivates **separate analyses for search and experience goods** across both the **predictive** and **configurational** phases of the study.

II. Literature Review

6. Synthesis of Research Gaps in Previous Studies

1. Visual Information Gap

Prior studies have mainly examined **image presence**, while the role of **aesthetic and technical image quality** remains less understood (Sun et al., 2019; Han et al., 2025).

2. Configurational Gap

There is limited understanding of how **textual and visual review conditions work together** to produce high review helpfulness, as configurational approaches remain scarce in this research stream (Kwon & Jahng, 2023).

3. Integrated Perspective Gap

Prior research has rarely combined **predictive and configurational perspectives** to examine both the contribution of review features and the pathways through which helpfulness emerges (Levallet et al., 2021).

These gaps motivate an integrated study of **image quality**, **configurational pathways**, and **complementary predictive-configurational insights** in multimodal review helpfulness.

III. Research Framework

This study develops a **two-phase analytical framework** applied to reviews collected from **MercadoLibre Chile**:

Phase 1: Data Collection & Variable Operationalization

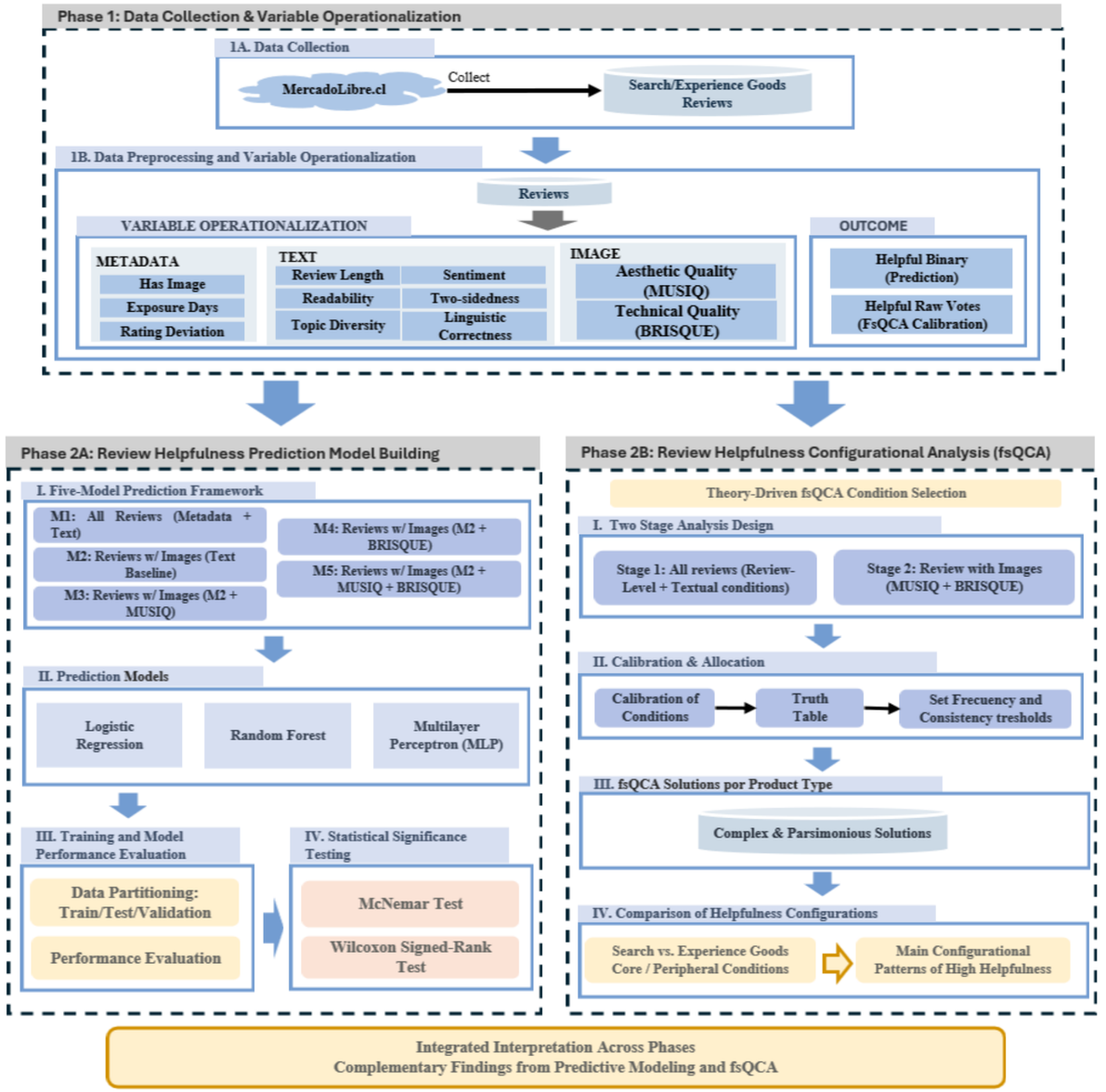
- Reviews classified as search or experience goods
- Variables operationalized across metadata, textual features, and visual quality (MUSIQ, BRISQUE)

Phase 2A: Review Helpfulness Prediction

- Three classifiers (Logistic Regression, Random Forest, MLP) under a five-model progression (M1–M5) isolating visual quality contributions
- Evaluated separately by product type using 10-fold cross-validation

Phase 2B: Review Helpfulness Configurational Analysis (fsQCA)

- Two-stage design: Stage 1 (all reviews: metadata + text); Stage 2 (image reviews: + MUSIQ + BRISQUE)
- Identifies product-type-specific pathways to high helpfulness



[Figure 2] Research Framework

IV. Experiments – Data Collection

Source: MercadoLibre Chile (MercadoLibre.cl)
Method: Automated Web Crawling by using Selenium library

Product Selection

- Based on **search vs. experience goods classification** (Nelson, 1974; Huang et al., 2009)
- Products selected across multiple categories and manually classified

Data Extraction:

- **80,145 reviews** collected across 593 products
- Review metadata (rating, dates, helpfulness votes)
- Review text
- Image-related fields (URLs/files)

Multimodal Context

- Reviews combine **textual and visual information**
- Image-enabled reviews available since 2023

[Figure 3] Example of a review on MercadoLibre platform.

IV. Experiments – Data Preprocessing & Sample Construction

Data Cleaning & Standardization
- Column standardization and type conversion
- Duplicate checks and data consistency validation

Product Matching
- Reviews matched to a validated product list
- Ensures classification into **search vs. experience goods**

Missing Values Handling
- Text and image fields treated as meaningful absence
- No missing values in key variables (dates, helpful votes)

Feature Preparation
- Basic variables constructed (e.g., text length, image indicators)
- Supports subsequent feature extraction

Filtering Criteria
- Reviews from **2023 onwards**
- Reviews with **at least one helpful vote** and reviews with **valid text and product type** (Mudambi & Schuff, 2010; Sun et al., 2019)

Final Dataset
- Final analytical dataset**: 14,603 reviews from 579 products**
- **Two outcome variables defined for different analytical approaches:**
  - helpful_binary (prediction models)
  - helpful_raw (fsQCA analysis)

IV. Experiments – Target Variable

Review helpfulness is operationalized **differently across the two analytical phases** to align with each method's underlying logic.

Phase 2A: Predictive Modeling (helpful_binary)

- Reviews binarized using a **product-type-specific median threshold** (Sun et al., 2019)
- Reviews above median = high helpfulness (1); below = low helpfulness (0)
- Accounts for differences in vote accumulation patterns across search vs. experience goods

Phase 2B: Configurational Analysis (fsQCA) (helpful_raw)

- Raw helpful vote count retained and **calibrated as a fuzzy-set membership score**
- Captures varying *degrees* of helpfulness rather than a strict binary cut
- Consistent with fsQCA calibration logic (Ragin, 2008; Pappas & Woodside, 2021)

[Table 2] Predictive Modeling - Outcome Variable Distribution

Subset	Count	Mean	std	Helpful (1)	Not Helpful (0)	% Helpful	% Not Helpful
All reviews	14603	0.416	0.493	6074	8529	41.60%	58.40%
Search	7506	0.341	0.474	2560	4946	34.10%	65.90%
Experience	7097	0.495	0.500	3514	3583	49.50%	50.50%

IV. Experiments – Variable Operationalization

[Table 3] Operationalization of Multimodal Review Variables

ID	Type	Variable Name	Operationalization	Studies
y1	Outcome	Helpfulness votes	Raw number of helpful votes received by each review	Mudambi & Schuff, 2010
y2	Outcome	Review helpfulness (Binary)	1 if helpful votes exceed the product-type-specific median; 0 otherwise	Sun et al., 2019; Kim et al., 2013
v1	Textual	Topic diversity	Proportion of Garvin-based product-quality dimensions detected in the review	Garvin, 1984; Sun et al., 2019; Li et al., 2022
v2	Textual	Readability	Spanish-adapted readability score based on the Índice de Flesch-Szigriszt (IFSZ)	Szigriszt Pazos, 1993
v3	Textual	Review length	Total word count of the cleaned review text	Mudambi & Schuff, 2010; Liu & Park, 2015
v4	Textual	Sentiment	Continuous polarity score computed as positive minus negative sentiment probability using pysentimiento / RoBERTuito	Pérez et al., 2024; Pérez et al., 2022; Hutto & Gilbert, 2014
v5	Textual	Two-sidedness	Degree of co-occurrence of positive and negative evaluative content in the review	Fan et al., 2022; Pérez et al., 2024
v6	Textual	Linguistic correctness	Normalized spelling-based measure derived from misspelled words in the review text	Ghose & Ipeirotis, 2011; Srivastava & Kalro, 2019
v7	Visual	Aesthetic image quality (MUSIQ)	Mean MUSIQ score across all images associated with a review	Ke et al., 2021; Han et al., 2025
v8	Visual	Technical image quality (BRISQUE)	Mean BRISQUE score across review images, capturing blur, noise, and visual distortions	Mittal et al., 2012
v9	Metadata	Exposure days	Number of days between review posting date and data collection date	Mudambi & Schuff, 2010; Kuan et al., 2015; Siering et al., 2018
v10	Metadata	Image presence	Binary indicator: 1 if the review includes at least one user-generated image	Sun et al., 2019; Chatterjee, 2020; Lee et al., 2021; Kübler et al., 2024
v11	Metadata	Rating deviation	Absolute deviation between the review rating and the product's average rating	Yin et al., 2016; Kuan et al., 2015; Lee et al., 2021

IV. Experiments – Variable Operationalization

Operationalized Variables Descriptive Statistics

[Table 4] Operationalized Variables Descriptive Statistics

ID	Variable	N	Missing (n)	Missing (%)	Mean	Std. Dev.	Min	25%	Median	75%	Max
y1	Helpfulness Votes	14603	0	0	6.476	35.304	1	1	1	4	2762
y2	Review Helpfulness (Binary)	14603	0	0	0.416	0.493	0	NaN	0	NaN	1
v1	Topic Diversity	14603	0	0	0.183	0.132	0	0.111	0.111	0.222	0.889
v2	Readability	14603	0	0	58.314	31.614	0	45.085	67.696	80.925	100
v3	Review Length	14603	0	0	19.195	19.927	1	6	13	25	241
v4	Sentiment	14603	0	0	0.59	0.549	-0.984	0.338	0.892	0.96	0.984
v5	Two-sidedness	14603	0	0	0.136	1.03	0	0	0	0	55.97
v6	Linguistic Correctness	14603	0	0	0.981	0.018	0.8	0.97	0.98	1	1
v7	Aesthetic Image Quality (MUSIQ)	4847	9756	66.81	67.505	7.564	21.109	63.556	69.125	73.242	79.084
v8	Technical Image Quality (BRISQUE)	4847	9756	66.81	38.369	17.502	-22.633	26.315	37.583	49.047	206.688
v9	Exposure Days	14603	0	0	456.913	304.729	2	169	428	707	1176
v10	Image Presence	14603	0	0	0.332	0.471	0	NaN	0	NaN	1
v11	Rating Deviation	14603	0	0	0.232	0.471	0	0.015	0.054	0.127	3.821

Note: Missing values for MUSIQ and BRISQUE correspond to reviews without images, as image-quality scores were computed only for reviews containing at least one image.

IV. Experiments – Prediction Model Building

I. Five-Model Progression (M1–M5)

Systematically evaluates **the incremental predictive contribution** of each visual quality dimension beyond the image-review baseline.

[Table 7] Five-model prediction design.

Model	Dataset	Feature set
M1	All reviews	Metadata + Textual
M2	Image-inclusive reviews	Metadata + Textual
M3	Image-inclusive reviews	M2 + MUSIQ (Aesthetic image quality)
M4	Image-inclusive reviews	M2 + BRISQUE (Technical image quality)
M5	Image-inclusive reviews	M2 + MUSIQ + BRISQUE (Full visual quality)

II. Classifiers

- **Logistic Regression:** linear baseline
- **Random Forest:** non-linear ensemble; n_estimators = 100, max_features = sqrt
- **Multilayer Perceptron (MLP):** 2 hidden layers (32 neurons), ReLU, Adam, 30 epochs, batch = 256

III. Training & Evaluation

- Stratified 80/10/10 train/validation/test Split
- 10-fold cross-validation applied to training set (Kohavi, 1995)
- Primary metric: F1, complementary metric: AUC-ROC
- Secondary metrics: Accuracy, Precision, Recall
- Statistical significance: McNemar test (binary predictions), Wilcoxon signed-rank test (fold-level AUC)

IV. Experiments – Configurational Modeling (fsQCA)

I. Two-Stage Analytical Design

Stage 1: All reviews | Review-level + Textual conditions

Stage 2: Image-inclusive reviews only | Adds MUSIQ & BRISQUE as conditions

Theory-driven condition selection applied before calibration.

II.A Calibration & Set Construction

- Anchors defined using p25/p50/p75 distribution-based thresholds (Pappas & Woodside, 2021)
- Theory-based anchors applied where substantively justified (Ragin, 2008)
- Image quality conditions calibrated exclusively on image-inclusive subsample
- Outcome calibrated using product-type-specific anchors (Ragin, 2008)

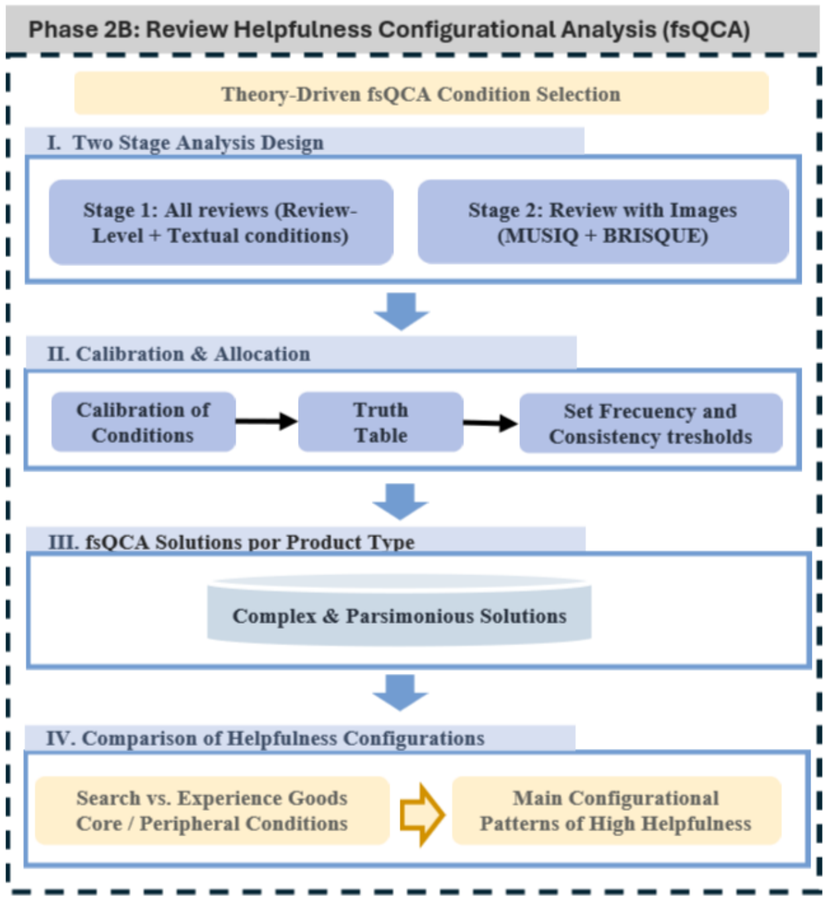
II.B Thresholds

- **Frequency cutoff:** 90th percentile of cumulative cases per analytical subset
- **Raw consistency:** Stage 1 ≥ 0.80 / Stage 2 ≥ 0.90 (Ragin, 2008)
- **PRI consistency:** Stage 1 ≥ 0.60 / Stage 2 ≥ 0.65 (Schneider & Wagemann, 2012)
- **Intermediate solutions not reported:** prior theory does not provide sufficient **directional guidance** for counterfactual assumptions, particularly for technical image quality.

III. Solutions for Product Type

IV. Comparison of Helpfulness Configurations

- Complex and parsimonious solutions identified separately for **search vs. experience goods**
- Configurations compared across **stages and product types** to support the interpretation of recurring and product-type-specific patterns



V. Results – Phase 2A: Prediction Model Performance

Visual quality provides **modest and selective gains** beyond the image-review baseline. The contribution is more visible for **experience goods**, but the gains are not uniformly additive across M3–M5.

- Relative to **M2, the image-review baseline**, visual-quality models provide **small but clearer predictive gains for experience goods** than for search goods.
- For **experience goods**, RF F1 increases from **82.3% in M2** to **83.0% in M3**, suggesting that aesthetic quality adds a small predictive contribution.
- For **search goods**, RF F1 increases only slightly from **62.2% in M2** to **62.6% in M3**, while M5 does not improve F1.

- **AUC provides a complementary check:** experience-goods AUC reaches **83.4% in M4/M5**; Wilcoxon tests show that visual-quality models outperform M2 more consistently for **experience goods** than for search goods.

- **McNemar test:** for experience goods, **M5 differs from M3 and M4** ($p = 0.001$; $p = 0.013$), suggesting that combining MUSIQ and BRISQUE changes some helpfulness predictions despite small F1/AUC gains.

[Table 5] Random Forest Performance across Visual-Quality Model Progressions

Product Type	Metric	M2 (image baseline)	M3 (+MUSIQ)	M4 (+BRISQUE)	M5 (+Both)
Search goods	F1	62.2%	**62.6%**	62.4%	61.8%
	AUC	75.0%	76.4%	75.9%	**76.7%**
Experience goods	F1	82.3%	**83.0%**	82.6%	81.5%
	AUC	82.4%	83.3%	**83.4%**	**83.4%**

[Table 6] ΔF1 Relative to the M2 Baseline — Random Forest

Product type	M3 (+ MUSIQ)	M4 (+ BRISQUE)	M5 (+Both)
Search	+0.4 pp	+0.2 pp	-0.4 pp
Experience	+0.7 pp	+0.3 pp	-0.8 pp

V. Results – Phase 2B: fsQCA Stage 1 Configurations

Absent linguistic correctness is the only core condition shared across both product types in Stage 1, suggesting the relevance of **authentic, colloquial consumer writing**.

- **Absent linguistic correctness** is the only core condition shared across both product types, suggesting that more **natural and informal consumer writing** forms part of sufficient pathways to high helpfulness (Srivastava & Kalro, 2019).
- For search goods, **image presence, review length, and topic diversity** are also core. helpful reviews are informationally rich and multimodal
- For experience goods, **absent linguistic correctness is the only core condition.** Unlike search goods, no condition reaches core status in its present form, suggesting that what defines helpful experience goods reviews is more about writing style than informational richness

Condition	Search Goods (S1a)	Experience Goods (E1a)
Has Image	●	•
Sentiment	•	•
Review Length	●	•
Readability	•	⊗
Rating Deviation	⊗	⊗
Linguistic Correctness	⊗	⊗
Topic Diversity	●	•
Consistency	0.852	0.819
Raw Coverage	0.059	0.076
Unique Coverage	0.059	0.076
Overall Consistency	0.852	0.819
Overall Coverage	0.059	0.076

●/⊗ = core present/absent; •/⊗ = peripheral present/absent. Frequency cutoff: Search ≥28; Experience ≥25. Raw consistency cutoff: ≥0.80. PRI minimum: ≥0.60.

[Figure 4] Sufficient configurations for high helpfulness - Stage 1

V. Results – Phase 2B: fsQCA Stage 2 Configurations

Absent high positive sentiment is the sole core condition across all configurations and both product types; once images are present, a **more restrained evaluative tone** becomes the most consistent configurational pattern.

- **Absent high positive sentiment** and **sufficient review length** are the only two core conditions shared across all five search-goods configurations. When images are present, reviews with a more restrained evaluative tone and adequate length consistently achieve high helpfulness.
- For experience goods, **absent sentiment is the sole core condition,** once visual information is available, evaluative tone dominates over writing style
- Visual quality conditions appear exclusively in **peripheral and absent roles** across all seven configurations. Image quality contributes to the **configurational context** of helpful reviews, but does not define the main pathways to high helpfulness.

	Search goods					Experience goods	
Condition	**S2a**	**S2b**	**S2c**	**S2d**	**S2e**	**E2a**	**E2b**
Sentiment	⊗	⊗	⊗	⊗	⊗	⊗	⊗
Review Length	●	●	●	●	●	•	•
Readability		⊗	•	⊗	•		⊗
Rating Deviation	⊗	⊗	⊗	•	•	⊗	⊗
Linguistic Correctness	•	⊗	•	•	⊗	•	•
Topic Diversity	•	⊗	⊗	⊗	⊗	•	•
Aesthetic Quality	⊗	⊗	⊗	⊗		⊗	
Technical Quality	•	⊗	⊗	⊗	⊗	⊗	⊗
Consistency	**0.972**	**0.963**	**0.966**	**0.969**	**0.988**	**0.973**	**0.969**
Coverage	**0.108**	**0.109**	**0.109**	**0.110**	**0.102**	**0.110**	**0.110**
Overall Consistency			**0.912**			**0.965**	
Overall Coverage			**0.134**			**0.115**	

●/⊗ = core present/absent; •/⊗ = peripheral present/absent. Frequency cutoff: Search ≥4; Experience ≥5. Raw consistency cutoff: ≥0.90. PRI minimum: ≥0.65.

[Figure 5] Sufficient configurations for high helpfulness - Stage 2

VI. Conclusions – Summary of Findings

RQ1: To what extent do aesthetic and technical image quality features contribute to review helpfulness beyond image presence and textual features, and how does this contribution vary across search and experience goods?

- Visual quality provides **modest and selective gains** beyond image presence and textual variables, with a clearer contribution for **experience goods**.
- Under Random Forest, experience-goods F1 increases from **82.3% in M2** to **83.0% in M3**, while search-goods F1 rises only slightly from **62.2% to 62.6%**.
- These results show that moving beyond a binary image-presence indicator is useful, but its contribution is **not uniform across product types or model configurations**.

RQ2: What configurational combinations of visual and textual review conditions lead to high review helpfulness, and how do these pathways differ across search and experience goods?

- **No single condition is necessary** for high review helpfulness; it emerges from **configurations of conditions**.
- In **Stage 1**, **absent high linguistic correctness** is the only core condition shared across both product types.
- In **Stage 2**, **absent high positive sentiment** becomes the sole core condition across both product types once images are present.
- Visual quality appears only in **peripheral or absent roles**, rather than as a core condition.

RQ3: What complementary insights emerge from integrating predictive and configurational analyses regarding how visual and textual information shape review helpfulness?

- **Predictive contribution** and **configurational centrality** are not equivalent forms of importance.
- Visual quality can help **identify helpful reviews**, especially for experience goods, without defining the **main pathways** through which high helpfulness is achieved.

Visual quality provides **predictive value** in Phase 2A, but it does not become a **core condition** in the configurational pathways of Phase 2B. Its role is **supportive rather than determinative**.

VI. Conclusions – Contributions

Theoretical Contribution

- Extends review helpfulness literature by showing that predictive importance and configurational centrality do not necessarily align: image quality can improve prediction without constituting a core configurational driver of helpfulness
- Visual quality contributes as a **meaningful predictive signal for experience goods** but is more limited for search goods, consistent with Nelson's (1974) search/experience distinction
- **Writing authenticity** and a **more balanced evaluative tone** consistently characterize helpful reviews, refining theoretical accounts of how informational and stylistic cues shape helpfulness in multimodal environments.

Methodological Contribution

- Develops an **integrated analytical framework** combining predictive modeling and fsQCA within a unified design
- Demonstrates that the **two methods answer different questions** about the same phenomenon. Visual quality matters for prediction but not for configuration, a result neither method could have produced alone (Ragin, 2008; Greckhamer et al., 2018)
- Applies **large-N fsQCA** to a multimodal IS dataset, demonstrating scalability of configurational methods (Meurer et al., 2025)

Practical Contribution

- The two analytical phases support **different but complementary platform decisions**:
  Phase 2A informs **review ranking and visibility**, while **Phase 2B** informs **review-writing support and interface prompts**.
- For ranking, platforms can use **visual quality selectively** as an additional signal, especially for **experience goods**, rather than simply rewarding image presence.
- For review-writing support, platforms can encourage users to provide **concrete product details**, **balanced evaluations**, and **useful images** when they help others understand the product.

VI. Conclusions – Limitations & Further Research Directions

- **Helpfulness vote filter constrains the analytical sample:** only reviews with at least one helpful vote were included (**14,603 of 80,145 reviews; ~18%**), which may exclude helpful reviews with low exposure. **Further research** may use exposure-adjusted helpfulness measures, alternative sampling strategies, or platforms with more complete voting data (Mudambi & Schuff, 2010; Ma et al., 2018).
- **No reviewer-level data:** platform constraints prevented the inclusion of reviewer expertise, reputation, or review history, which may also shape helpfulness evaluations. **Further research** may incorporate reviewer-level characteristics and examine how they interact with multimodal review content.
- **Single-platform:** results are based on **MercadoLibre Chile**, where product categories, review formats, and voting behaviors may differ from other platforms. **Further research** may test the framework across other Latin American and global e-commerce environments
- **Low fsQCA solution coverage:** coverage remained low across all four subsets (**0.059–0.134**), likely reflecting the limited variation of helpfulness votes and the small number of evaluated reviews for many products. **Future research** may examine platforms with more symmetric voting systems, apply exposure-adjusted outcomes, and use alternative calibration anchors and consistency thresholds (Schneider & Wagemann, 2012; Greckhamer et al., 2018).

Thank You for your attention!

Appendix A - fsQCA Truth Tables

[Table A 1] Truth Tables for High Review Helpfulness in Search and Experience Goods – Stage 1

Search Goods – Stage 1

Has Image	Sentiment	Review Length	Readability	Rating Deviation	Linguistic Correctness	Topic Diversity	N	High Review Helpfulness	RAW consistency	PRI consistency
1	1	1	1	0	0	1	40	1	0.852	0.613
1	1	1	0	0	0	1	94	0	0.810	0.598
1	1	1	0	0	1	1	401	0	0.728	0.571
1	1	1	1	0	1	1	238	0	0.757	0.569
1	1	1	0	0	1	0	78	0	0.770	0.552
1	1	1	0	0	0	0	30	0	0.820	0.546
1	1	1	0	1	1	1	59	0	0.847	0.542
0	0	1	1	0	0	0	32	0	0.879	0.525
1	1	1	1	0	1	0	81	0	0.762	0.521
1	1	1	1	1	1	1	36	0	0.864	0.521
0	0	1	1	0	1	0	48	0	0.839	0.475
0	1	1	0	1	0	1	35	0	0.901	0.468
0	1	1	1	0	0	0	74	0	0.783	0.467
0	0	1	0	0	0	0	28	0	0.878	0.464

Experience Goods – Stage 1

Has Image	Sentiment	Review Length	Readability	Rating Deviation	Linguistic Correctness	Topic Diversity	N	High Review Helpfulness	RAW consistency	PRI consistency
1	1	1	0	0	0	1	87	1	0.819	0.620
1	1	1	0	0	0	0	50	0	0.814	0.598
1	1	1	1	0	0	1	54	0	0.803	0.554
0	1	1	1	0	0	1	91	0	0.805	0.456
0	0	1	0	0	1	0	31	0	0.869	0.460
0	1	1	0	0	0	1	119	0	0.814	0.480
0	0	1	0	0	0	0	26	0	0.892	0.490
1	1	1	1	1	1	0	27	0	0.853	0.492
1	1	1	1	1	1	1	39	0	0.875	0.502
1	1	1	0	1	1	1	42	0	0.890	0.553
1	1	0	0	0	1	1	84	0	0.810	0.445
0	1	1	0	1	1	0	27	0	0.893	0.452
0	1	1	0	1	1	1	42	0	0.908	0.438
0	1	1	1	1	1	1	60	0	0.888	0.443
0	0	1	1	0	1	0	64	0	0.816	0.426

Appendix A - fsQCA Truth Tables

[Table A 2] Truth Tables for High Review Helpfulness in Search and Experience Goods – Stage 2

Search Goods – Stage 2

Sentiment	Review Length	Readability	Rating Deviation	Linguistic Correctness	Topic Diversity	Aesthetic Quality	Technical Quality	Number	High Review Helpfulness	RAW consistency	PRI consistency
0	1	1	1	0	0	0	1	4	1	0.987	0.661
0	1	1	0	1	0	0	0	4	1	0.966	0.662
0	1	0	0	0	0	0	0	7	1	0.963	0.664
0	1	1	0	1	1	0	1	5	1	0.983	0.671
0	1	0	0	1	1	0	1	8	1	0.982	0.695
0	1	0	1	1	0	0	0	6	1	0.969	0.715
0	0	1	1	1	0	0	0	5	0	0.974	0.642
0	1	0	1	0	0	0	0	5	0	0.969	0.637
0	1	0	1	0	1	0	0	4	0	0.975	0.636
1	1	1	0	0	1	0	0	10	0	0.928	0.633
1	1	1	0	0	1	1	0	20	0	0.904	0.633
0	1	0	0	0	0	1	0	4	0	0.965	0.630
0	1	1	0	1	1	0	0	4	0	0.965	0.624
0	1	1	0	0	1	1	0	4	0	0.968	0.618
0	1	0	0	0	0	0	1	4	0	0.977	0.616

Experience Goods – Stage 2

Sentiment	Review Length	Readability	Rating Deviation	Linguistic Correctness	Topic Diversity	Aesthetic Quality	Technical Quality	Number	High Review Helpfulness	RAW consistency	PRI consistency
0	1	0	0	1	1	0	0	6	1	0.978	0.737
0	1	0	0	1	1	1	0	8	1	0.979	0.757
0	1	1	0	1	1	0	0	5	1	0.983	0.794
1	1	0	1	1	1	1	0	15	0	0.946	0.624
1	1	0	1	1	0	0	0	7	0	0.948	0.608
1	1	1	1	0	1	1	0	6	0	0.963	0.602
0	1	0	1	1	1	0	0	12	0	0.968	0.598
1	1	1	1	1	1	0	0	21	0	0.939	0.586
0	1	1	1	1	1	0	0	6	0	0.970	0.572
1	1	0	0	1	1	1	1	18	0	0.943	0.572
1	1	1	0	0	1	1	1	7	0	0.961	0.570
1	1	0	0	0	0	0	1	5	0	0.944	0.562
1	1	1	0	0	1	0	1	11	0	0.931	0.553
1	1	1	1	1	1	1	0	9	0	0.942	0.550
1	1	1	1	1	0	1	0	13	0	0.922	0.546

Appendix B - Phase 2A Performance Results: Secondary Metrics

[Table B 1] Search goods: Secondary performance metrics

Metric	Clf.	M1	M2	M3 (+MUSIQ)	M4 (+BRISQUE)	M5 (+Both)
Accuracy	LR	63.8%	**68.2%**	**68.2%**	67.4%	68.0%
Accuracy	RF	**73.3%**	69.5%	70.1%	70.1%	69.5%
Accuracy	MLP	65.9%	67.8%	67.6%	67.7%	**68.3%**
Precision	LR	47.6%	**63.4%**	**63.4%**	62.2%	63.0%
Precision	RF	66.4%	65.6%	66.8%	**67.0%**	66.3%
Precision	MLP	50.1%	62.7%	62.6%	62.4%	**63.0%**
Recall	LR	59.7%	59.8%	**60.2%**	59.4%	60.1%
Recall	RF	44.1%	**59.1%**	**59.1%**	58.3%	58.1%
Recall	MLP	59.1%	60.6%	59.1%	60.9%	**62.2%**

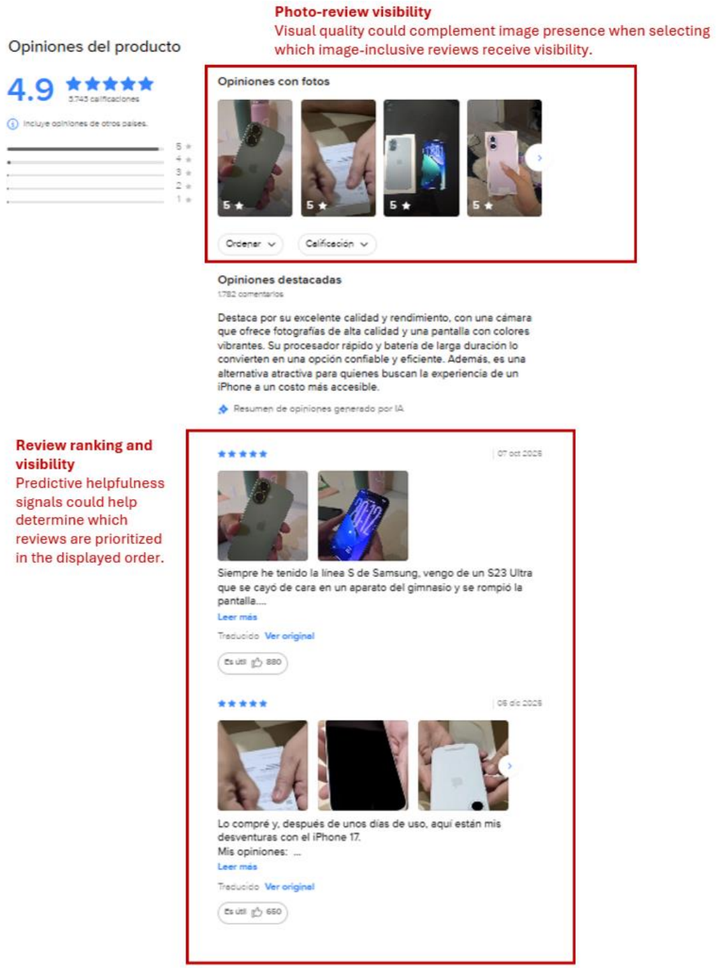
[Table B 2] Experience goods: Secondary performance metrics

Metric	Clf.	M1	M2	M3 (+MUSIQ)	M4 (+BRISQUE)	M5 (+Both)
Accuracy	LR	64.7%	66.8%	67.0%	67.0%	**67.1%**
Accuracy	RF	70.8%	77.7%	**78.6%**	78.2%	76.8%
Accuracy	MLP	66.2%	67.6%	67.6%	67.9%	**68.3%**
Precision	LR	65.0%	75.5%	75.6%	75.5%	**75.8%**
Precision	RF	70.5%	78.1%	**78.8%**	78.5%	77.7%
Precision	MLP	67.3%	**76.7%**	75.9%	75.9%	**76.7%**
Recall	LR	62.4%	65.6%	**65.8%**	**65.8%**	**65.8%**
Recall	RF	70.7%	87.0%	**87.7%**	87.3%	85.8%
Recall	MLP	61.5%	65.8%	67.1%	**67.6%**	67.4%

Appendix C - Illustrative Platform Application of the Study Findings

The purpose of these figures is not to propose an implemented platform design, but to illustrate how the study's empirical findings may be applied in practice.

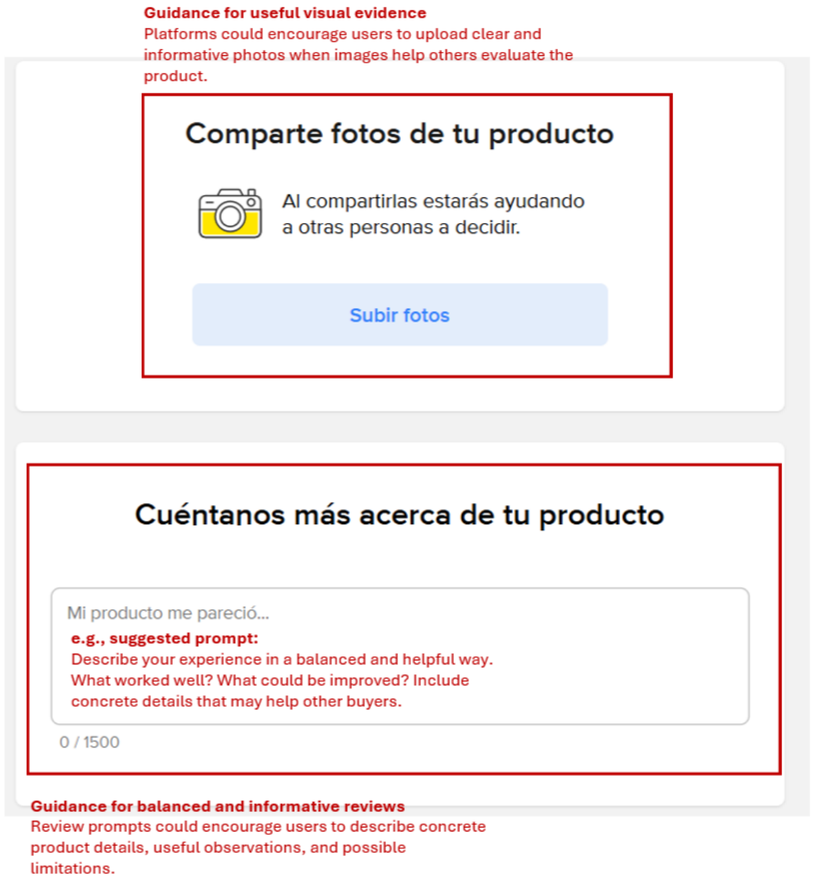
- Reviews may be prioritized using a combination of textual and visual signals, with greater attention to visual quality in experience-goods contexts. The figure highlights, in red, where a platform could incorporate helpfulness-related ranking logic beyond simple image presence.



[Figure C 1] Illustrative application to review ranking and visibility

Appendix C - Illustrative Platform Application of the Study Findings

The highlighted areas show where a platform could provide prompts encouraging balanced evaluation, authentic consumer-style writing, informative textual content, and useful images when relevant. These prompts reflect the combinations of conditions more consistently associated with high helpfulness in the configurational analysis.



[Figure C 2] Illustrative application to review-writing guidance