



HAL
open science

Detection of Branded Posts in User-Generated Content

Clement Levallois

► **To cite this version:**

Clement Levallois. Detection of Branded Posts in User-Generated Content. International Conference on Human-Computer Interaction, Jul 2024, Washington DC, United States. pp.304-312, 10.1007/978-3-031-61305-0_21 . hal-04638619

HAL Id: hal-04638619

<https://hal.science/hal-04638619v1>

Submitted on 8 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Detection of Branded Posts in User-Generated Content

Clement Levallois¹[0000-0002-2482-1319]

emlyon business school, 15 boulevard Diderot, 75012 Paris, France
levalllois@em-lyon.com

Abstract. User-generated content (UGC) is a fundamental source of information for the study of consumer behavior, product development, and to assess the quality of service. The expansion of branded content, published and mixed with “ordinary” UGC on the same online platforms, blurs the notions of which content should be considered for these studies. This contribution draws on the notion of “authenticity” to offer a taxonomy distinguishing “branded” from “organic” content and presents a computational method to detect branded content in UGC.

Keywords: user-generated content, branded content, text mining

1 Introduction: Defining UGC and “Authenticity”

“User - generated content” (UGC) is a broad denomination covering the textual and visual elements produced by individuals and published online, typically on social media in the form of private and public conversations, posts, comments and on e-commerce platforms in the form of reviews, ratings and comments. UGC is often perceived to be a distinctively “authentic” form of content when compared to alternative sources of information, which explains its strong appeal to consumers to inform their purchase decisions and to researchers investigating consumer sentiment. In this contribution, we take a second look at the authenticity of UGC, by offering a matrix crossing six components of authenticity with different types of UGC – including branded content posted by influencers. Based on this taxonomy, we then present a method to improve the detection and removal of branded UGC from a corpus.

UGC has been deemed valuable as a source for text analysis in market research [4, 14] as it can provide near real-time, high volume customer sentiment and customer feedback on their experiences at reduced cost, which classic sources for market research such as focus group, poll or surveys struggle to achieve. Product reviews have served identifying customer needs [38] and the product features which can meet these needs [7, 9]. Reviews can help assess the quality of service, for instance in the hospitality industry [6, 25, 30].

A (sometimes implicit) assumption made when performing analysis on UGC is that, besides the convenience to collect it, it derives a special value from a series of qualities, which can be summed up as “authenticity”. “Authenticity” of UGC would stem from the fact that it is contributed in an unsolicited manner (as opposed to opinions shared in response to polls or surveys), in natural settings (when contrasted with

the experimental environment of a focus group), and independently from editorial supervision and professional hierarchies (contrary to the content produced by traditional news media). As such, UGC is considered a trustworthy reflection of consumer knowledge and sentiment [17, 37, 39]. When UGC is analyzed in market research with text mining, which “can reveal patterns of attention or focus of which the speaker may not be conscious” [14], the assumption is reinforced that UGC would grant an unmediated access to the candid, “authentic” state of mind of the content producer.

The assumption of the unqualified authenticity of UGC can however be easily questioned. Is the post by an influencer promoting a product as “authentic” as the post by an ordinary, amateur individual discussing their recent experience of the same product? Disentangling the issue requires first to unpack the multifaceted notion of “authenticity”. The concept is usefully decomposed from the perspective of the consumer’s experience by Nunes et al. [28], through their qualitative fieldwork examining data collected directly from consumers. Their study validates a definition of authenticity as a “holistic consumer assessment determined by six component judgments (accuracy, connectedness, integrity, legitimacy, originality, and proficiency)” (see **Table 1**).

Table 1. Components of authenticity, from Nunes et al. [28]. Used with permission.

Table 2. Components of Authenticity.	
Component	Definition
1. Accuracy	The extent to which a provider is perceived as transparent in how it represents itself and its products and/or services and, thus, reliable in terms of what it conveys to customers.
2. Connectedness	The extent to which a customer feels engaged, familiar with, and sometimes even transformed by a source and/or its offering.
3. Integrity	The extent to which a provider is perceived as being intrinsically motivated, not acting out of its own financial interest, while acting autonomously and consistently over time.
4. Legitimacy	The extent to which a product or service adheres to shared norms, standards, rules, or traditions present in the market.
5. Originality	The extent to which a product or service stands out from mainstream offerings present in the market and does so without unnecessary embellishments.
6. Proficiency	The extent to which a provider is perceived as properly skilled, exhibiting craftsmanship and/or expertise.

A second step to comprehend better how UGC relates to the notion of authenticity requires to trace a clear and meaningful delineation between the different types of motivations driving the creation of otherwise similar content. In the following, we will designate the content created by authors who post without an ulterior motive as “organic UGC”: authors are intrinsically motivated, and the content is not created for (typically) monetary gain.

In contrast, “influencers” will be defined as authors who publish content for extrinsic motives, deriving typically from a contractual commercial partnership with a brand [34] or from the participation to a referral program [24]. A common extrinsic motive is monetary or material gain, which rejoins established definitions of “branded content” in the industry: content which is at least partially funded and created for marketing purposes [2, 11]. Hence, we will define influencers as producers of branded content.

2 A Taxonomy of UGC and Their Authenticity

With these conceptual and terminological clarifications at hand, we can draw a taxonomy of different types of UGC according to the six components of authenticity identified by Nunes et al. [28]. The taxonomy includes “organic UGC” and “branded content” as defined above with three other types of content (user generated or not) which are typically also posted and mixed with UGC on the same online platforms: advertisements, spam and fake reviews [15, 19, 29] (Table 2).

Table 2. Taxonomy of different types of UGC (plus advertising) and their relative strengths on the components constituting “authenticity” (as defined by Nunes et al. [28]), from the point of the view of the consumer’s experience.

	licit types of content			illicit or illegitimate types of content	
	“organic UGC”	Branded content	Advertisement	Spam	Fake reviews
Accuracy	Medium to strong	Weak to medium	Medium	Weak	Weak
Connectedness	Medium to strong	Medium to strong	Highly dependent on the campaign	Weak	Weak
Integrity	Strong	Weak to medium	Weak to medium	Weak	Weak
Legitimacy	Weak to medium	Medium	Strong	Weak	Weak
Originality	Highly dependent on the individual	Highly dependent on the influencer	Highly dependent on the campaign	Weak	Weak
Proficiency	Highly dependent on the individual	Medium to strong	Strong	Weak	Weak

The table illustrates the contrasted scoring of different kinds of UGC on the six components forming authenticity. Spam and fake reviews stand out as uniformly weak on all components. The remaining types follow different logics. Organic UGC scores the highest on integrity and accuracy, but the lowest on legitimacy. Advertisements score the highest on legitimacy and proficiency, but the lowest on integrity.

A perspective per type of UGC provides an interesting view on a possible reason why the market for branded content and influencers is growing exponentially [11]: branded content scores high on connectedness and proficiency, all while remaining relatively legitimate (more than an advertisement) as they produce content in a more established, standard style than the ordinary, amateur user posting on the same topic.

3 The Case for Isolating Branded Content from UGC for the Study of Authentic Consumer Sentiment

Are all types of UGC to be considered for all types of inquiries? Studies concerned with the magnitude and spread of (dis)information typically include all forms of UGC – including spam and fake reviews - to feed into their analysis [10, 26, 39]. In contrast, studies investigating consumer sentiment would be tainted by the inclusion of spam and fake reviews. Marketers and researchers are reminded that they “should vigorously identify and isolate these predatory comments from the analysis” [31]. A large research stream has detection methods for the removal of spam and fake reviews [15, 16, 29, 35]¹.

For the purpose of investigating the *formation* of consumer sentiment, branded content and advertisements must naturally be included in the investigations mapping the information flowing along the ties connecting taste makers and susceptible members in online social networks [1], contributing to identifying the dynamics of information sharing and the determinants of virality [34, 36].

The characterization of *the state* of consumer sentiment would presumably require a different approach. While consumer sentiment is indeed shaped by the variety of online contents it coexists with – branded content, ads, spams and fake reviews – it arguably relates to a distinct subset of UGC, which we named “organic content”: the content posted by ordinary, intrinsically motivated individuals sharing their thoughts transparently. This subset matches most of the components of the definition of authenticity offered by Nunes et al. [28], which is coherent with the notion that consumer sentiment would reflect the “authentic” expression of consumers.

Detecting branded content would allow for more accurate consumer sentiment analysis, centered on organic UGC, reflecting consumer’s thoughts with higher fidelity. This would benefit market research tasks performed on UGC which hinge on the premise that UGC is the “authentic” reflection of consumer expression. In the following section, we offer a preliminary version of a method for the identification of organic and branded content.

4 Detection of Branded Posts

We design a method for detection which follows a rule-based approach derived from Umigon, which is a model for sentiment analysis developed by the author [21-22]. The model can be decomposed in:

1. Pre-processing of the UGC: tokenization, detection of phrases, ngrams, removal of stopwords.
2. for each phrase, each n-gram is compared with entries from a lexicon.

¹ In practice, the step of spam removal is unfrequently performed (and if so, not documented in detail) in marketing research papers on online consumer sentiment.

3. in case of a match, zero or several Boolean conditions attached to the lexicon entry are evaluated and return a result: branded content or not.

4.1 Pre-processing

Tokenization. Tokenization is performed with a custom parser². The parser splits the text at white spaces and identifies a series of categories of tokens: words, emojis, emoticons, punctuation signs, “non words” (onomatopes, abbreviations) and white spaces. Indices of each token in the text are recorded.

Phrases. A phrase is defined as a relatively independent proposition in a text. Periods, commas, exclamation and question marks all delineate phrases within a text. Parentheses and quotes also play this role in a more elaborated way (with a logic of opening and closing characters). Identifying phrases can be useful to ignore irrelevant fragments (such as content in parentheses), and to identify tokens at positions with special significance in the text – such as a word at the beginning of a phrase, or a punctuation sign at the end of it.

Ngrams. Ngrams are consecutive sequence of terms. We use a custom implementation which leverages the tokenization and sentence fragment detection presented above: ngrams are detected within sentence fragments and are guaranteed not to include punctuation signs, emojis, emoticons or non words. Ngrams are also filtered to remove less frequent ones already included in lengthier ngrams (the trigram “United States of” will be removed if the quadrigram “United States of America” appears more frequently).

Stopword Removal. Stopwords are removed, only after processing ngrams so that stopwords embedded in ngrams will not be removed: the stopword “of” is removed but not when embedded in “United States of America”.

4.2 Matching with Lexicon Entries

Each ngram of the text is compared with entries of a lexicon. The lexicon is a list of ngrams which are frequently found in branded content. This list is curated by the author and expands steadily (with 104 entries as of February 2024)³. Entries are selected by culling “manually” social media posts. UGC which appears to be a form of branded content is scrutinized for typical semantic features of branded content:

² The source code of all steps of the method is available under an Creative Commons Attribution 4.0 International Public License license at <https://github.com/seinecle/umigon-family>

³ https://github.com/seinecle/umigon-static-files/blob/main/src/main/resources/net/clementvallois/umigon/lexicons/en/9_commercial%20tone.txt

- calls to action. Organic UGC rarely includes calls to action while branded content often does.
- addresses to an audience, from the perspective of an unspecified “we” (“we offer...”, “you will appreciate”).
- use of vocabulary typical of a corporate voice. Organic UGC rarely uses phrases like “complimentary”, “for a chance to”, “stoked to”, “boasts” and similar expressions, which are characteristic of an intent to promote or embellish a product or service.

4.3 Evaluating Boolean Conditions

Lexicon entries alone are frequently insufficiently specific to afford a precise characterization of the content. Each entry can be supplemented by several Boolean conditions which help capture the context of use of the lexicon entry. For instance, the expression “find out” is not a specific marker of branded content in itself, however when used as an imperative at the beginning of a phrase it evokes a clear call to action. The condition `isStartOfSegment` attached to the lexicon entry allows to assess this element of context. 40 Boolean conditions have been designed to assist in qualifying lexicon entries⁴.

5 Evaluation

Four brands have been selected to cover a variety of products and services from a hedonist, utilitarian, B2B and B2C contexts: MongoDB (database), Club Med (French travel and tourism operator), Veuve Cliquot (Champagne house) and HP printers. A search for 200 tweets has been performed in February 2024 for each of them, using the `twscrape` library⁵. A number of tweets were removed as the search term was included in the twitter handle, not in the content of the post. The dataset was annotated by the author for organic vs branded tone. The coding procedure followed these heuristics:

- a post with a call to action, or promoting a product or an event, or addressing an impersonal audience (“you...”) from an impersonal standpoint (“we...”), will tend to be annotated as “branded”.
- other posts will tend to be annotated as “organic”

The tweets were analyzed with the method presented in this paper and implemented in the platform Nocode Functions (<https://nocodefunctions.com>).

⁴ <https://github.com/seinecle/umigon-lexicons/tree/main/src/main/java/net/clementvallois/umigon/heuristics/booleanconditions>

⁵ <https://github.com/vladkens/twscrape>. The search on MongoDB returned 218 tweets and the search on HP printer returned 206 results, despite the parameter set.

Table 3. Evaluation of the method.

	n	True positives	False positives	False negatives	F1
MongoDB	218	13	1	55	0.32
Club Med	132	12	0	12	0.67
Veuve Cliquot	181	5	2	31	0.23
HP printer	205	4	0	51	0.14

F1 scores are noticeably low. The method offers however promising perspectives, for three reasons.

The method is inherently interpretable, as the internals of rule-based systems can in principle be decomposed to reveal the chain of causal processes which led to the classification outcome [32]. The system presented here can effectively be used to share with the analyst a full report of the rules which were followed to arrive at the classification of each piece of text⁶.

The rate of false positive is close to zero: the method is very precise. One of the strengths of lexicon-based methods, compared to models based on statistical learning, is that expert-crafted rules can be designed with great precision, avoiding the inherent element of randomness introduced by a supervised learning approach. The adjunction of Boolean conditions to further characterize lexicon entries limits even further the rate of false positives, as confounding elements of context can be evaluated and considered.

Finally, the recall is very low, due to the very weak coverage of the lexicons. A low number of lexicon entries implies that a large number of expressions typically characterizing “branded content” are not currently captured by the lexicon. This is sometimes described as a limiting factor of all lexicon-based methods [12]. However, such methods have proven to be able to achieve high levels of recall in a similar categorization task (sentiment analysis), leading to F1 scores competitive with large language models [23]⁷. The resource needed to achieve such levels of recall, and overall precision, is a continuous expansion of the lexicons to gradually expand their coverage. This traces an encouraging roadmap for future work.

⁶ Single tests can be performed on the homepage of <https://nocodefunctions.com>. Text files can be analyzed on the same platform, returning explanations for each of the results in a spreadsheet format. An API access is also available.

⁷ See also the public benchmark: <https://github.com/seinecle/umibench>

References

1. Aral, S., & Walker, D. (2012). Identifying Influential and Susceptible Members of Social Networks. *Science*, 337(6092), 337–341. <https://doi.org/10.1126/science.1215842>
2. Asmussen, B., Wider, S., Williams, R., Stevenson, N., & Whitehead, E. (2016). Defining Branded Content for the Digital Age: The Industry Experts' Views on Branded Content as a New Marketing Communications Concept (A Collaborative Research Project Commissioned by the Branded Content Marketing Association (BCMA), p. 42). Branded Content Marketing Association. <https://research.cbs.dk/en/publications/defining-branded-content-for-the-digital-age-the-industry-experts>
3. BalaAnand, M., Karthikeyan, N., Karthik, S., Varatharajan, R., Manogaran, G., & Sivaparthipan, C. B. (2019). An enhanced graph-based semi-supervised learning algorithm to detect fake users on Twitter. *The Journal of Supercomputing*, 75(9), 6085–6105. <https://doi.org/10.1007/s11227-019-02948-w>
4. Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the Tribes: Using Text for Marketing Insight. *Journal of Marketing*, 84(1), 1–25. <https://doi.org/10.1177/0022242919873106>
5. Calder, B. J. (1977). Focus Groups and the Nature of Qualitative Marketing Research. *Journal of Marketing Research*, 14(3), 353–364. <https://doi.org/10.1177/002224377701400311>
6. Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment Classification of Consumer-Generated Online Reviews Using Topic Modeling. *Journal of Hospitality Marketing & Management*, 26(7), 675–693. <https://doi.org/10.1080/19368623.2017.1310075>
7. Decker, R., & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*, 27(4), 293–307. <https://doi.org/10.1016/j.ijresmar.2010.09.001>
8. Griffin, A., & Hauser, J. R. (1993). The Voice of the Customer. *Marketing Science*, 12(1), 1–27. <https://doi.org/10.1287/mksc.12.1.1>
9. Haddara, M., Hsieh, J., Fagerstrøm, A., Eriksson, N., & Sigurdsson, V. (2020). Exploring customer online reviews for new product development: The case of identifying reinforcers in the cosmetic industry. *Managerial & Decision Economics*, 41(2), 250–273. <https://doi.org/10.1002/mde.3078>
10. Hajli, N., Saeed, U., Tajvidi, M., & Shirazi, F. (2022). Social Bots and the Spread of Disinformation in Social Media: The Challenges of Artificial Intelligence. *British Journal of Management*, 33(3), 1238–1253. <https://doi.org/10.1111/1467-8551.12554>
11. Hardy, J., Karagiorgou, I., Keddo, N., Moise, R., Sujon, Z., & Yesiloglu, S. (2023). *The UK Branded Content Industry: Report and Survey* (p. 97). Branded Content Research Hub.
12. Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, 36(1), 20–38. <https://doi.org/10.1016/j.ijresmar.2018.09.009>
13. Heinonen, K. (2011). Consumer activity in social media: Managerial approaches to consumers' social media behavior. *Journal of Consumer Behaviour*, 10(6), 356–364. <https://doi.org/10.1002/cb.376>
14. Humphreys, A., & Wang, R. J.-H. (2018). Automated Text Analysis for Consumer Research. *Journal of Consumer Research*, 44(6), 1274–1306. <https://doi.org/10.1093/jcr/ucx104>
15. Jindal, N., & Liu, B. (2007). Analyzing and Detecting Review Spam. *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, 547–552. <https://doi.org/10.1109/ICDM.2007.68>

16. Kaddoura, S., Chandrasekaran, G., Elena Popescu, D., & Duraisamy, J. H. (2022). A systematic literature review on spam content detection and classification. *PeerJ Computer Science*, 8, e830. <https://doi.org/10.7717/peerj-cs.830>
17. Kannan, P. K., & Li, H. (2017). Digital marketing: A framework, review and research agenda. *International Journal of Research in Marketing*, 34(1), 22–45. <https://doi.org/10.1016/j.ijresmar.2016.11.006>
18. Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H. (2020). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*, 90, 523–537. <https://doi.org/10.1016/j.indmarman.2019.08.003>
19. Lappas, T., Sabnis, G., & Valkanas, G. (2016). The Impact of Fake Reviews on Online Visibility: A Vulnerability Assessment of the Hotel Industry. *Information Systems Research*, 27(4), 940–961. <https://doi.org/10.1287/isre.2016.0674>
20. Lee, T. Y., & Bradlow, E. T. (2011). Automated Marketing Research Using Online Customer Reviews. *Journal of Marketing Research*, 48(5), 881–894. <https://doi.org/10.1509/jmkr.48.5.881>
21. Levallois, C. (2013). Umigon: Sentiment analysis for tweets based on terms lists and heuristics. *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, 414–417. <https://aclanthology.org/S13-2068>
22. Levallois, C. (2023). *Reintroducing qualitative insights in big data: The case of “sentiment” in textual analysis. Under review.*
23. Levallois, C. (n.d.). Umigon-lexicon: A contribution to inherently interpretable sentiment analysis. *Language Resources and Evaluation* (forthcoming)
24. Lobel, I., Sadler, E., & Varshney, L. R. (2017). Customer Referral Incentives and Social Media. *Management Science*, 63(10), 3514–3529. <https://doi.org/10.1287/mnsc.2016.2476>
25. Luo, J., Huang, S. (Sam), & Wang, R. (2021). A fine-grained sentiment analysis of online guest reviews of economy hotels in China. *Journal of Hospitality Marketing & Management*, 30(1), 71–95. <https://doi.org/10.1080/19368623.2020.1772163>
26. Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional Reviews: An Empirical Investigation of Online Review Manipulation. *American Economic Review*, 104(8), 2421–2455. <https://doi.org/10.1257/aer.104.8.2421>
27. Melander, L. (2019). Customer involvement in product development: Using Voice of the Customer for innovation and marketing. *Benchmarking: An International Journal*, 27(1), 215–231. <https://doi.org/10.1108/BIJ-04-2018-0112>
28. Nunes, J. C., Ordanini, A., & Giambastiani, G. (2021). The Concept of Authenticity: What It Means to Consumers. *Journal of Marketing*, 85(4), 1–20. <https://doi.org/10.1177/0022242921997081>
29. Paul, H., & Nikolaev, A. (2021). Fake review detection on online E-commerce platforms: A systematic literature review. *Data Mining and Knowledge Discovery*, 35(5), 1830–1881. <https://doi.org/10.1007/s10618-021-00772-6>
30. Pizam, A., Shapoval, V., & Ellis, T. (2016). Customer satisfaction and its measurement in hospitality enterprises: A revisit and update. *International Journal of Contemporary Hospitality Management*, 28(1), 2–35. <https://doi.org/10.1108/IJCHM-04-2015-0167>
31. Rambocas, M., & Pacheco, B. G. (2018). Online sentiment analysis in marketing research: A review. *Journal of Research in Interactive Marketing*, 12(2), 146–163. <https://doi.org/10.1108/JRIM-05-2017-0030>

32. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), Article 5. <https://doi.org/10.1038/s42256-019-0048-x>
33. Rundin, K., & Colliander, J. (2021). Multifaceted Influencers: Toward a New Typology for Influencer Roles in Advertising. *Journal of Advertising*, 50(5), 548–564. <https://doi.org/10.1080/00913367.2021.1980471>
34. Schulze, C., Schöler, L., & Skiera, B. (2014). Not All Fun and Games: Viral Marketing for Utilitarian Products. *Journal of Marketing*, 78(1), 1–19. <https://doi.org/10.1509/jm.11.0528>
35. Spirin, N., & Han, J. (2012). Survey on web spam detection: Principles and algorithms. *ACM SIGKDD Explorations Newsletter*, 13(2), 50–64. <https://doi.org/10.1145/2207243.2207252>
36. Subramani, M. R., & Rajagopalan, B. (2003). Knowledge-sharing and influence in online social networks via viral marketing. *Communications of the ACM*, 46(12), 300–307. <https://doi.org/10.1145/953460.953514>
37. Timoshenko, A., & Hauser, J. R. (2019). Identifying Customer Needs from User-Generated Content. *Marketing Science*, 38(1), 1–20. <https://doi.org/10.1287/mksc.2018.1123>
38. van Dieijen, M., Borah, A., Tellis, G. J., & Franses, P. H. (2020). Big Data Analysis of Volatility Spillovers of Brands across Social Media and Stock Markets. *Industrial Marketing Management*, 88, 465–484. <https://doi.org/10.1016/j.indmarman.2018.12.006>
39. Wu, Y., Ngai, E. W. T., Wu, P., & Wu, C. (2020). Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems*, 132, 113280. <https://doi.org/10.1016/j.dss.2020.113280>
40. Xiao, L., Li, X., & Zhang, Y. (2023). Exploring the factors influencing consumer engagement behavior regarding short-form video advertising: A big data perspective. *Journal of Retailing and Consumer Services*, 70, 103170. <https://doi.org/10.1016/j.jretconser.2022.103170>