Journal of Retailing and Consumer Services 93 (2026) 104888

Contents lists available at ScienceDirect

Journal of Retailing and Consumer Services

journal homepage: www.elsevier.com/locate/jretconser

AI-driven sentiment analysis of grape wine consumers: A comparative study between China and Italy

Qiankun Liu [a,*], Muhabaiti Pareti [a,b]

[a] Alma Mater Studiorum - University of Bologna, Bologna, Italy
[b] Xinjiang Agricultural University, Urumqi, China

ARTICLE INFO

Keywords:
Sentiment analysis
Wine consumption
Consumer behavior
Big data analytics
Text mining
China and Italy

ABSTRACT

As global e-commerce expands, cross-cultural consumer sentiment analysis using native-language deep learning remains underexplored in online retail contexts. This study analyzes 33,464 online wine reviews from China (JD.com) and Italy (Amazon.it) to examine how culturally embedded review patterns are associated with consumer emotional expression in digital environments. By integrating Self-Determination Theory (SDT) and Hofstede's cultural dimensions within a Transformer-based natural language processing (NLP) framework, the study links theoretical constructs to observable sentiment patterns. The results reveal a systematic divergence between platform-generated star ratings and underlying textual sentiment. Chinese consumers emphasize functional and risk-related attributes, whereas Italian consumers focus more on experiential, social and aesthetic dimensions, suggesting that online reviews may serve as indicators of culturally embedded evaluation tendencies rather than purely platform-driven behaviors. Building on these findings, this study develops a cross-cultural digital sentiment matrix that conceptualizes consumer behavior along two dimensions: motivation and expression. This framework provides a transferable analytical reference for interpreting cross-cultural consumer sentiment and supporting data-driven retail decision-making.

1. Introduction

With the acceleration of global digitalization, online shopping platforms provide consumers with abundant information through online reviews, facilitating purchase decisions (Pocchiari et al., 2025). The expansion of online retail allows food consumers in different countries to purchase wine more conveniently (Rui et al., 2025). Wine consumption demonstrates shifts in consumer preferences and values (Moscovici et al., 2022). Along with the growth of online wine retail, China and Italy, as major wine markets, provide a reference for the sentiment analysis of wine consumers across different countries. Italy represents a paragon of the "old market" for wine, where wine is deeply rooted in daily life, and consumption behavior appears to fundamentally align with intrinsic motivation, aesthetics, and cultural heritage (Bazzani et al., 2024). China, as an emerging wine market, often demonstrates indicative patterns of localized pragmatism, where consumption behavior appears to be more readily influenced by external factors (Duan et al., 2022; Dang-Van et al., 2026). Understanding how these differences are expressed through platform-mediated consumer sentiment is therefore critical for effective retail strategy.

The foundation of product market performance is consumer trust (Grabner-Kraeuter, 2002; Wang et al., 2024). As a channel for consumer feedback, online reviews provide abundant information and sentiments (Guo et al., 2017), influencing product reputation and potential consumers (Basu et al., 2024). Despite these advances, existing research has not sufficiently explained how post-purchase emotional expressions are culturally embedded and manifested across different markets. Although digital sensory marketing can effectively shape purchasing responses through product displays (Dong et al., 2025), how post-purchase emotional expressions are culturally encoded and subsequently uploaded to online retail platforms remains largely opaque (Moradi et al., 2023). This limitation is particularly critical in cross-cultural e-commerce contexts. Such discrepancies in consumer sentiment become more pronounced under varying wine consumption habits and cultural backgrounds, particularly in the current context where wine retailers rely on big data to assist in product positioning (Morimura and Sakagawa, 2023; Bai and Yu-Buck, 2025).

Cross-cultural sentiment analysis aids in understanding the diversification of consumer evaluations (Přibáň et al., 2024). Research on Chinese cross-border e-commerce demonstrates that leveraging artificial

* Corresponding author.
E-mail addresses: qiankun.liu2@unibo.it (Q. Liu), muhabaiti86@xjau.edu.cn (M. Pareti).

https://doi.org/10.1016/j.jretconser.2026.104888
Received 31 January 2026; Received in revised form 15 May 2026; Accepted 22 May 2026
Available online 26 May 2026

intelligence to analyze e-commerce data can better formulate marketing strategies (Guo et al., 2025). To construct an effective comparative framework, it is necessary to clarify the sentiment distribution across different cultures, deconstruct the degree of consumer attention toward products and additional attributes such as services, and verify the effectiveness of the selected analytical methods in handling platform noise. Based on these requirements, the following research questions are proposed:

RQ1:What differences exist in the sentiment distribution and expression of online wine consumers between China and Italy?
RQ2:What differences exist in the degree of consumer attention toward products and additional attributes such as services?
RQ3:How effectively can native-language deep learning models reveal culturally embedded linguistic and emotional differences?

To address these research questions, this study applies artificial intelligence technology to decode the sentiment differences of online wine consumers in China and Italy. Rather than establishing 'true generalizability', this study aims to contribute transferable analytical insights for cross-cultural sentiment analysis by identifying recurring behavioral patterns across digital retail contexts. By analyzing online wine retail data from JD in China and Amazon in Italy, this study constructs a comparative sentiment analysis framework based on Transformer architectures (RoBERTa for Chinese and UmBERTo for Italian), and ultimately integrates consumer behavior theories to analyze the underlying causes of these sentiment differences. Furthermore, a dual-layer supplementary validation framework is introduced to examine cross-platform consistency and analytical consistency across methods.

The remainder of this paper is structured as follows: Section 2 reviews the relevant literature and constructs the theoretical framework; Section 3 details the methodology; Section 4 reports the research results and robustness tests; Section 5 discusses the research findings, implications, limitations and proposes future research directions; Section 6 provides the conclusion.

2. Literature review

2.1. Cross-cultural wine consumption in digital environments

Driven by the trade facilitation of online shopping, the global wine market is increasingly influenced by digitalization, multiculturalism, and shifting consumer habits. As consumer behavior internationalizes, understanding how consumers from diverse backgrounds evaluate wine and express their consumption sentiment has become crucial. Religion, politics, and social planning may systematically influence per capita wine consumption (Agnoli and Outreville, 2021). Recent cross-country studies on wine e-commerce indicate that in emerging markets like China, wine consumption is deeply intertwined with evolving cultural tastes, consumer sentiments, and shifts in social identity (Rui et al., 2025). As established in Consumer Culture Theory (CCT), this transition requires researchers to examine not just what consumers buy, but the culturally embedded frameworks they use to assign value to their purchases (Arnould and Thompson, 2005). These culturally embedded frameworks manifest in online reviews, where sentiment serves as a measurable proxy for consumer cognition and trust.

Online reviews, functioning as an evolution of traditional word-of-mouth, transcend geographical boundaries and mitigate perceived shopping risks for e-commerce consumers (Wistedt, 2024). Consequently, acknowledging that cultural backgrounds may influence market preferences is essential for analyzing post-purchase consumer satisfaction (Croitoru et al., 2024). However, online wine retailing also entails perceived risks, particularly concerning product authenticity and food safety, which influence consumer trust and purchase intentions within emerging markets (Garaus and Treiblmaier, 2021; Ladwein and Sánchez Romero, 2021). These complex trust dynamics and risk perceptions are ultimately articulated as measurable sentiments within online reviews (Wang et al., 2022).

2.2. Psychological and cultural drivers of consumer motivation

To concretize and theorize these deep emotional drivers, this study integrates a dual psychological and cultural analytical framework. While consumer emotions and trust significantly impact product cognition and post-purchase evaluations (Bagozzi et al., 1999), this trust mechanism is fundamentally shaped by the joint effects of diverse cultural backgrounds and e-commerce platform environments (Hallikainen and Laukkanen, 2018; Kim et al., 2018; Kim et al., 2023). To systematically decode how distinct cultures influence consumer trust and motivate online review behaviors, the psychological dimension of this framework is anchored in Self-Determination Theory (SDT) (Ryan and Deci, 2000).

Based on SDT, individual consumer behavior is categorized into extrinsic and intrinsic motivation. Extrinsic motivation encompasses external incentives or functional pressures, while intrinsic motivation involves internal satisfaction or evaluative standards (Ryan and Deci, 2000). When these motivations are mapped onto Hofstede's cultural dimensions, observable tendencies rather than deterministic patterns emerge. In cultures with high uncertainty avoidance, consumers typically perceive higher online shopping risks and thus rely more heavily on extrinsic cues, such as online reviews, to guide their decisions (Filieri and Mariani, 2021; Kim et al., 2023). In the specific context of Chinese wine e-commerce, this pragmatic risk aversion manifests clearly, driving consumers to depend heavily on external platform incentives rather than purely intrinsic evaluations (Duan et al., 2022; Rui et al., 2025). E-commerce platforms design external incentive mechanisms specifically considering this trait to stimulate user engagement; consequently, the review behaviors of Chinese consumers are predominantly driven by extrinsic motivation. Such external platform incentives likely contribute to a higher concentration of positive star ratings (Román et al., 2024). Furthermore, according to Hofstede's cultural dimensions theory, the degree of uncertainty avoidance exerts a profound impact on the establishment of consumer trust and risk perception mechanisms within the digital retail sector (Kim et al., 2023).

Conversely, Italian wine consumers are deeply embedded within a Mediterranean cultural context that emphasizes hedonic consumption, sociality, and aesthetic appreciation (Agnoli and Outreville, 2021). Consumers share their evaluations primarily for intrinsic self-expression and social connection rather than for external incentives. This cultural backdrop naturally aligns with intrinsic motivation, which is expected to result in a more evenly distributed spectrum of star ratings. By mapping the pathway from "cultural and platform context" to "consumer motivation" and ultimately to "review characteristics", this study elucidates how diverse cultural backgrounds shape consumers' emotional expressions. This structured approach provides a robust theoretical foundation for the subsequent empirical analysis (Kim et al., 2018; Filieri and Mariani, 2021). However, while the theoretical link between culture and consumer motivation is clear, effectively capturing and quantifying these culturally embedded emotional expressions from massive, unstructured online reviews requires advanced computational tools. This creates a critical need to bridge cultural theory with machine learning methodologies.

2.3. AI-driven sentiment analysis in consumer research

2.3.1. Sentiment analysis in consumer research

To address this methodological need and operationalize the aforementioned theoretical framework, the expansion of big data and artificial intelligence has fundamentally transformed traditional marketing analytics, providing the capability to decode complex consumer sentiments at scale (Bai and Yu-Buck, 2025). Although natural language processing (NLP) and sentiment mining have been widely proven effective across multiple industries (such as the hospitality and service sectors) (Ahani et al., 2019; Nilashi et al., 2022; Zhao and Huang, 2024), existing wine literature remains confined to chemical composition analysis or small-scale survey data (Sáenz-Navajas et al., 2013; Spence, 2020; Sinesio et al., 2021). Although research on consumer purchasing decisions is relatively extensive, there remains a significant gap concerning post-purchase satisfaction and emotional expression specifically within the wine sector (Mosikyan et al., 2024). While recent research has demonstrated the utility of machine learning in predicting beer flavor profiles (Schreurs et al., 2024), similar large-scale deep learning methodologies are seldom applied to wine sentiment analysis. Conducting sentiment analysis using big data entails complexity.

Analyzing consumer reviews in the food and beverage domain poses unique analytical challenges due to complex contextual nuances and domain-specific jargon, which traditional statistical sentiment models often struggle to decode (Barbierato et al., 2021; Gupta and Katarya, 2024). However, recent research in the consumer domain has gradually adopted the Transformer architecture, which possesses contextual understanding capabilities (Xiao et al., 2022). Research by Darraz et al. (2025) employing the BERT model in the e-commerce sector demonstrates that deep learning outperforms traditional lexicon-based methods in identifying consumers' implicit emotional drivers. Recent studies demonstrate the efficacy of Transformer architectures utilizing dynamic attention weights within e-commerce systems, these models are capable of identifying multi-dimensional sentiments and effectively mitigating the impact of review noise (Praveen et al., 2024; Ray and Singh, 2025).

2.3.2. Methodological challenges in cross-lingual settings

Despite the proven effectiveness of machine learning in sentiment recognition, current literature predominantly remains focused on monolingual environments or bulk commodities (Acheampong et al., 2021; Xiao et al., 2022). Cultural backgrounds and analytical methods influence the accuracy of sentiment models (Kim et al., 2018; Filieri and Mariani, 2021). To address the complexities of big data sentiment analysis, researchers have developed deep learning models such as convolutional LSTM (Co-LSTM) (Behera et al., 2021). However, difficulties persist in the analysis and comparison of online reviews across different languages. Comparisons between Chinese and English product reviews indicate that Chinese expressions are often implicit, whereas English expressions are relatively direct. Lo et al. (2017) and Brand and Reith (2022) emphasize that "high-context" and "low-context" cultures require the adoption of different language models to prevent the loss of sentiment data; directly translating consumer review data may overlook subtle linguistic differences.

Furthermore, platform incentive mechanisms can skew quantitative ratings, leading to a proliferation of extreme positive reviews and resulting in a "J-shaped" distribution of consumer star ratings (Schoenmueller et al., 2020; Buil et al., 2024). Consequently, analyzing semantic density offers a more objective proxy for reflecting consumers' intrinsic motivation and cognitive effort, effectively circumventing the rating inflation induced by platform incentives (Leung and Cho, 2024; Liu et al., 2024). For instance, Rui et al. (2025) compared the sentiment differences among online wine consumers in China, the United Kingdom, and the United States, which may neglect the variations in emotional expression within non-English speaking countries. As recent studies indicate, translating consumer review datasets frequently results in the loss of domain-specific terminology and cultural contexts inherent to the native language (Přibáň et al., 2024). Furthermore, when using artificial intelligence to analyze consumer behavior, it is necessary to resolve consumer trust issues (Kim et al., 2021) to ensure that models accurately and promptly translate consumer emotions (Otter et al., 2021). Consequently, research on multilingual NLP models for decoding cross-border consumer behavior indicate the necessity of pre-training and fine-tuning these models (Dashtipour et al., 2016; Ullah et al., 2025). Recent research has proven the effectiveness of machine learning in sentiment recognition and extraction directly within their native linguistic contexts (Zhao et al., 2021; Alantari et al., 2022; Darraz et al., 2025; Ray and Singh, 2025).

2.3.3. NLP as a theoretical bridge

To systematically address these methodological challenges, NLP methods can serve as a bridge to reveal the true theoretical drivers of consumer behavior. Traditional quantitative data derived from consumer review star ratings is highly susceptible to interference from platform incentives. By classifying unstructured consumer review data into positive, neutral, and negative sentiment categories, NLP facilitates consumer sentiment segmentation (Ahani et al., 2019; Behera et al., 2021; Xiao et al., 2022). Integrating these methods with consumer sentiment theory enables a more effective understanding and prediction of consumer behavior (Bagozzi et al., 1999; Yang et al., 2024).

Distinct from traditional product star ratings, NLP technology can identify specific semantic drivers that influence consumers' emotional expressions, such as price, quality, logistics, and taste (Praveen et al., 2024; Liu et al., 2024b; Ray and Singh, 2025). It can also effectively extract, quantify, and analyze these behavioral factors (Xiang et al., 2015; Guo et al., 2017). Ultimately, NLP technology can systematically decouple intrinsic emotional expressions from extrinsic, platform-driven functional feedback. By integrating NLP with consumer sentiment theory, this methodology effectively operationalizes the theoretical framework, thereby overcoming the inherent research biases associated with traditional star ratings (Antioco and Coussement, 2018; Wang et al., 2024; Baier et al., 2025).

2.4. Research gaps and proposed conceptual framework

Despite the theoretical potential of NLP outlined above, substantial methodological limitations persist in current literature. Given the limitations, a research gap remains in the online wine retail sector regarding the use of native language deep learning models for comparing consumer behavior across different backgrounds, particularly between China and Italy, which possess distinct wine market characteristics. This gap hinders the understanding of how different cultural backgrounds influence and shape consumer sentiments, as well as the underlying reasons for these emotional differences. To overcome this limitation, this study develops an integrative conceptual framework linking cultural context and consumer motivation within digital environments (see Fig. 1). Grounded in Hofstede's dimensions, the framework captures fundamental differences where the Chinese context exhibits high uncertainty avoidance and pragmatic orientation, while the Italian context reflects Mediterranean hedonism. These cultural orientations are embedded within distinct digital environments, which shape rather than solely determine the expression of consumer behavior.

Building on this foundation, the theoretical core is conceptualized through SDT, distinguishing between extrinsic motivation (driven by functional goals) and intrinsic motivation (driven by self-expression and aesthetics). By employing native language deep learning models (specifically, Chinese RoBERTa and Italian UmBERTo) as a methodological bridge, this research operationalizes these constructs into measurable empirical patterns. Consequently, motivations manifest in two layers: quantitatively, through distinct rating distributions (e.g., highly concentrated versus evenly distributed ratings), and qualitatively, through semantic focus on functional and pragmatic concerns versus emotional and aesthetic expressions. This structured mapping provides a theoretically grounded foundation for AI-driven sentiment analysis while highlighting how digital environments mediate culturally embedded behavior. However, the observed patterns should be interpreted as contextually embedded tendencies shaped by both cultural orientations and platform environments, rather than culture-only effects.

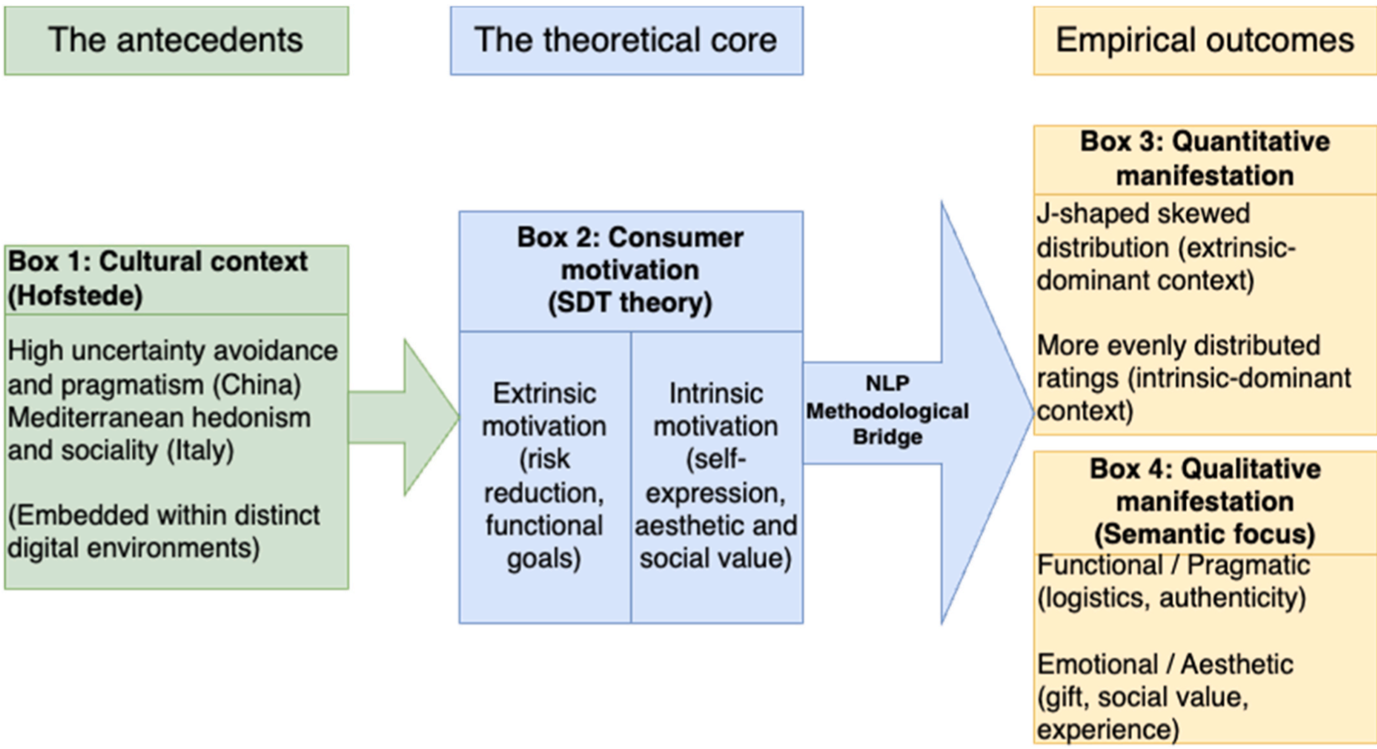

Fig. 1. Conceptual framework: Mapping cultural context to consumer motivation and review characteristics.

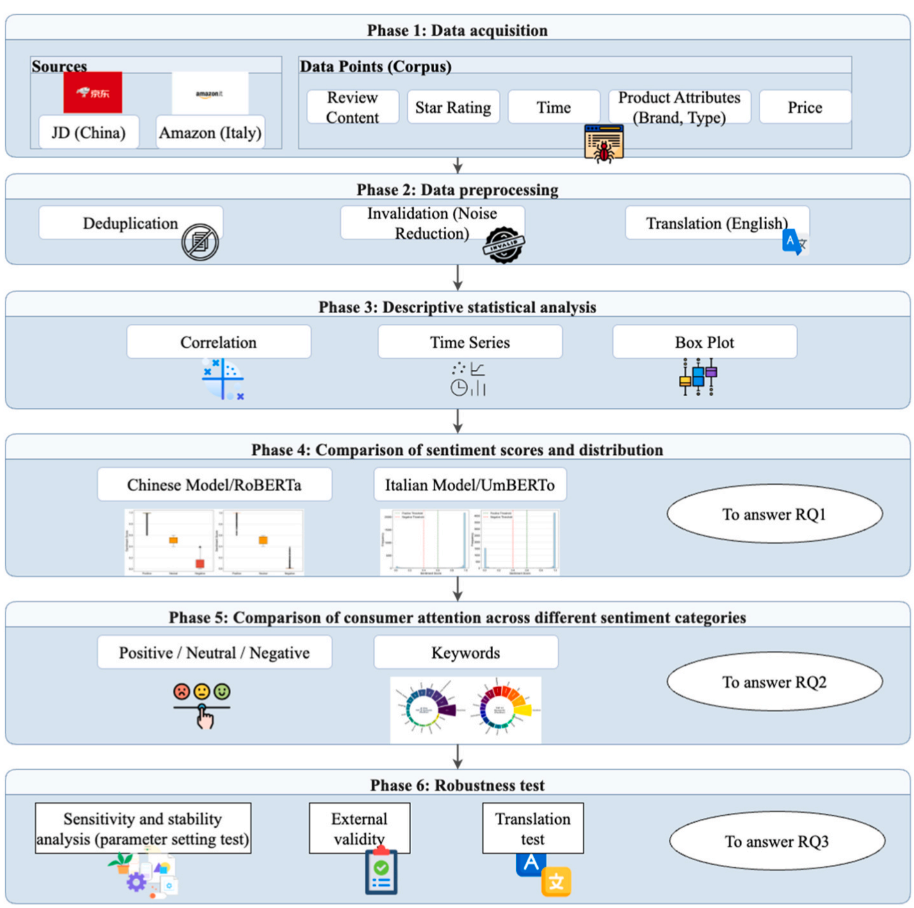

Fig. 2. Research methodology framework.

3. Methodology

In order to compare the consumer emotional expression differences in online wine retail, the Chinese and Italian wine markets are selected as comparison objects. This selection does not aim to verify consumer taste heterogeneity, which has been fully proven in beverage market research, but rather ensures the comparability of cross-cultural sentiment analysis. Italy represents a mature traditional market with a well-established wine evaluation system and low context language expression. China, as a rapidly expanding, highly digitalized emerging market, associates wine consumption with social and cultural capital, presenting high context language expression characteristics. By comparing these two distinct wine markets, the characteristics of consumer emotional expression under different cultural backgrounds can be researched utilizing artificial intelligence native language models, thereby forming a robust methodological framework applicable to other product categories.

3.1. Analytical framework

To address the research questions proposed in Section 1 and ensure the transparency and replicability of the empirical process, this study develops a transparent and replicable analytical pipeline. By systematically decoupling intrinsic emotional expressions from extrinsic, platform-driven functional feedback, this methodology overcomes the inherent research biases associated with traditional star ratings. To ensure theoretical operationalization, the framework explicitly maps computational NLP metrics to psychological and cultural constructs, strategically aligning technical execution with the study's core objectives. As illustrated in the research methodology framework (Fig. 2), the analytical pipeline is rigorously structured into six interconnected phases.

The initial stages focus on data preparation and foundational analysis. In Phase 1 (Data acquisition), consumer review data is acquired from JD.com (China) and Amazon. it (Italy), which primarily encompasses review content, star ratings, timestamps, prices, and product attributes for online retail wine. Following this, Phase 2 (Data preprocessing) executes rigorous deduplication, dynamic noise reduction, and English translation to facilitate subsequent robustness testing (in Section 4.3.4). Once the datasets are cleaned, Phase 3 (Descriptive statistical analysis) examines the correlations, distributional characteristics, and time-series trends of the raw review data, establishing a solid baseline for deeper computational modeling.

The subsequent stages deploy advanced computational techniques to extract theoretical insights and validate the empirical findings. To directly address RQ1, Phase 4 (Comparison of sentiment scores and distribution) employs the pre-trained Chinese RoBERTa model and the pre-trained Italian UmBERTo model for the JD and Amazon datasets respectively, outputting sentiment classifications and continuous sentiment scores. Next, to explicitly answer RQ2, Phase 5 (Comparison of consumer attention across different sentiment categories) integrates TF-IDF techniques (Salton and Buckley, 1988) with the sentiment subsets to extract and weight emotion-specific keywords. Specifically, Phase 4 and Phase 5 are no longer merely technical procedures but are designed to support the interpretation of theoretically relevant patterns: (1) keyword clusters related to functional attributes are interpreted as indicative signals potentially associated with uncertainty-related concerns; (2) semantic density and rating skewness are used as exploratory indicators related to the extrinsic and intrinsic motivation spectrum within SDT. Finally, to evaluate model efficacy and address RQ3, Phase 6 (Robustness test) conducts statistical and model parameter tests, alongside cross-language translation benchmarking, to confirm that the observed emotional differences stem from actual consumer behavior rather than native language model bias. In addition to the six-phase primary analytical pipeline, a supplementary validation using secondary wine-focused platforms was conducted to examine cross-platform and cross-method semantic consistency, as reported in Section 4.5.

3.2. Analysis tools and parameter settings

To extract the potential cognitive drivers of wine consumers, this study utilizes a hybrid and theory-driven analytical pipeline integrating deep learning for latent sentiment classification with an interpretable layer for explicit feature extraction (Acheampong et al., 2021). Transformer-based models were selected because they can capture contextual sentiment patterns more effectively than traditional keyword-based approaches, particularly when dealing with subtle emotional expressions and complex review language (Qiu et al., 2020; Otter et al., 2021).

For each market, this study did not train models from scratch. Instead, it deployed existing native-language models that had already been pretrained and fine-tuned for sentiment analysis. For the Chinese market, this study adopted a pre-trained Chinese RoBERTa model from the Hugging Face ecosystem. Specifically, we utilized the “uer/roberta-base-finetuned-jd-binary-chinese” model developed by the UER team, which was fine-tuned explicitly for the massive e-commerce review corpus of JD.com (architectural parameters and pre-training configurations are publicly accessible via https://huggingface.co/uer/roberta-base-finetuned-jd-binary-chinese). For the Italian market, the Italian UmBERTo model was adopted, specifically the “MilaNLProc/feel-it-italian-sentiment” model (architectural parameters and pre-training configurations are publicly accessible via https://huggingface.co/MilaNLProc/feel-it-italian-sentiment). Compared to the original multilingual BERT, native-language Transformer-based models trained on massive national corpora can more accurately identify specific linguistic characteristics and subtle emotional nuances (Přibáň et al., 2024). The phenomenon where UmBERTo outputs sentiment distributions clustered near 0 and 1 is a known mathematical feature of softmax saturation within specific fine-tuned Transformer architectures. This structural clustering indicates high model confidence in bipolar classification tasks, rather than a loss of analytical reliability (Gawlikowski et al., 2023).

The model deployment for both wine markets utilized the PyTorch pipeline to achieve standardization. Regarding parameter settings, the models were instantiated through the Hugging Face pipeline API, ensuring identical parameter configurations across both models. The maximum sequence length was set to 512 tokens, supporting dynamic truncation, parallel inference across multiple CPU cores, and enabling smart hardware acceleration (CUDA or Apple Silicon MPS). To operationalize the theoretical framework in a transparent and replicable manner, the analytical process was structured into three methodological steps:

(1) Step 1: Latent sentiment classification. The Transformer-based models (RoBERTa and UmBERTo) classify each review into positive, neutral, or negative categories. To ensure robustness in classification, strict probability thresholds were established: reviews with scores $\geq$0.6 were marked as positive, those $\leq$0.4 were marked as negative, and intermediate values were considered neutral.
(2) Step 2: Text preprocessing and lexical filtering. Following sentiment classification, language-specific tokenizers (Jieba for Chinese and SpaCy for Italian) are applied to segment the text, remove stop words, and retain evaluation-relevant parts of speech (primarily nouns and adjectives).
(3) Step 3: Keyword salience identification via TF-IDF. The TF-IDF algorithm is then applied within each sentiment subset to identify and rank the most salient terms. While more advanced techniques such as attention-based keyword extraction or embedding clustering could be applied, TF-IDF is intentionally selected to enhance interpretability and ensure transparency in identifying actionable lexical drivers. This approach provides

transparent lexical anchors, such as “Logistics” and “Packaging”, that can be directly linked to consumer interpretation and managerial discussion.

To address potential comparability issues arising from utilizing two distinct pretrained models (RoBERTa and UmBERTo), it is critical to clarify that our cross-cultural comparison does not rely on comparing the absolute raw continuous scores between the two models. Due to differing training corpora and architectural calibrations, direct numerical comparison of raw probabilities is methodologically flawed. Instead, our comparative framework is built upon the discrete sentiment categories (Positive, Neutral, Negative). Theoretical literature suggests that introducing a symmetric margin around the absolute neutral point (e.g., 0.4 to 0.6) effectively filters out ambiguous, mixed, or objective expressions, thereby enhancing the precision of polar classification (Salehan and Kim, 2016; Yadav and Vishwakarma, 2020). The fixed thresholds (0.6 and 0.4) were established not to force numerical equivalence, but to isolate functionally equivalent semantic subsets within each distinct linguistic environment. The validity of these thresholds was further assessed through the threshold sensitivity analysis reported in Section 4.4.2.

3.3. Data collection and cleaning

Data collection and cleaning were conducted using standardized and reproducible Python scripts, which helps reduce potential research bias. For JD. com (China), reviews were retrieved using the keyword “葡萄酒” (grape wine). The selection of JD. com and Amazon. it is justified by their status as the dominant, culturally representative e-commerce ecosystems in their respective nations. While platform mechanisms differ, these differences are considered part of the broader digital context in which culturally embedded review behaviors are expressed, rather than being treated as noise. For instance, JD's extrinsic reward system and Amazon's relatively organic review environment may shape how consumers express their experiences within their respective market contexts. By selecting these leading platforms, the study captures “platform-mediated culture”, which is the most authentic form of consumer behavior in the digital age. Based on default product recommendation rankings, data were obtained from the products listed in, Appendix A resulting in 43,353 raw reviews. Similarly, for Amazon. it (Italy), equivalent search parameters using the term “Vino” (grape wine) were applied to identify the products in Appendix B, from which 9864 raw reviews were extracted as a comparative corpus. To reduce product-level structural bias, product selection in both markets followed comparable platform default ranking logic and focused on mainstream retail wine categories rather than niche or premium-only segments.

Following the collection of raw data, multi-stage data cleaning and noise reduction processes were conducted. To ensure data timeliness, the dataset was restricted to reviews published in 2021 and onwards. Subsequently, duplicate data was removed; completely identical rows and duplicate content from the same user were entirely deleted. For duplicate content across different users, an algorithm was employed for intelligent discrimination; such content was deleted only if it exhibited at least two distinct characteristics (e.g., marketing templates and abnormal user behavior). After applying the temporal restriction and deduplication, a multi-dimensional scoring model was utilized to identify fraudulent or “brushing” behaviors (Hu et al., 2012; Luca and Zervas, 2016; Zhang et al., 2016; Kumar et al., 2018; Plotkina et al., 2020; Zhao et al., 2025). The six dimensions included: (1) high-frequency users with more than 10 reviews (weight 2); (2) repetitive marketing template reviews (weight 1.5); (3) generic phrase stuffing evaluated via dynamic length thresholds (weight 1); (4) lengthy, structured reviews exceeding 100 characters that matched marketing templates (weight 1); (5) extremely short five-star evaluations containing only 5 to 8 characters (weight 0.5); and (6) concentrated time periods where the daily review volume exceeded 50 (weight 0.5). Reviews with a cumulative score exceeding 3.5 were deleted. This 3.5 threshold was established after manual testing of 500 reviews, which confirmed that reviews scoring above this threshold had a 92% probability of being invalid.

Following the removal of invalid reviews, a language-adaptive noise filtering mechanism was applied to texts exceeding 50 tokens (characters for Chinese, words for Italian). This minimum length ensures noise detection is applied only to sufficiently informative texts. To eliminate generic phrase stuffing (e.g., repetitive lexical patterns), reviews were excluded if the noise proportion (operationalized as the relative frequency of repeated or redundant tokens within each review) exceeded 12% of the total tokens for Chinese and 15% for Italian. This calibration accounts for the higher semantic density of Chinese characters, ensuring functional equivalence across both linguistic contexts. Ultimately, the analytical corpus retained 27,126 JD. com reviews (62.57%) and 6338 Amazon. it reviews (64.25%).

3.4. Ground truth validation

To verify the reliability of the Transformer-based models and ensure that the sentiment categories (positive, neutral, negative) are comparable across the distinct linguistic architectures of RoBERTa and UmBERTo, a human-annotated benchmark validation was conducted. A stratified random sample of 400 reviews was extracted from the cleaned corpus, comprising 200 reviews from JD.com and 200 from Amazon.it. The stratification was based on the distribution of model-predicted sentiment categories to ensure balanced representation across positive, neutral, and negative classes, thereby reducing class imbalance bias in performance evaluation and providing a sample size sufficient to yield statistically stable estimates of classification performance metrics in NLP validation settings. To minimize cultural and linguistic bias, two independent annotators, each a native speaker of the respective language (Chinese and Italian) with bilingual proficiency, manually classified the reviews into three sentiment categories while remaining blinded to the model predictions. The annotation protocol was developed based on prior e-commerce sentiment classification studies and refined through pilot coding before formal labeling (Xiang et al., 2015; Ray and Singh, 2025).

Inter-rater reliability was assessed using *Cohen's Kappa* coefficient, yielding scores of 0.85 for the Chinese dataset and 0.83 for the Italian dataset, indicating strong agreement. In cases of disagreement, primarily involving implicit expressions or boundary ambiguity between neutral and polarized sentiment (Basu et al., 2024), a third senior researcher adjudicated to establish the final ground truth labels. The manually annotated labels were then compared with model predictions using predefined probability thresholds ($\geq$0.6 for positive, $\leq$0.4 for negative), which were determined a priori and further evaluated through sensitivity analyses using alternative cutoffs (e.g., 0.5 and 0.7), as detailed in Section 4.4.2. Model performance was evaluated using precision, recall, and macro-averaged *F1* scores (Birjali et al., 2021), with the Chinese RoBERTa model achieving an *F1* score of 89.5% (*precision* = 90.1%, *recall* = 88.9%) and the Italian UmBERTo model achieving an *F1* score of 87.2% (*precision* = 88.4%, *recall* = 86.1%).

These results demonstrate strong agreement between human annotations and model predictions within each linguistic context. Importantly, the use of language-specific annotators ensures that sentiment labels are grounded in native linguistic and cultural interpretation, rather than translation-based approximations (Přibáň et al., 2024). While this validation does not directly test cross-linguistic equivalence, the consistently high performance across both datasets provides supporting (rather than definitive) evidence that the classification thresholds capture comparable sentiment boundaries, thereby supporting the validity of subsequent cross-cultural comparisons.

4. Results of the analysis

4.1. Data overview and preliminary analysis

This section provides a statistical comparison of online consumption data between the Chinese and Italian platforms, focusing particularly on wine consumer review star ratings (1 to 5 stars), their temporal distribution, and evolutionary trends. First, the distribution of review star ratings in Fig. 3 reveals significant differences between the datasets of the two platforms. The JD dataset exhibits a skewed "J-shaped" distribution, predominantly consisting of five-star reviews, with all other star ratings falling below 10%. In contrast, the distribution of review star ratings in the Amazon dataset is relatively more even; although five-star reviews still constitute the majority, the proportions of three- and four-star reviews within the overall dataset are notably higher. Considering that consumers' online review star ratings are susceptible to the environmental factors of the respective sales platforms, direct numerical comparisons of these ratings may fail to yield valid insights. Therefore, to enhance the comparability of the consumption data, this study shifts its focus from comparing the absolute values of review star ratings to analyzing text sentiment and semantic density, thereby more accurately reflecting the consumer experience.

Second, regarding changes in temporal distribution, the directional outcomes of the review time-series tests for the two platforms from 2021 to 2025 diverge. The overall time series for JD wine consumer review star ratings demonstrates a statistically significant upward trend ($\tau = 0.5091$, $p < 0.001$). Examining the specific trajectory of the curve, JD experienced a noticeable decline at the end of 2021, rapidly rebounded in early 2022, and subsequently maintained a relatively high level from 2024 to 2025. Conversely, the overall time series for Amazon wine consumer review star ratings exhibits a statistically significant negative trend ($\tau = -0.4629$, $p < 0.001$). Although it remained relatively stable between 2021 and 2024, it began a gradual descent after 2024, followed by a slight rebound after a sharp drop in late 2025. Furthermore, the extreme star rating values observed in certain months may reflect small-sample variability. Therefore, cross-platform trend interpretations should be considered alongside the corresponding monthly sample sizes.

Thirdly, fundamental structural differences exist between the ideographic and phonetic languages utilized in the review datasets of the two platforms. Chinese and Italian differ in text length, measurement units, and tokenization methods, as ideographic languages record meaning directly through symbols, whereas phonetic languages spell out sounds using letters. To enhance the comparability of cross-lingual consumption reviews, this study employs character counts for the Chinese dataset and word counts for the Italian dataset to achieve redundancy normalization. The results (Fig. 4) indicate that the median semantic density of the JD dataset is below 30, peaking at 250; conversely, the Italian dataset exhibits a broader distribution range for semantic density, reaching approximately 400. These variations in linguistic expression are attributable to both inherent language characteristics and the respective platform ecosystems. Concurrently, the correlation analysis between semantic density and review star ratings reveals that the linear correlations for both platforms are relatively weak: the JD dataset exhibits a weak positive correlation ($r = 0.071$), while the Italian dataset presents a weak negative correlation ($r = -0.073$). Although this correlation is statistically significant, the massive sample size renders its practical effect size marginal. Consequently, this result should be interpreted merely as a weak behavioral tendency and does not dictate the actual length of the reviews.

Notably, when examining the distribution across review groups, the semantic density of JD reviews exhibits higher dispersion across different star rating contexts, accompanied by highly concentrated density outliers. In contrast, the main quartile ranges of Amazon reviews remain relatively stable and consistent overall. This phenomenon may be associated with extrinsic motivation as defined in SDT. JD encourages users to write reviews by offering external rewards, such as accumulable cash deductions for future purchases, to those who post consumption reviews. Conversely, Italian wine consumers may be more driven by internal self-expression rather than external incentives.

4.2. Sentiment score distribution and comparison

Following the analysis using native language machine learning models, Fig. 5 illustrates the sentiment classifications (positive, neutral, and negative) and sentiment scores of online wine reviews from JD.com and Amazon. it, thereby addressing RQ1. The sentiment scores of the JD dataset are predominantly concentrated on the positive end, whereas the negative sentiment scores exhibit significant fluctuations. This indicates that the RoBERTa model, deployed via Hugging Face, can relatively effectively capture consumers' emotions of dissatisfaction. Conversely, the sentiment score distribution for Amazon is dominated by 0 and 1, and neutral sentiment scores are extremely weak. The phenomenon where UmBERTo outputs sentiment distributions clustered near 0 and 1 is a known mathematical feature of softmax saturation within specific fine-tuned Transformer architectures. This structural clustering demonstrates that the model possesses high confidence in bipolar classification tasks, rather than indicating a loss of analytical reliability. An interesting finding is the pronounced discrepancy between the sentiment score distributions generated by the native

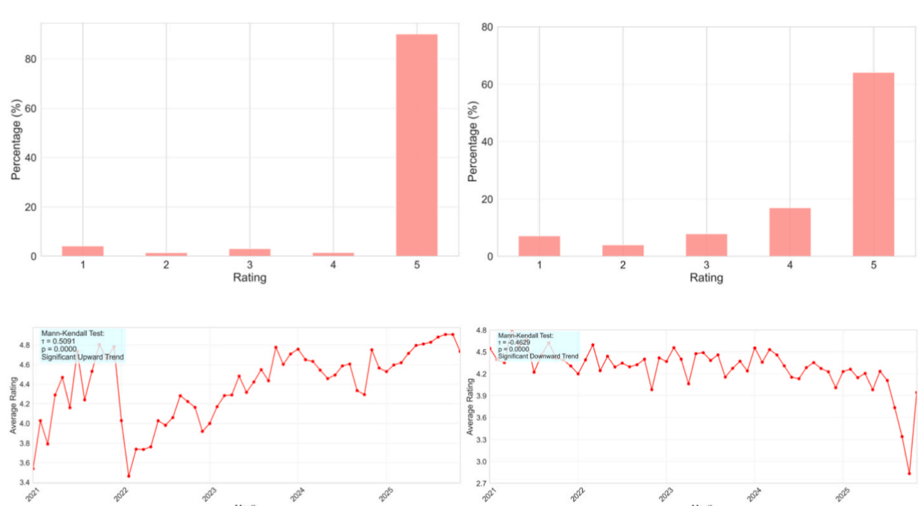

Fig. 3. Comparative analysis of rating distribution and temporal trends: JD (left) vs. Amazon (right).

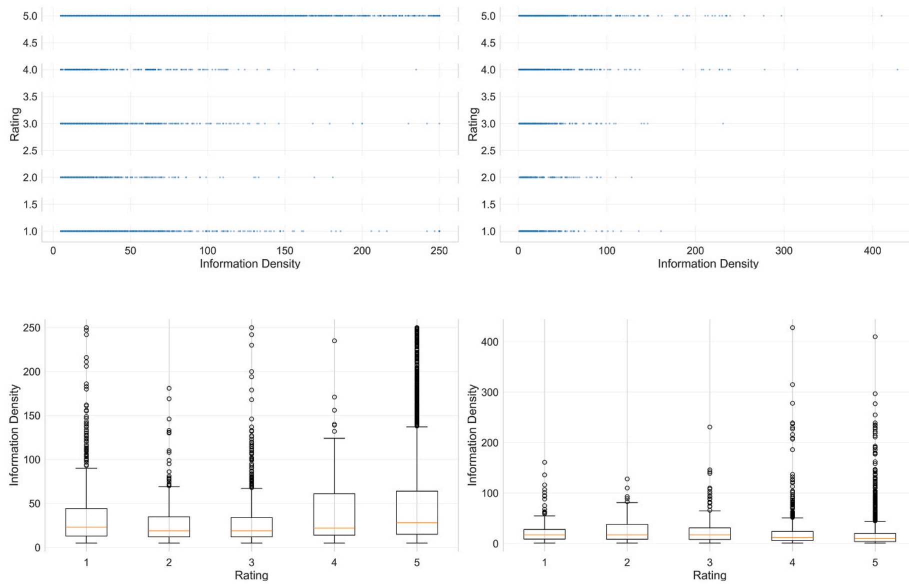

Fig. 4. Distribution characteristics of information density across ratings: JD (left) vs. Amazon (right)
Note on Metrics: Different metrics are applied to normalize information density across languages. Character Count is used for Chinese (logographic script), while Word Count is used for Italian (alphabetic script), as a single Chinese character often carries semantic weight comparable to a full Italian word.

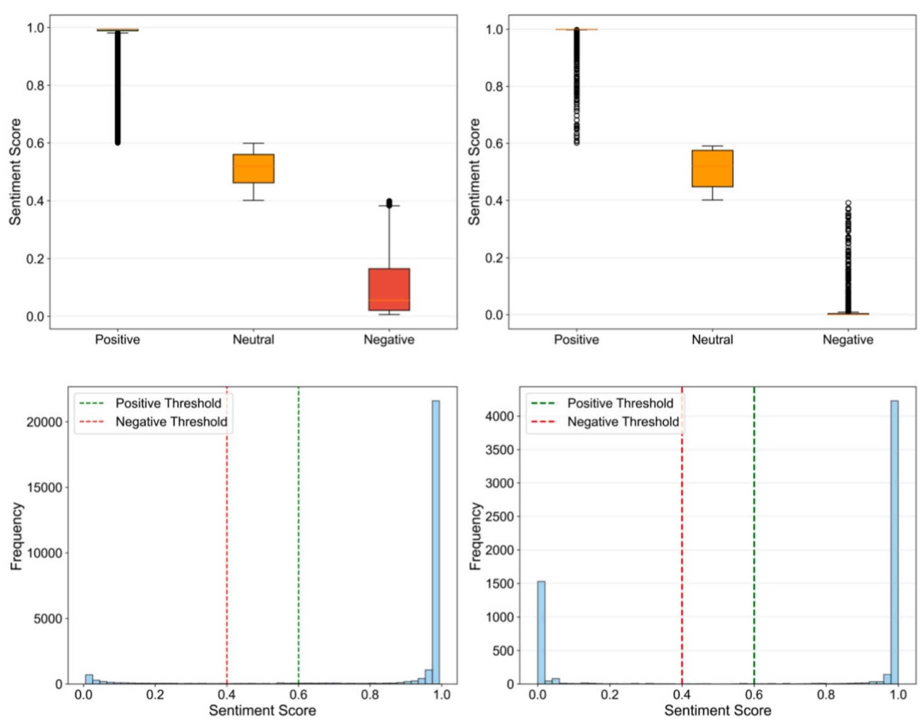

Fig. 5. Distribution and classification of sentiment scores (Addressing RQ1): a comparative analysis of JD (left) and Amazon (right).

language machine learning models and the consumers' raw review star ratings.

4.3. Semantic focus and TF-IDF keyword analysis

To investigate the focal points of online retail wine consumers under varying sentiments, this study employs the TF-IDF method to analyze keyword weights across different emotional states (as illustrated in Fig. 6, Figs. 7 and 8), thereby addressing RQ2. This comparative framework demonstrates the relative influence of different product attributes on consumer sentiments within distinct market environments. To ensure comparability across different consumer datasets, the linear min-max scaling method was utilized to standardize the raw TF-IDF score weights to a range of 0 to 100. This mathematical procedure eliminates discrepancies in data volume and linguistic density between datasets of different languages, enabling the comparison of consumer attention weights on an equivalent scale.

As illustrated in Fig. 6, within the positive sentiment distribution, JD wine consumers prioritize the product experience, which is primarily characterized by "Mouthfeel" (sensory dimension) and "Logistics"

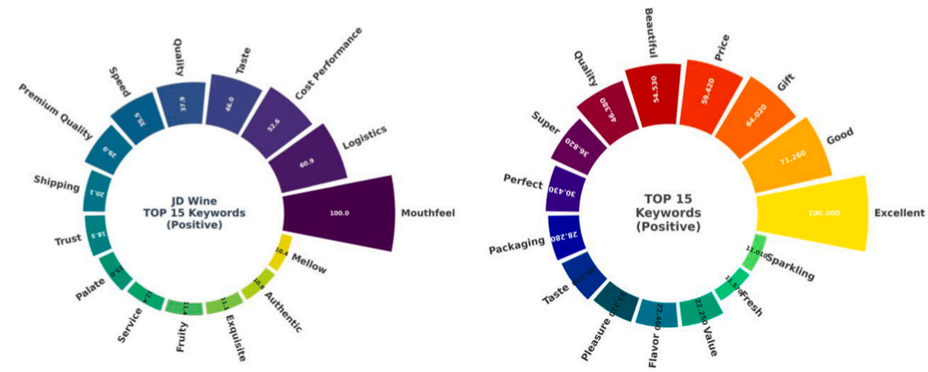

Fig. 6. Comparative analysis of Top 15 positive keywords based on TF-IDF (Addressing RQ2): JD (left) vs. Amazon (right).

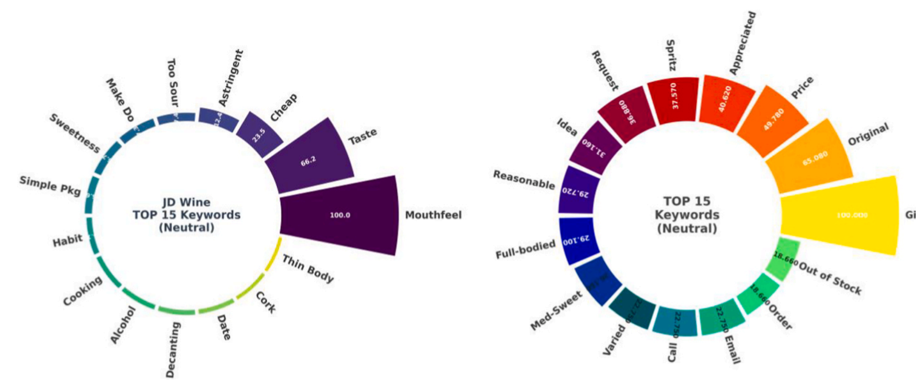

Fig. 7. Comparative analysis of Top 15 neutral keywords based on TF-IDF (Addressing RQ2): JD (left) vs. Amazon (right).

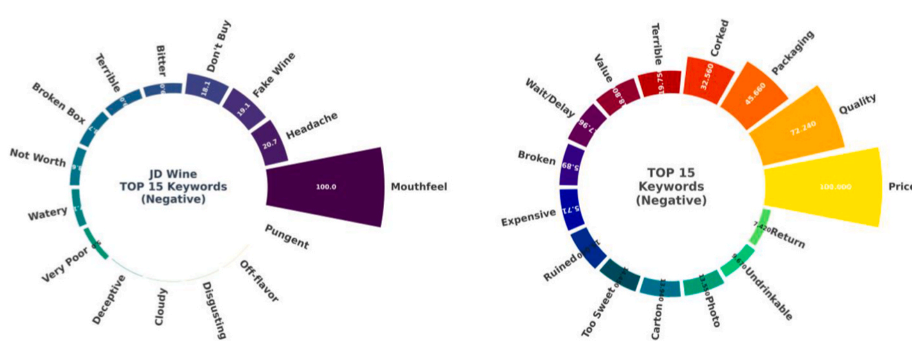

Fig. 8. Comparative analysis of Top 15 negative keywords based on TF-IDF (Addressing RQ2): JD (left) vs. Amazon (right).

(service dimension). Chinese wine consumers focus more on the functional attributes and logistical services of the products; this may reflect a stronger tendency toward function-oriented evaluation among Chinese wine consumers. Conversely, Amazon wine consumers emphasize the "aesthetic" dimension, where high-sentiment words such as "Excellent" and "Good" occupy dominant positions, followed by "Gift" and "Price". This tendency may be associated with the expressive habits of Italian wine consumers, or it may indicate that they place greater emphasis on the social attributes and aesthetic characteristics of the wine itself.

Although the sample size for neutral sentiment remains relatively constrained, consumers in both markets demonstrate stable priorities that align with the thematic orientations observed in other sentiment categories. As illustrated in Fig. 7, within the framework of neutral sentiment, wine consumers on the JD platform appear to remain primarily focused on "Mouthfeel", followed by specific sensory descriptors such as "taste", "astringent" and "sour". In contrast, consumers on Amazon emphasize service-oriented factors including "gifts", "orders" and "out of stock", which relate more to logistics and fulfillment than the intrinsic properties of the product. These findings are highly consistent with the consumer concerns identified in the positive sentiment analysis.

As illustrated in Fig. 8, the keyword rankings for JD and Amazon wine consumers under negative sentiment reveal distinct priorities. JD consumers exhibited significant concerns regarding food safety, with terms such as "Headache", "Fake Wine", and "Don't Buy" closely following "Mouthfeel". Conversely, Amazon wine consumers focused more heavily on economic and logistical factors; "Price", "Quality", and "Packaging" emerged as their primary concerns, demonstrating a low tolerance for quality and packaging issues coupled with pronounced price sensitivity. The factors driving these consumer discrepancies between the two countries may be associated with market maturity. The Italian wine consumption market appears more mature, whereas food safety issues currently remain the primary concern within the Chinese wine market.

4.4. Robustness test

4.4.1. Descriptive statistics and distributional characteristics

To further validate the reliability of the research findings, a robustness check was performed. Initially, descriptive statistical tests were conducted on the datasets. As presented in Table 1, the descriptive statistics for sentiment analysis indicate divergent score distributions between the two platforms. For JD.com, the sample mean sentiment score ($N = 27126$, $M = 0.9006$, $SD = 0.2499$) appears notably higher than that of Amazon.it ($N = 6338$, $M = 0.7190$, $SD = 0.4387$). Regarding data distribution, the JD dataset exhibits a negative skewness of -2.7864, indicating that the positive sentiments of wine consumers are highly concentrated. The Amazon dataset also demonstrates a negative skewness of -0.9786, though its positive sentiment distribution is relatively more even compared to JD.

The results in Table 2 reveal that both the JD and Amazon datasets deviate from a normal distribution (Jarque-Bera test ($p < 0.001$), Kolmogorov-Smirnov test ($D = 0.6248$, $p < 0.001$) and Mann-Whitney U test ($p < 0.001$)). These findings further indicate discrepancies in the sentiment distributions generated by the Chinese and Italian native language machine learning models; the two datasets represent fundamentally distinct statistical populations, with an observed effect size of *Cohen's* $d = 0.6152$. Unlike the previously observed weak correlation regarding semantic density, this medium effect size demonstrates that the emotional differences between the wine markets of the two countries are not merely a statistically significant artifact generated by massive data, but rather reflect genuine market differences. This supports that the observed sentiment distributions differ meaningfully across the two market datasets, indicating that the cross-market differences are not merely statistical artifacts.

4.4.2. Sensitivity and stability analysis

Sensitivity and stability analysis indicates that both native language models maintain robustness across different sentiment thresholds. In the initial sentiment score threshold settings, scores above 0.6 were classified as positive, and those below 0.4 were classified as negative. To verify the models' efficacy under varying sentiment score thresholds, this study conducted sensitivity tests ranging from 0.5 to 0.7 in increments of 0.1. The results in Table 3 demonstrate that the Chinese model consistently classifies a large proportion of reviews (89% to 91.2%) as positive, while the Italian model maintains a stable classification rate of 71.4% to 72%. Crucially, the substantial divergence in positive sentiment proportions between the two cohorts (approximately 17-20%) persists regardless of cross-model threshold permutations. Furthermore, the Bootstrap confidence intervals in Table 1 provide additional support for the stability of the sentiment scores across both models. This suggests that the machine learning sentiment analysis results are likely to reflect meaningful emotional differences among online retail wine consumers, rather than being mere artifacts resulting from cross-model calibration disparities or baseline threshold selections.

Table 1
Descriptive statistics and Bootstrap analysis of sentiment scores.

Statistic	JD (RoBERTa)	Amazon (UmBERTo)
Sample Characteristics		
Observations (N)	27,126	6338
Mean	0.9006	0.719
Median	0.9918	0.9994
Std. Deviation	0.2499	0.4387
Skewness	−2.7864	−0.9786
Kurtosis	6.2802	−1.0122
Bootstrap Analysis 95% CI		
Mean Estimate	0.9006	0.7189
Lower Bound	0.8977	0.7083
Upper Bound	0.9035	0.7298
CI Width	0.0058	0.0215

Note: CI = Confidence Interval. Bootstrap analysis was performed with 1000 resamples.

Table 2
Statistical tests for distributional differences and normality.

Test	Platform	Statistic	p-value	Result
Normality Tests				
Jarque-Bera	JD	79,678.47	<0.001	Non-normal
Jarque-Bera	Amazon	1282.26	<0.001	Non-normal
Comparative Tests				
Kolmogorov-Smirnov	JD vs. Amz	0.6248	<0.001	Distinct Dist.
Mann-Whitney U	JD vs. Amz	5.99E+07	<0.001	Median Diff.
Effect Size				
Cohen's d	JD vs. Amz	0.6152	–	Medium Effect

Table 3
Sensitivity of sentiment classification to threshold variations.

Threshold	JD (RoBERTa)			Amazon (UmBERTo)		
	Pos(%)	Neg(%)	Neu(%)	Pos(%)	Neg(%)	Neu(%)
0.5	91.2	8.8	0	72	28	0
0.6	90.3	8.0	1.7	71.8	27.7	0.5
0.7	89	7.3	3.7	71.4	27.5	1.1

Note: Classification rules: Positive (≥Threshold), Negative (<1− Threshold). Neutral values near zero reflect overlapping boundaries at low thresholds.

4.4.3. Contamination test

Table 4 presents the correlation analysis between the fake review detection scores obtained during data cleaning and the sentiment scores from the two native language models. The results demonstrate that neither the JD ($r = -0.0036$, $p = 0.55$) nor the Amazon ($r = -0.0193$, $p = 0.12$) datasets exhibit statistically significant correlations. This indicates that the sentiment scoring models are statistically independent of the fake review detection algorithm. Regardless of whether a review is classified as fake, the sentiment analysis remains unbiased, thereby providing additional support for the robustness of the sentiment analysis results.

4.4.4. Translation test

To further verify whether the observed sentiment differences stem from model bias, the reviews from both datasets were translated into English and analyzed using the same English model (BERT-base-uncased); these results were subsequently compared with those generated by the native language models. The results in Table 5 demonstrate that the magnitude of change in sentiment differences is minimal; the sentiment discrepancy between the two languages increased from 18.5% using the original models to 19.8% using the English model. This suggests that the sentiment analysis results likely originate from meaningful differences in consumer behavior and cultural backgrounds, rather than being artifacts of processing biases within the native language models.

Table 4
Pearson correlation between sentiment scores and brushing (fake review) Scores.

Platform	Correlation (r)	p-value	Significance
JD (RoBERTa)	−0.0036	0.55	Not Significant
Amazon (UmBERTo)	−0.0193	0.124	Not Significant

Table 5
Consistency of sentiment distribution across models (Addressing RQ3).

Consumer Group	Original Model (Positive %)	English-Unified Model (Positive %)	Absolute Difference
Chinese Consumers	90.30%	84.90%	5.40%
Italian Consumers	71.80%	75.70%	3.90%
Cross-Cultural Gap	**18.50%**	**19.80%**	**1.30%**

This further validates the efficacy of employing native language models, thereby addressing RQ3.

4.5. Supplementary validation using wine-focused secondary platforms

To examine whether the core semantic orientations identified in the primary JD.com and Amazon.it comparison remain observable beyond large general-purpose e-commerce platforms, a supplementary validation was conducted using two wine-focused digital platforms: Jiuxian.com in China ($n = 1441$) and Vivino.it in Italy ($n = 1786$). The detailed URLs of these validation datasets are provided in Appendix C. Jiuxian.com represents a native vertical wine retail platform, whereas Vivino.it represents a wine-focused review and retail ecosystem. To ensure methodological comparability, the data cleaning and noise reduction protocols applied to these secondary datasets were kept consistent with those used for the primary platforms. Therefore, this validation was not intended to fully isolate platform effects, but to assess whether the higher-order semantic patterns remain observable in more wine-specialized digital environments.

Step 1: Cross-platform validation using TF-IDF. By applying the primary TF-IDF methodology to these datasets, this study examined the consistency of the semantic focus. As presented in Table 6, Chinese consumers on Jiuxian.com consistently emphasized dimensions associated with logistics reliability (“Shipping fast”), packaging integrity (“Packing”), and quality assurance (“Quality wine”). These findings remain broadly consistent with the primary TF-IDF analysis, where “Logistics” and “Fake wine” emerged as prominent concerns. This suggests that, for Chinese consumers, structural integrity and transactional reliability are closely linked to overall product evaluation. Conversely, Italian consumers on Vivino.it continued to exhibit a sensory-aesthetic orientation, such as “Champagne”, “Citrus” and “Fruity”, remaining broadly consistent with the patterns identified in the Amazon.it dataset.

Step 2: Methodological validation using ABSA. An ABSA model was deployed to further examine these semantic tendencies (Zhao et al., 2021). Table 7 indicates that the extracted aspect-polarity pairs reflect distinct evaluative orientations. On Jiuxian.com, aspects related to quality and transactional concerns, including “Quality”, “Shipping” and “Packing”, showed a high concentration of positive sentiment, reflecting a risk-sensitive evaluation orientation centered on authenticity and transactional reliability. In contrast, Italian consumers on Vivino.it extracted specific intrinsic attributes, including “Fruity”, “Aroma” and “Acidity”, with a high proportion of neutral sentiment. This indicates that the Italian discourse in this specialized environment reflects a descriptive sensory evaluation style, rather than relying on predominantly satisfaction-oriented evaluations (Danner et al., 2017).

Table 6
Step 1 comparison of top TF-IDF keywords on secondary platforms.

Rank	Jiuxian.com (China)	TF-IDF Score	Vivino.it (Italy)	TF-IDF Score
1	Shipping fast	0.1023	Champagne	0.0177
2	Packing	0.0954	Citrus	0.017
3	Repeated purchase	0.0803	Excellent	0.0166
4	Gift for parents	0.0764	Fruity	0.0125
5	Quality wine	0.0719	Fresh	0.0118

Taken together, the TF-IDF and ABSA results provide convergent evidence that the higher-order semantic orientations identified in the main analysis remain observable in the supplementary datasets. While specific terminology shifted towards more vertical-specific concerns, the fundamental focus of Chinese consumers remained centered on risk-sensitive evaluation concerns, whereas Italian consumers consistently prioritized a sensory-descriptive evaluation style. These findings could be interpreted as supplementary evidence of semantic consistency across the primary and secondary platform settings rather than as full-scale external validation or a strict control for platform effects.

5. Discussion

5.1. Integrative framework and answers to research questions

By decoupling textual sentiment scores from platform star ratings using native-language NLP models, this study addresses RQ1 and RQ2 by identifying differences in sentiment distribution and semantic focus between Chinese and Italian online wine consumers.

Regarding rating distributions, JD consumers exhibit a “J-shaped” pattern with slight increases over time and numerous semantic density outliers. These outliers stem from extreme evaluation distributions typical of e-commerce (Schoenmueller et al., 2020) and external platform rewards (Zaman et al., 2023), which heavily skew behaviors toward extreme positive reviews (Liang et al., 2025). Consequently, the presence of these outliers underscores that Chinese consumers' rating behaviors are largely compliance-driven for incentives, creating a disconnect between linguistic expressions and inflated star ratings. Integrating SDT with Hofstede's cultural dimensions, this phenomenon aligns with external reward-seeking (Ryan and Deci, 2000) and cultural norms of interpersonal harmony (Kim et al., 2018). Conversely, Amazon consumers show a more even rating distribution that declines slightly over time, exhibiting a weak negative correlation with semantic density. Their relatively stable semantic density may suggest a comparatively stronger role of intrinsic motivation and self-expression.

In terms of sentiment scores and distributions, JD consumers focus on sensory and functional indicators like “Mouthfeel” and “Logistics”, often accompanied by anxieties over “Fake wine”. This semantic focus is broadly consistent with a stronger tendency toward uncertainty avoidance, rather than serving as direct evidence of it. Consumer evaluation tendencies appear to be associated with broader cultural and market contexts; in Asian e-commerce, concerns about fraudulent information are frequently linked to stronger negative sentiments (Dang-Van et al., 2026). Consequently, consumers in these emerging markets appear to place greater emphasis on risk mitigation than on aesthetic experience (Kendall et al., 2019; Kim et al., 2023). In contrast, Amazon consumers emphasize “Gift”, “Taste”, and overall aesthetic appreciation, which SDT aligns with the psychological dimension of Autonomy. Driven by intrinsic motivation, Italian consumers use detailed sensory vocabulary to express personal aesthetic values and cultural heritage (Agnoli and Outreville, 2021; Bazzani et al., 2024). Their preference for terms like “Excellent” and “Good” likely ties to cultural norms of high-intensity emotional expression (Román et al., 2024).

From the perspective of cultural differences, these divergent behaviors may be associated with the interaction between cultural contexts and platform environments, rather than platform structures alone. Consequently, this study delineates these findings as “platform-mediated cultural expression”. While platform heterogeneity may shape review behaviors, the observed cross-cultural differences are unlikely to be explained solely by platform artifacts. Our robustness tests, using an English-unified model, demonstrate that core emotional divergences persist even when standardized across languages (RQ3). Furthermore,

Table 7
Step 2 comparison of top aspect sentiment distribution from ABSA validation.

Rank	Jiuxian.com (China)	Frequency	Positivity (%)	Negativity (%)	Vivino.it (Italy)	Frequency	Positivity (%)	Negativity (%)
1	Quality	971	97.80	1.10	Quality	198	87.90	0.50
2	Shipping	873	91.30	0.50	Fruity	177	10.70	3.40
3	Packing	845	98.50	0.20	Aroma	159	5.00	0.60
4	Price	421	92.40	0.20	Acidity	104	9.60	6.70
5	Color	376	89.90	0.00	Taste	68	16.20	1.50
6	Aroma	364	89.30	0.00	Complexity	57	14.00	3.50
7	Fruity	362	89.00	0.00	Price	55	43.60	3.60
8	Tannins	338	87.60	0.30	Body	52	9.60	1.90
9	Origin	299	87.60	0.30	Color	48	6.20	4.20
10	Taste	228	86.00	1.80	Packing	26	7.70	3.80

Note: The remaining percentage for each aspect represents neutral sentiment. The high proportion of neutral sentiment in the Italian dataset reflects a descriptive sensory evaluation style rather than predominantly satisfaction-oriented evaluations.

the supplementary validation conducted on Jiuxian.com and Vivino.it provides additional support for the consistency of the identified semantic orientations across the primary general-purpose platforms and the supplementary wine-focused platforms. Crucially, in the Chinese e-commerce context, sensory descriptors such as “Mouthfeel” should not be interpreted solely as hedonic tasting expressions, but also as indicators associated with product authenticity and consumption reliability. Together with concerns regarding logistics and counterfeit products, these terms collectively reflect a broader risk-sensitive evaluation orientation potentially associated with uncertainty avoidance tendencies (Zaman et al., 2023). This convergence suggests that the observed semantic orientations may reflect relatively consistent culturally embedded evaluation tendencies that extend beyond the characteristics of a single platform environment. Therefore, while platforms provide the digital context in which reviews are expressed, the recurring semantic patterns observed across datasets may also reflect broader culturally embedded evaluation tendencies (Filieri and Mariani, 2021). Because these nuanced cultural expressions are often obscured by surface-level quantitative metrics, this study reduces reliance on biased star ratings by capturing semantic-level emotional signals. The use of native language models further reduces potential distortion caused by skewed rating distributions, suggesting that this framework may have broader applicability for analyzing consumer reviews across different linguistic and platform contexts (Kusal et al., 2023; Poria et al., 2023).

This study provides an additional comparative approach for sentiment analysis in non-English online wine markets by establishing a transferable, comparative framework (Pocchiari et al., 2025). It demonstrates the disparities in consumer sentiment between Chinese and Italian platforms, illustrating how digital emotional expressions are jointly constrained by internal cultural factors and external platform mechanisms (Verma et al., 2023). To improve conceptual clarity, Table 8 provides a concise mapping between the main theoretical constructs, the observed semantic indicators, and the supporting NLP evidence. This mapping clarifies how the abstract constructs used in the discussion are operationally interpreted in this study, while recognizing that they should be understood as text-based indicative associations rather than directly measured psychological traits.

Table 8
Mapping of theoretical constructs, semantic indicators, and NLP evidence.

Theoretical construct	Observed semantic indicators	Supporting NLP evidence
Risk-sensitive evaluation orientation (associated with extrinsic motivation and uncertainty-related concerns)	Review attention is directed toward product reliability, authenticity, logistics, packaging, and consumption safety.	**Primary TF-IDF evidence:** JD reviews show salient terms such as “Mouthfeel”, “Logistics”, “Fake wine”, “Headache” and “Don't Buy”. **Supplementary TF-IDF evidence:** Jiuxian.com reviews highlight “Shipping fast”, “Packing” and “Quality wine”. **ABSA evidence:** Jiuxian.com reviews show high positive sentiment for functional aspects such as “Quality” (97.8% positive), “Packing” (98.5% positive) and “Shipping” (91.3% positive).
Sensory-descriptive evaluation orientation (associated with intrinsic motivation and sensory-aesthetic expression)	Review attention is directed toward sensory descriptions, tasting experience, social value, gifting, and aesthetic appreciation.	**Primary TF-IDF evidence:** Amazon.it reviews show salient terms such as “Gift”, “Excellent”, “Good”, “Taste”, “Packaging” and “Price”. **Supplementary TF-IDF evidence:** Vivino.it reviews highlight “Champagne”, “Citrus”, “Excellent”, “Fruity” and “Fresh”. **ABSA evidence:** Vivino.it reviews show high neutral proportions for sensory attributes such as “Aroma” (94.4% neutral) and “Acidity” (83.7% neutral), indicating a more descriptive sensory evaluation style.

5.2. *Theoretical and managerial implications*

To move beyond a context-specific comparison between China and Italy, this study develops a cross-cultural digital sentiment matrix (Fig. 9) as a transferable analytical reference framework for global marketing. This framework maps consumer behavior along two empirically derived dimensions: consumer motivation (functional/extrinsic versus hedonic/intrinsic) and sentiment expression pattern (restrained versus expressive), both identified through NLP-based sentiment and semantic analysis.

Based on these two dimensions, four analytically distinct review engagement profiles are proposed, positioned along a continuous behavioral spectrum rather than as isolated categories. The “Vocal Utilitarian”, located at the intersection of high expressiveness and functional orientation, is typically associated with strong platform incentives, resulting in high textual output but lower emotional authenticity. This configuration highlights the need for algorithmic filtering to distinguish genuine feedback from incentivized noise. At the opposite end of the expressive spectrum, the “Pragmatic Utilitarian” reflects a restrained yet functionally driven orientation, characteristic of risk-averse markets such as China, where consumers emphasize product authenticity and logistics reliability. This implies that trust-building mechanisms and operational transparency should be prioritized. Shifting to the hedonic dimension, the “Expressive Hedonist” occupies the quadrant of high expressiveness and intrinsic motivation, exemplified

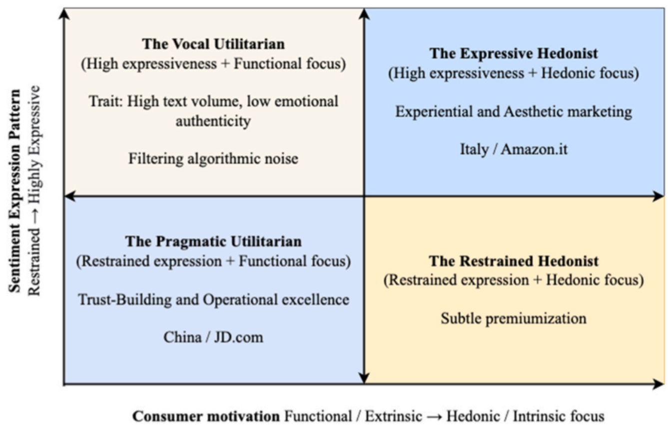

Fig. 9. Typological matrix of consumer review engagement and strategy.

by the Italian market, where reviews function as channels for self-expression and social interaction. Here, experiential marketing and user-generated content strategies become central. Completing the framework, the "Restrained Hedonist" represents a low expressiveness yet intrinsically motivated profile, in which consumers value product experience but exhibit limited public engagement, requiring subtle premiumization and personalized service strategies.

It is important to clarify that the proposed matrix currently represents a theoretically informed classification framework rather than a statistically validated typology model. Although its foundational dimensions are derived from NLP-based sentiment and semantic patterns, the resulting quadrants should be interpreted as heuristic categories for organizing cross-cultural review characteristics rather than as predictive classifications of individual consumers. This clarification helps avoid overextending the framework beyond the empirical evidence provided in this study. Rather than treating China and Italy as isolated cases, these markets serve as empirical anchors for illustrating a broader analytical logic. The proposed framework enables practitioners to map new markets onto a comparative analytical space by applying native-language NLP models to local review data. This provides a structured reference for supporting more data-informed localization strategies beyond static qualitative assessments.

From a managerial perspective, the study demonstrates how empirical NLP outputs can inform forward-looking strategic suggestions by linking specific sentiment keywords to concrete retail interventions. While the specific operational interventions discussed below, such as algorithmic optimization and sensory visualization interfaces, were not empirically tested in this study, they should be understood as data-informed strategic suggestions rather than validated managerial interventions. Specifically, as a forward-looking strategy, platform interfaces and recommendation systems can be dynamically optimized based on dominant semantic drivers. For instance, when addressing negative sentiments, the prominence of keywords such as "Fake wine" and "Headache" in the Chinese market suggests that platforms could proactively implement anti-counterfeit tracing systems and display prominent food safety certifications on product landing pages. Conversely, the concentration of negative Italian reviews around "Packaging" and "Price" points toward the potential value of targeted downstream fulfillment and operational interventions, focused on optimizing delivery materials and introducing dynamic promotional alerts. In addition, the framework supports predictive risk monitoring through keyword dynamics. By tracking shifts in positive and neutral lexical clusters, retailers can tailor their front-end marketing interfaces. When functional and sensory keywords such as "Mouthfeel" and "Sour" dominate consumer attention, managers could deploy detailed sensory visualization charts to precisely manage product expectations. When aesthetic and social terms including "Excellent" and "Gift" emerge as primary positive drivers, recommendation algorithms could prioritize social-sharing features, premium gift-wrapping options, and aesthetic presentations. Taken together, these findings suggest that sentiment analysis can move beyond a descriptive tool and provide a data-informed basis for aligning digital platform strategies with culturally embedded consumer behavior.

5.3. Limitations

Although this study provides useful insights into cross-cultural consumer sentiment in digital retail environments, several limitations define the boundaries of the current research design. First, relying primarily on textual review data excludes multimodal information such as images or visual product cues, which may critically shape consumer emotional responses in hedonic product categories. Second, while native-language Transformer-based models improve sentiment accuracy, the analytical pipeline remains constrained by semantic granularity. Specifically, the use of TF-IDF prioritizes interpretability over deeper contextual embedding, while regional linguistic variations and dialectal nuances may not be fully captured. Third, the empirical setting is limited to JD.com and Amazon.it, which, although representative, may not fully encompass the diversity of digital wine consumption behaviors, as platform algorithms may still mediate how culturally embedded preferences are expressed. While secondary datasets from wine-focused digital platforms were introduced for supplementary validation to mitigate this concern, the sample sizes for these validation sets remain relatively modest compared with the primary corpus. Therefore, this secondary validation primarily supports directional consistency rather than full-scale external validation. Moreover, Vivino. it contains both review-community and retail-related functions, meaning that this supplementary validation serves as evidence of semantic consistency rather than as a strict control for platform effects. Finally, because this study infers cultural orientations from textual sentiment patterns rather than directly measuring individual-level cultural values or actual purchasing behavior, the findings represent theoretically informed associations rather than direct causal claims.

5.4. Future research directions

Building on these limitations, several avenues for future research may further strengthen the explanatory power and applicability of this framework. First, to address the absence of visual inputs, future studies could adopt multimodal sentiment analysis approaches that integrate textual, visual, and audiovisual data to capture a more comprehensive range of consumer emotional expression. Second, to enhance semantic precision, future research could incorporate more advanced regionally fine-tuned ABSA and Transformer-based language models, thereby improving the ability to detect dialectal nuances and localized slang. Third, expanding the empirical scope to include multiple platforms, countries, and digital retail channels at a larger scale would help disentangle platform effects from culturally embedded consumer motivations and assess the broader applicability of the proposed framework. Finally, to strengthen causal inference, future research could combine text-based sentiment analysis with mixed-methods designs that integrate behavioral sales data (e.g., purchase conversion and repurchase behavior) and direct psychometric surveys. This would enable more precise validation of the relationships between linguistic patterns, consumer behavior, and underlying cultural orientations.

6. Conclusions

By employing native-language machine learning models (RoBERTa for Chinese and UmBERTo for Italian), this study conducted a cross-cultural sentiment analysis of online retail wine consumers on JD.com (China) and Amazon.it (Italy). Based on 33,464 online reviews, the findings revealed systematic differences in how consumers in these two markets express sentiment in digital retail environments. Chinese consumers placed greater emphasis on functional and risk-related attributes, such as logistics, mouthfeel, and fake wine, whereas Italian consumers focused more strongly on experiential, social, and aesthetic dimensions, including gifting, packaging, and price-related value. These findings suggest that online review behavior is indicative of culturally embedded motivational tendencies while also interacting with digital platform contexts. In addition, the results reveal a systematic divergence between platform-generated star ratings and underlying textual sentiment. Moreover, the results indicate that digital platforms do not fully determine consumer behavior, but may mediate how culturally embedded preferences are expressed in online reviews, a pattern this study conceptualizes as platform-mediated cultural expression.

The principal contribution of this study lies in integrating SDT and Hofstede's cultural dimensions into a native-language NLP framework, thereby linking abstract psychological and cultural constructs to observable sentiment patterns. By decoupling textual sentiment from platform-generated star ratings, this study suggests that native-language sentiment analysis can help capture the emotional structures embedded in consumer reviews. Building on these findings, this study develops a cross-cultural digital sentiment matrix as a transferable analytical framework for interpreting consumer behavior across heterogeneous digital markets. For practitioners, this framework provides a strategic basis for localized digital marketing, by translating specific NLP semantic drivers into concrete operational interventions, ranging from verifiable anti-counterfeit mechanisms in risk-sensitive markets to interactive sensory interfaces in hedonic contexts.

Future research may further strengthen this framework by incorporating multimodal sentiment analysis, regionally fine-tuned language models, and broader platform settings across multiple countries. In addition, integrating text-based sentiment analysis with behavioral sales data and direct psychometric measures would improve both causal interpretability and practical applicability. Ultimately, by bridging qualitative cultural interpretation with quantitative deep learning analytics, this study offers a scalable methodological pathway for understanding consumer behavior in global digital retail environments.

AI-driven sentiment analysis of grape wine consumers

A Comparative Study between China and Italy.

Funding

This work was supported by the Fundamental Research Funds for the University of Bologna and the Chinese Scholarship Council.

CRediT authorship contribution statement

Qiankun Liu: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Visualization, Writing – original draft. **Muhabaiti Pareti:** Formal analysis, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the Fundamental Research Funds for the University of Bologna and the Chinese Scholarship Council.

Appendix A. JD (China) product lists

(1) https://item.jd.com/100055004376.html (2) https://item.jd.com/100014769863.html (3) https://item.jd.com/100125453481.html (4) https://item.jd.com/1304924.html (5) https://item.jd.com/100050195286.html (6) https://item.jd.com/2236617.html (7) https://item.jd.com/100013853378.html (8) https://item.jd.com/25595595514.html (9) https://item.jd.com/100009858143.html (10) https://item.jd.com/6115649.html (11) https://item.jd.com/100018272508.html (12) https://item.jd.com/100024190510.html (13) https://item.jd.com/3188580.html (14) https://item.jd.com/10030370467232.html (15) https://item.jd.com/100251340406.html (16) https://item.jd.com/58438089801.html (17) https://item.jd.com/4984030.html (18) https://item.jd.com/100027834626.html (19) https://item.jd.com/10069609569.html (20) https://item.jd.com/100015396604.html (21) https://item.jd.com/100016660334.html (22) https://item.jd.com/100006024847.html (23) https://item.jd.com/100132491045.html (24) https://item.jd.com/100006759536.html (25) https://item.jd.com/100022875624.html (26) https://item.jd.com/100173987637.html (27) https://item.jd.com/100006517135.html (28) https://item.jd.com/1916616.html (29) https://item.jd.com/100063643756.html (30) https://item.jd.com/41351055086.html (31) https://item.jd.com/100043901827.html (32) https://item.jd.com/3188602.html (33) https://item.jd.com/3967652.html (34) https://item.jd.com/100139008826.html (35) https://item.jd.com/2830770.html (36) https://item.jd.com/2235838.html (37) https://item.jd.com/1590122.html (38) https://item.jd.com/3564062.html (39) https://item.jd.com/100045095303.html (40) https://item.jd.com/1145739624.html (41) https://item.jd.com/100051812046.html (42) https://item.jd.com/100045095279.html (43) https://item.jd.com/100069375696.html (44) https://item.jd.com/100147308782.html (45) https://item.jd.com/1182927.html (46) https://item.jd.com/10112136178812.html (47) https://item.jd.com/5811180.html (48) https://item.jd.com/100259240442.html (49) https://item.jd.com/1180763.html (50) https://item.jd.com/1271843.html (51) https://item.jd.com/100009640313.html (52) https://item.jd.com/100117702822.html (53) https://item.jd.com/100004032176.html (54) https://item.jd.com/10156628668984.html (55) https://item.jd.com/100125976962.html (56) https://item.jd.com/10024829634467.html (57) https://item.jd.com/10055725851076.html (58) https://item.jd.com/10057868999063.html (59) https://item.jd.com/100145844034.html (60) https://item.jd.com/100145135616.html (61) https://item.jd.com/100018641164.html (62) https://item.jd.com/10038500881607.html (63) https://item.jd.com/100018690383.html (64) https://item.jd.com/10104937851363.html (65) https://item.jd.com/2150242.html (66) https://item.jd.com/100021636711.html (67) https://item.jd.com/100167679363.html (68) https://item.jd.com/100012702275.html (69) https://item.jd.com/100030417216.html (70) https://item.jd.com/100017175796.html (71) https://item.jd.com/100048883652.html (72) https://item.jd.com/100104440878.html (73) https://item.jd.com/100177018485.html (74) https://item.jd.com/100109204049.html (75) https://item.jd.com/100191703232.html (76) https://item.jd.com/2939097.html (77) https://item.jd.com/100060703259.html (78) https://item.jd.com/10027723528905.html (79) https://item.jd.com/100228984354.html (80) https://item.jd.com/100109769443.html (81) https://item.jd.com/100043869479.html (82) https://item.jd.com/100151994127.html (83) https://item.jd.com/100070824034.html (84) https://item.jd.com/100189638068.html (85) https://item.jd.com/100008656123.html (86) https://item.jd.com/100016848279.html (87) https://item.jd.com/100167922016.html (88) https://npcitem.jd.hk/10184958311021.html (89) https://item.jd.com/100065089729.html (90) https://item.jd.com/100041152484.html (91) https://item.jd.com/100102256859.html (92) https://item.jd.com/100037942804.html (93) https://item.jd.com/100191677253.html (94) https://item.jd.com/100130692149.html (95) https://item.jd.com/100030417224.html (96) https://item.jd.com/10026615627204.html.

Appendix B. Amazon (Italy) product lists

(1) https://www.amazon.it/dp/B08691RZTJ; (2) https://www.amazon.it/dp/B005LABCOC (3) https://www.amazon.it/dp/B01GGBSQIE; (4) https://www.amazon.it/dp/B09BQVNZFD (5) https://www.amazon.it/dp/B08NV3MSD5; (6) https://www.amazon.it/dp/B08M728GZF (7) https://www.amazon.it/dp/B007X377IK; (8) https://www.amazon.it/dp/B005WPDFPK (9) https://www.amazon.it/dp/B08C3PMR2F; (10) https://www.amazon.it/dp/B09FXXQM8H (11) https://www.amazon.it/dp/B088CCDQTR; (12) https://www.amazon.it/dp/B017D1BLSI (13) https://www.amazon.it/dp/B00XUHGA1S; (14) https://www.amazon.it/dp/B088CCCPDK (15) https://www.amazon.it/dp/B008U7SWXC; (16) https://www.amazon.it/dp/B07THVC5JK (17) https://www.amazon.it/dp/B017T0YXG0; (18) https://www.amazon.it/dp/B004EAL5B0 (19) https://www.amazon.it/dp/B07P15HBLR; (20) https://www.amazon.it/dp/B09DZGM8F5 (21) https://www.amazon.it/dp/B0BLST2DRP; (22) https://www.amazon.it/dp/B08692WFWY (23) https://www.amazon.it/dp/B00DQR49QA; (24) https://www.amazon.it/dp/B007T29DA0 (25) https://www.amazon.it/dp/B088CD8M3H; (26) https://www.amazon.it/dp/B00I1XY0KO (27) https://www.amazon.it/dp/B08XJZDTGV; (28) https://www.amazon.it/dp/B0915VBKRX (29) https://www.amazon.it/dp/B086919WPW; (30) https://www.amazon.it/dp/B08H94QPG4 (31) https://www.amazon.it/dp/B07791JXM6; (32) https://www.amazon.it/dp/B085DPNWKD (33) https://www.amazon.it/dp/B071YLTCRX; (34) https://www.amazon.it/dp/B07DNR179D (35) https://www.amazon.it/dp/B01DUAAO4S; (36) https://www.amazon.it/dp/B08MMCS4WK (37) https://www.amazon.it/dp/B08J4N43JD; (38) https://www.amazon.it/dp/B09YRHP7P1 (39) https://www.amazon.it/dp/B0916R5Z5G; (40) https://www.amazon.it/dp/B09HQ6VF23 (41) https://www.amazon.it/dp/B00LJ7E6YO; (42) https://www.amazon.it/dp/B07LC2SN7X (43) https://www.amazon.it/dp/B08M739BM6; (44) https://www.amazon.it/dp/B004EAMAKA (45) https://www.amazon.it/dp/B017D1BN5O; (46) https://www.amazon.it/dp/B07252WF1X (47) https://www.amazon.it/dp/B001P4YY0C; (48) https://www.amazon.it/dp/B07ZRQSW6Q (49) https://www.amazon.it/dp/B003P3MR6E; (50) https://www.amazon.it/dp/B00P7DW5EE (51) https://www.amazon.it/dp/B085P1B29C; (52) https://www.amazon.it/dp/B016B1DM60 (53) https://www.amazon.it/dp/B086928V6S; (54) https://www.amazon.it/dp/B07DNRD59Z (55) https://www.amazon.it/dp/B07YSGLP7J; (56) https://www.amazon.it/dp/B01GI1HNAE (57) https://www.amazon.it/dp/B0BD7Z2Y6G; (58) https://www.amazon.it/dp/B00WAR19O2 (59) https://www.amazon.it/dp/B0BWFYRBT3; (60) https://www.amazon.it/dp/B0822WDLCK (61) https://www.amazon.it/dp/B06XH94171; (62) https://www.amazon.it/dp/B09Y9WLMPW (63) https://www.amazon.it/dp/B0BD2LWTVR; (64) https://www.amazon.it/dp/B01K1MDBW0 (65) https://www.amazon.it/dp/B01061V63W; (66) https://www.amazon.it/dp/B088CCBB6W (67) https://www.amazon.it/dp/B0876QL3P9; (68) https://www.amazon.it/dp/B004EAFFIE (69) https://www.amazon.it/dp/B088MC8RK4; (70) https://www.amazon.it/dp/B08C3PFXJC (71) https://www.amazon.it/dp/B005WPD9SI; (72) https://www.amazon.it/dp/B07DNRD125 (73) https://www.amazon.it/dp/B001TP8O4E; (74) https://www.amazon.it/dp/B0756BNP5G (75) https://www.amazon.it/dp/B0DYVJQXRF; (76) https://www.amazon.it/dp/B09N22C3CL (77) https://www.amazon.it/dp/B00XURT5CE; (78) https://www.amazon.it/dp/B0D1RLHS8F (79) https://www.amazon.it/dp/B08CB8X65Z; (80) https://www.amazon.it/dp/B072232VCY (81) https://www.amazon.it/dp/B0CKJ5HVJ3; (82) https://www.amazon.it/dp/B004ZX74PM (83) https://www.amazon.it/dp/B00FCA7BP0; (84) https://www.amazon.it/dp/B0872YMHGN (85) https://www.amazon.it/dp/B09F9T815X; (86) https://www.amazon.it/dp/B08XZJNWVS (87) https://www.amazon.it/dp/B07XYVWJLN; (88) https://www.amazon.it/dp/B09NDSJ9R2 (89) https://www.amazon.it/dp/B0BXT6MST1; (90) https://www.amazon.it/dp/B00XUGQQUY (91) https://www.amazon.it/dp/B009A5B7TY; (92) https://www.amazon.it/dp/B0BXY4JHBC (93) https://www.amazon.it/dp/B08CBGGL9J; (94) https://www.amazon.it/dp/B0CGV3MS57 (95) https://www.amazon.it/dp/B00AQF2UKM; (96) https://www.amazon.it/dp/B00NA02UC8 (97) https://www.amazon.it/dp/B00NP6PSE4; (98) https://www.amazon.it/dp/B086928QZ1 (99) https://www.amazon.it/dp/B08693G79F; (100) https://www.amazon.it/dp/B08CBDSZ1C (101) https://www.amazon.it/dp/B0BD82HQJR; (102) https://www.amazon.it/dp/B0CCDPJQK2 (103) https://www.amazon.it/dp/B007BV7FMW; (104) https://www.amazon.it/dp/B01GI1DPTC (105) https://www.amazon.it/dp/B096RDDN1P; (106) https://www.amazon.it/dp/B0BG2YP14W (107) https://www.amazon.it/dp/B00XUH2WE2; (108) https://www.amazon.it/dp/B07RYGL2B1 (109) https://www.amazon.it/dp/B0BG281G8Y; (110) https://www.amazon.it/dp/B09994CN9P; (111) https://www.amazon.it/dp/B002MX9W5C (112) https://www.amazon.it/dp/B0191BYUO0; (113) https://www.amazon.it/dp/B09SB148Q9 (114) https://www.amazon.it/dp/B01844N9Q0; (115) https://www.amazon.it/dp/B09HQ7H6S8 (116) https://www.amazon.it/dp/B00Y8BXARM; (117) https://www.amazon.it/dp/B071NDVFFN (118) https://www.amazon.it/dp/B086927JSR; (119) https://www.amazon.it/dp/B092J816Y4 (120) https://www.amazon.it/dp/B00XURNBJ2; (121) https://www.amazon.it/dp/B0972T4546 (122) https://www.amazon.it/dp/B07PP3JDTT; (123) https://www.amazon.it/dp/B081RMYXGF (124) https://www.amazon.it/dp/B0BDZLWC1Z; (125) https://www.amazon.it/dp/B0BDZKKP72 (126) https://www.amazon.it/dp/B00XURN7DM; (127) https://www.amazon.it/dp/B094Y6L3FQ (128) https://www.amazon.it/dp/B08RZ4HN3N; (129) https://www.amazon.it/dp/B07FKB6LR1 (130) https://www.amazon.it/dp/B004YXFA5Y; (131) https://www.amazon.it/dp/B004YXFA5Y (132) https://www.amazon.it/dp/B0C2CSH56W; (133) https://www.amazon.it/dp/B0CZPBKBH9 (134) https://www.amazon.it/dp/B08CBDQSB3; (135) https://www.amazon.it/dp/B01GI1G3H8 (136) https://www.amazon.it/dp/B07DMWPHT8; (137) https://www.amazon.it/dp/B0C5XK7C6B.

Appendix C. External validation data sources

Validation dataset 1: Chinese market (Jiuxian.com)

(1) https://www.jiuxian.com/goods-55671.html#answerArea (2) https://www.jiuxian.com/goods-1975504.html (3) https://www.jiuxian.com/goods-525847.html (4) https://www.jiuxian.com/goods-39751.html (5) https://www.jiuxian.com/goods-37739.html (6) https://www.jiuxian.com/goods-89345.html (7) https://www.jiuxian.com/goods-11564.html (8) https://www.jiuxian.com/goods-745520.html (9) https://www.jiuxian.com/goods-1903824.html (10) https://www.jiuxian.com/goods-27030.html.

Validation dataset 2: Italian market (Vivino.it)

(1) https://www.vivino.com/en/st-michael-eppan-sanct-valentin-chardonnay/w/2398590?year=2023&price_id=39916052 (2) https://www.vivino.com/it/pavonero-four-volte-25-rosso/w/1174493?price_id=34240849 (3) https://www.vivino.com/it/brigitte-delmotte-blanc-de-noirs-brut-champagne-champagne/w/8915271?price_id=34412403 (4) https://www.vivino.com/it/st-michael-eppan-sanct-valentin-sauvignon/w/1644452?year=2024&price_id=39410130 (5) https://www.vivino.com/it/tenuta-moraia-vermentino-maremma/w/7450105?year=2024&price_id=39957221 (6) https://www.vivino.com/en/la-volpe-one-uno-primitivo-di-manduria-riserva/w/6404545?year=2022&price_id=39844967 (7) https://www.vivino.com/en/marlborough-sun-sauvignon-blanc-marlborough/w/1624286?year=2025&price_id=41054199 (8) https://www.vivino.com/it/arthur-metz-cuvee-michel-leon-gewurztraminer/w/1823790?year=2024&price_id=39176560 (9) https://www.vivino.com/it/chapuy-l-espri

t-de-chapuy-extra-brut-blanc-de-blancs-champagne-grand-cru/w/1915384?price_id=34320352 (10) https://www.vivino.com/it/bastioni-della-rocca-appassimento/w/8528334?year=2025&price_id=41203056.

Data availability

Data will be made available on request.

References

Acheampong, F.A., Nunoo-Mensah, H., Chen, W., 2021. Transformer models for text-based emotion detection: a review of BERT-based approaches. Artif. Intell. Rev. 54 (8), 5789–5829. https://doi.org/10.1007/s10462-021-09958-2.

Agnoli, L., Outreville, J.F., 2021. Wine consumption and culture: a cross-country analysis. Appl. Econ. Perspect. Pol. 43 (3), 1101–1124. https://doi.org/10.1002/aepp.13097.

Ahani, A., et al., 2019. 'Market segmentation and travel choice prediction in spa hotels through TripAdvisor's online reviews'. Int. J. Hospit. Manag. 80, 52–77. https://doi.org/10.1016/j.ijhm.2019.01.003.

Alantari, H.J., et al., 2022. An empirical comparison of machine learning methods for text-based sentiment analysis of online consumer reviews. Int. J. Res. Market. 39 (1), 1–19. https://doi.org/10.1016/j.ijresmar.2021.10.011.

Antioco, M., Coussement, K., 2018. Misreading of consumer dissatisfaction in online product reviews: writing style as a cause for bias. Int. J. Inf. Manag. 38 (1), 301–310. https://doi.org/10.1016/j.ijinfomgt.2017.10.009.

Arnould, E.J., Thompson, C.J., 2005. Consumer Culture Theory (CCT): twenty years of research. J. Consum. Res. 31 (4), 868–882. https://doi.org/10.1086/426626.

Bagozzi, R.P., Gopinath, M., Nyer, P.U., 1999. The role of emotions in marketing. J. Acad. Market. Sci. 27 (2), 184–206. https://doi.org/10.1177/0092070399272005.

Bai, Y., Yu-Buck, G., 2025. Identifying targeted needs from online marketer- and user-generated data. J. Retailing Consum. Serv. 84, 104245. https://doi.org/10.1016/j.jretconser.2025.104245.

Baier, D., Karasenko, A., Rese, A., 2025. Measuring technology acceptance over time using transfer models based on online customer reviews. J. Retailing Consum. Serv. 85, 104278. https://doi.org/10.1016/j.jretconser.2025.104278.

Barbierato, E., Bernetti, I., Capecchi, I., 2021. Analyzing TripAdvisor reviews of wine tours: an approach based on text mining and sentiment analysis. Int. J. Wine Bus. Res. 34 (2), 212–236. https://doi.org/10.1108/IJWBR-04-2021-0025.

Basu, B., Sebastian, M.P., Kar, A.K., 2024. What affects the promoting intention of mobile banking services? Insights from mining consumer reviews. J. Retailing Consum. Serv. 77, 103695. https://doi.org/10.1016/j.jretconser.2023.103695.

Bazzani, C., et al., 2024. Exploring the effect of naturalness on consumer wine choices: evidence from a survey in Italy. Food Qual. Prefer. 113, 105062. https://doi.org/10.1016/j.foodqual.2023.105062.

Behera, R.K., et al., 2021. Co-LSTM: convolutional LSTM model for sentiment analysis in social big data. Inf. Process. Manag. 58 (1), 102435. https://doi.org/10.1016/j.ipm.2020.102435.

Birjali, M., Kasri, M., Beni-Hssane, A., 2021. A comprehensive survey on sentiment analysis: approaches, challenges and trends. Knowl. Base Syst. 226, 107134. https://doi.org/10.1016/j.knosys.2021.107134.

Brand, B.M., Reith, R., 2022. Cultural differences in the perception of credible online reviews – the influence of presentation format. Decis. Support Syst. 154, 113710. https://doi.org/10.1016/j.dss.2021.113710.

Buil, I., Catalán, S., Oliveira, T., 2024. Encouraging altruistic user-generated content in gamified review platforms. Internet Res. 36 (1), 263–291. https://doi.org/10.1108/INTR-02-2024-0211.

Croitoru, G., et al., 2024. A cross-cultural analysis of perceived value and customer loyalty in restaurants. Eur. Res. Manag. Bus. Econ. 30 (3), 100265. https://doi.org/10.1016/j.iedeen.2024.100265.

Dang-Van, T., Doan, L.T.M., Huynh, Q., 2026. Psychological pathways from fake reviews to consumer switching in e-tailing: a cross-cultural analysis. J. Retailing Consum. Serv. 90, 104700. https://doi.org/10.1016/j.jretconser.2025.104700.

Danner, L., et al., 2017. "I like the sound of that!" wine descriptions influence consumers' expectations, liking, emotions and willingness to pay for Australian white wines. Food Res. Int. 99, 263–274. https://doi.org/10.1016/j.foodres.2017.05.019.

Darraz, N., et al., 2025. Integrated sentiment analysis with BERT for enhanced hybrid recommendation systems. Expert Syst. Appl. 261, 125533. https://doi.org/10.1016/j.eswa.2024.125533.

Dashtipour, K., et al., 2016. Multilingual sentiment analysis: state of the art and independent comparison of techniques. Cogn. Comput. 8 (4), 757–771. https://doi.org/10.1007/s12559-016-9415-7.

Dong, Y., Fan, A., Ma, C., 2025. Leveraging digital sensory marketing in online wine sales – a moderated serial mediation model. Int. J. Contemp. Hospit. Manag. 37 (13), 234–251. https://doi.org/10.1108/IJCHM-04-2025-0534.

Duan, Y., Liu, T., Mao, Z., 2022. How online reviews and coupons affect sales and pricing: an empirical study based on e-commerce platform. J. Retailing Consum. Serv. 65, 102846. https://doi.org/10.1016/j.jretconser.2021.102846.

Filieri, R., Mariani, M., 2021. The role of cultural values in consumers' evaluation of online review helpfulness: a big data approach. Int. Mark. Rev. 38 (6), 1267–1288. https://doi.org/10.1108/IMR-07-2020-0172.

Garaus, M., Treiblmaier, H., 2021. The influence of blockchain-based food traceability on retailer choice: the mediating role of trust. Food Control 129, 108082. https://doi.org/10.1016/j.foodcont.2021.108082.

Gawlikowski, J., et al., 2023. A survey of uncertainty in deep neural networks. Artif. Intell. Rev. 56 (1), 1513–1589. https://doi.org/10.1007/s10462-023-10562-9.

Grabner-Kraeuter, S., 2002. The role of consumers' trust in online-shopping. J. Bus. Ethics 39 (1), 43–50. https://doi.org/10.1023/A:1016323815802.

Guo, Y., Barnes, S.J., Jia, Q., 2017. Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation. Tour. Manag. 59, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009.

Guo, Y., Chen, Y., Singh, S.K., 2025. Cross-border platforms' coupling of artificial intelligence adoption and strategic agility for upgraded ESMEs' global value chains: international marketing capability perspectives. Int. Mark. Rev. 42 (6), 1250–1283. https://doi.org/10.1108/IMR-10-2024-0418.

Gupta, G., Katarya, R., 2024. A computational approach towards food-wine recommendations. Expert Syst. Appl. 238, 121766. https://doi.org/10.1016/j.eswa.2023.121766.

Hallikainen, H., Laukkanen, T., 2018. National culture and consumer trust in e-commerce. Int. J. Inf. Manag. 38 (1), 97–106. https://doi.org/10.1016/j.ijinfomgt.2017.07.002.

Hu, N., et al., 2012. Manipulation of online reviews: an analysis of ratings, readability, and sentiments. Decis. Support Syst. 52 (3), 674–684. https://doi.org/10.1016/j.dss.2011.11.002.

Kendall, H., et al., 2019. Chinese consumer's attitudes, perceptions and behavioural responses towards food fraud. Food Control 95, 339–351. https://doi.org/10.1016/j.foodcont.2018.08.006.

Kim, J., Giroux, M., Lee, J.C., 2021. When do you trust AI? The effect of number presentation detail on consumer trust and acceptance of AI recommendations. Psychol. Market. 38 (7), 1140–1155. https://doi.org/10.1002/mar.21498.

Kim, J.M., Jun, M., Kim, C.K., 2018. The effects of culture on consumers' consumption and generation of online reviews. J. Interact. Market. 43, 134–150. https://doi.org/10.1016/j.intmar.2018.05.002.

Kim, Y., Srite, M., Zhao, H., 2023. Trust antecedents in online reviews across national cultures. Decis. Support Syst. 173, 113998. https://doi.org/10.1016/j.dss.2023.113998.

Kumar, N., et al., 2018. Detecting review manipulation on online platforms with hierarchical supervised learning. J. Manag. Inf. Syst. 35 (1), 350–380. https://doi.org/10.1080/07421222.2018.1440758.

Kusal, S., et al., 2023. A systematic review of applications of natural language processing and future challenges with special emphasis in text-based emotion detection. Artif. Intell. Rev. 56 (12), 15129–15215. https://doi.org/10.1007/s10462-023-10509-0.

Ladwein, R., Sánchez Romero, A.M., 2021. The role of trust in the relationship between consumers, producers and retailers of organic food: a sector-based approach. J. Retailing Consum. Serv. 60, 102508. https://doi.org/10.1016/j.jretconser.2021.102508.

Leung, K., Cho, V., 2024. Motivation for writing long online reviews: a big data analysis of an anime community. Internet Res. 34 (5), 1845–1871. https://doi.org/10.1108/INTR-07-2022-0548.

Liang, W.-Y., et al., 2025. The impact of mandatory disclosure on rewarding online reviews based on S-O-R theory. Asia Pac. J. Mark. Logist. 37 (8), 2500–2516. https://doi.org/10.1108/APJML-08-2024-1131.

Liu, X., et al., 2024a. eWOM information richness and online user review behavior: evidence from TripAdvisor. J. Theor. Appl. Electron. Commer. Res. 19 (2), 880–898. https://doi.org/10.3390/jtaer19020046.

Liu, Y., et al., 2024b. Unveiling consumer preferences in automotive reviews through aspect-based opinion generation. J. Retailing Consum. Serv. 77, 103605. https://doi.org/10.1016/j.jretconser.2023.103605.

Lo, S.L., et al., 2017. Multilingual sentiment analysis: from formal to informal and scarce resource languages. Artif. Intell. Rev. 48 (4), 499–527. https://doi.org/10.1007/s10462-016-9508-4.

Luca, M., Zervas, G., 2016. Fake it till you make it: reputation, competition, and yelp review fraud. Manag. Sci. 62 (12), 3412–3427. https://doi.org/10.1287/mnsc.2015.2304.

Moradi, M., Dass, M., Kumar, P., 2023. Differential effects of analytical versus emotional rhetorical style on review helpfulness. J. Bus. Res. 154, 113361. https://doi.org/10.1016/j.jbusres.2022.113361.

Morimura, F., Sakagawa, Y., 2023. The intermediating role of big data analytics capability between responsive and proactive market orientations and firm performance in the retail industry. J. Retailing Consum. Serv. 71, 103193. https://doi.org/10.1016/j.jretconser.2022.103193.

Moscovici, D., et al., 2022. Consumer preferences for organic wine - global analysis of people and place. J. Clean. Prod. 368, 133215. https://doi.org/10.1016/j.jclepro.2022.133215.

Mosikyan, S., et al., 2024. A systematic literature review and future research agenda to study consumer acceptance of novel foods and beverages. Appetite 203, 107655. https://doi.org/10.1016/j.appet.2024.107655.

Nilashi, M., et al., 2022. Factors impacting customer purchase intention of smart home security systems: social data analysis using machine learning techniques. Technol. Soc. 71, 102118. https://doi.org/10.1016/j.techsoc.2022.102118.

Otter, D.W., Medina, J.R., Kalita, J.K., 2021. A survey of the usages of deep learning for natural language processing. IEEE Transact. Neural Networks Learn. Syst. 32 (2), 604–624. https://doi.org/10.1109/TNNLS.2020.2979670.

Přibáň, P., et al., 2024. A comparative study of cross-lingual sentiment analysis. Expert Syst. Appl. 247, 123247. https://doi.org/10.1016/j.eswa.2024.123247.

Plotkina, D., Munzel, A., Pallud, J., 2020. Illusions of Truth—experimental insights into human and algorithmic detections of fake online reviews. J. Bus. Res. 109, 511–523. https://doi.org/10.1016/j.jbusres.2018.12.009.

Pocchiari, M., Proserpio, D., Dover, Y., 2025. Online reviews: a literature review and roadmap for future research. Int. J. Res. Market. 42 (2), 275–297. https://doi.org/10.1016/j.ijresmar.2024.08.009.

Poria, S., et al., 2023. Beneath the tip of the iceberg: current challenges and new directions in sentiment analysis research. IEEE Trans. Affective Comput. 14 (1), 108–132. https://doi.org/10.1109/TAFFC.2020.3038167.

Praveen, S.V., et al., 2024. Crafting clarity: leveraging large language models to decode consumer reviews. J. Retailing Consum. Serv. 81, 103975. https://doi.org/10.1016/j.jretconser.2024.103975.

Qiu, X., et al., 2020. Pre-trained models for natural language processing: a survey. Sci. China Technol. Sci. 63 (10), 1872–1897. https://doi.org/10.1007/s11431-020-1647-3.

Ray, R.K., Singh, A., 2025. From online reviews to smartwatch recommendation: an integrated aspect-based sentiment analysis framework. J. Retailing Consum. Serv. 82, 104059. https://doi.org/10.1016/j.jretconser.2024.104059.

Román, S., Riquelme, I.P., Iacobucci, D., 2024. Antecedents and consequences of perceived helpfulness of extremely positive and exaggerated reviews. J. Retailing Consum. Serv. 80, 103907. https://doi.org/10.1016/j.jretconser.2024.103907.

Rui, M., et al., 2025. Exploring consumer sentiments and opinions in wine E-commerce: a cross-country comparative study. J. Retailing Consum. Serv. 82, 104097. https://doi.org/10.1016/j.jretconser.2024.104097.

Ryan, R.M., Deci, E.L., 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am. Psychol. 55 (1), 68–78. https://doi.org/10.1037//0003-066x.55.1.68.

Salehan, M., Kim, D.J., 2016. Predicting the performance of online consumer reviews: a sentiment mining approach to big data analytics. Decis. Support Syst. 81, 30–40. https://doi.org/10.1016/j.dss.2015.10.006.

Salton, G., Buckley, C., 1988. Term-weighting approaches in automatic text retrieval. Inf. Process. Manag. 24 (5), 513–523. https://doi.org/10.1016/0306-4573(88)90021-0.

Schoenmueller, V., Netzer, O., Stahl, F., 2020. The polarity of online reviews: prevalence, drivers and implications. J. Mark. Res. 57 (5), 853–877. https://doi.org/10.1177/0022243720941832.

Schreurs, M., et al., 2024. Predicting and improving complex beer flavor through machine learning. Nat. Commun. 15 (1), 2368. https://doi.org/10.1038/s41467-024-46346-0.

Sáenz-Navajas, M.-P., et al., 2013. Sensory drivers of intrinsic quality of red wines: effect of culture and level of expertise. Food Res. Int. 54 (2), 1506–1518. https://doi.org/10.1016/j.foodres.2013.09.048.

Sinesio, F., et al., 2021. Influence of wine traits and context on liking, intention to consume, wine-evoked emotions and perceived sensory sensations. Food Qual. Prefer. 93, 104268. https://doi.org/10.1016/j.foodqual.2021.104268.

Spence, C., 2020. Wine psychology: Basic & applied. Cognit. Res.: Princip. Implications 5 (1), 22. https://doi.org/10.1186/s41235-020-00225-6.

Ullah, F., et al., 2025. Prompt-based fine-tuning with multilingual transformers for language-independent sentiment analysis. Sci. Rep. 15 (1), 20834. https://doi.org/10.1038/s41598-025-03559-7.

Verma, D., et al., 2023. A meta-analysis of antecedents and consequences of eWOM credibility: investigation of moderating role of culture and platform type. J. Bus. Res. 154, 113292. https://doi.org/10.1016/j.jbusres.2022.08.056.

Wang, L., Chen, L., Li, C., 2024. Research on strategies for improving green product consumption sentiment from the perspective of big data. J. Retailing Consum. Serv. 79, 103802. https://doi.org/10.1016/j.jretconser.2024.103802.

Wang, Q., et al., 2022. Effect of online review sentiment on product sales: the moderating role of review credibility perception. Comput. Hum. Behav. 133, 107272. https://doi.org/10.1016/j.chb.2022.107272.

Wistedt, Ugné, 2024. Consumer purchase intention toward POI-retailers in cross-border E-commerce: An integration of technology acceptance model and commitment-trust theory. Journal of Retailing and Consumer Services 81, 104015. https://doi.org/10.1016/j.jretconser.2024.104015, 104015. https://linkinghub.elsevier.com/retrieve/pii/S0969698924003114.

Xiang, Z., et al., 2015. What can big data and text analytics tell us about hotel guest experience and satisfaction? Int. J. Hospit. Manag. 44, 120–130. https://doi.org/10.1016/j.ijhm.2014.10.013.

Xiao, Y., et al., 2022. User preference mining based on fine-grained sentiment analysis. J. Retailing Consum. Serv. 68, 103013. https://doi.org/10.1016/j.jretconser.2022.103013.

Yadav, A., Vishwakarma, D.K., 2020. Sentiment analysis using deep learning architectures: a review. Artif. Intell. Rev. 53 (6), 4335–4385. https://doi.org/10.1007/s10462-019-09794-5.

Yang, S., et al., 2024. Deep learning mechanism and big data in hospitality and tourism: developing personalized restaurant recommendation model to customer decision-making. Int. J. Hospit. Manag. 121, 103803. https://doi.org/10.1016/j.ijhm.2024.103803.

Zaman, M., et al., 2023. Motives for posting fake reviews: evidence from a cross-cultural comparison. J. Bus. Res. 154, 113359. https://doi.org/10.1016/j.jbusres.2022.113359.

Zhang, D., et al., 2016. What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews. J. Manag. Inf. Syst. 33 (2), 456–481. https://doi.org/10.1080/07421222.2016.1205907.

Zhao, H., et al., 2021. A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. Inf. Process. Manag. 58 (5), 102656. https://doi.org/10.1016/j.ipm.2021.102656.

Zhao, Y., et al., 2025. AI vs. human: a large-scale analysis of AI-generated fake reviews, human-generated fake reviews and authentic reviews. J. Retailing Consum. Serv. 87, 104400. https://doi.org/10.1016/j.jretconser.2025.104400.

Zhao, X., Huang, Z., 2024. A method for exploring consumer satisfaction factors using online reviews: a study on anti-cold drugs. J. Retailing Consum. Serv. 81, 103895. https://doi.org/10.1016/j.jretconser.2024.103895.