Journal of
Theoretical and Applied
Electronic Commerce Research

MDPI

Article

Measuring Social Media Content Differentiation: R&D Investment and Strategic Positioning

Jixin Ye [1], Ling Liu [2,*], Yao He [2] and Aoxuan Liu [3]

[1] School of Business Administration, Southwestern University of Finance and Economics, Chengdu 611130, China
[2] School of Management Science and Engineering, Southwestern University of Finance and Economics, Chengdu 611130, China
[3] Xgate Co., Ltd., Chengdu 610041, China
* Correspondence: lingliu@swufe.edu.cn

Abstract

Firms can distinguish themselves through innovation and through the ways they communicate their value, yet how these dimensions relate in social media remains less understood. Using 420,100 posts from the verified Sina Weibo accounts of 365 listed Chinese firms during 2018–2020, this study develops *ContentDiff*, a firm-quarter measure that uses word embeddings to calculate the semantic distance between a firm's content and a contemporaneous industry benchmark. The measure provides a content-based perspective on firms' communication positions that complements expenditure- and engagement-based indicators. Panel regressions show a negative association between R&D investment and content differentiation, with a stronger association among larger firms and those with higher ESG performance. These findings extend the strategic-emphasis discussion of the R&D–marketing interface to social media content, suggesting that innovation investment and communicative distinctiveness do not necessarily move together. The negative association is consistent with a compensatory-positioning interpretation, although differences in business models and communication priorities may also contribute. Managers can use industry-relative content comparisons to identify how their messages differ from those of peers and examine whether those differences reflect their intended value proposition.

Keywords: social media content differentiation; digital marketing communication; firm-generated content; competitive positioning; strategic emphasis; R&D investment; word embeddings

Check for updates

Academic Editors: Brighton Nyagadza and Abu Bashar

Received: 27 September 2026
Revised: 3 November 2026
Accepted: 6 November 2026
Published: 10 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

In today's competitive business environment, firms continually seek ways to distinguish themselves from their rivals and build sustainable advantages [1,2]. They can pursue this goal by developing distinctive products and technologies through research and development (R&D), or by communicating their value to customers in distinctive ways [1,3–5]. Research on strategic emphasis and the R&D–marketing interface has long considered how firms balance these activities. They may reinforce one another [2,6–8], or they may reflect competing priorities in firms' use of resources [9–12]. This debate also raises a question about firms' communication content. How does R&D investment relate to the distinctiveness of the messages they share?

Social media provides a useful setting for exploring this question. It enables firms to communicate directly with their audiences [13–16], while their messages compete for

attention amid a continuous stream of content [17,18]. While R&D investments mainly support substantive differentiation at the product or technological level, social media reflects a communicative dimension of differentiation through what firms share and how their messages compare with those of competitors. Such differences have been documented in firms' social media content strategies, including among close competitors [19–21]. Firm-generated posts thus offer an observable basis for examining firms' relative communication positions in digital markets [20,22–25].

Examining these differences empirically requires measures of what firms communicate. Conventional indicators such as marketing expenditure, posting activity, and audience responses capture resource commitments, communication intensity, and audience reactions [18,26–28], but provide limited information about the substance of firms' messages [17,23,29]. Text-based research offers a complementary perspective, examining consumer mindset and competitive behavior [19,30,31], as well as content similarity between specific rivals [20,21,32]. Because differentiation is inherently comparative, the choice of reference group shapes how distinctiveness is understood. A firm may communicate differently from a particular competitor while still following themes common across its industry. Looking across industry peers can reveal how far its communication stands apart from the broader competitive conversation, complementing comparisons with individual rivals.

To make this comparison, we develop *ContentDiff*, a content-based measure of social media differentiation that applies Natural Language Processing (NLP) to firm-generated communication. Using word embeddings [33,34] to represent semantic relationships, *ContentDiff* quantifies the distance between a firm's content and a contemporaneous industry benchmark.

We then use this measure to examine how R&D investment relates to the distinctiveness of firms' social media communication, an aspect of the R&D–marketing relationship that remains less understood. Drawing on 420,100 posts from the verified Sina Weibo accounts of 365 listed Chinese firms between 2018 and 2020, we find a negative association between R&D investment and *ContentDiff*. This association is stronger among larger firms and those with higher ESG performance. The results are consistent with a compensatory-positioning interpretation, although differences in business models and communication priorities may also contribute to the observed pattern.

This study contributes to research on digital marketing communication and firm differentiation in two ways. First, *ContentDiff* provides a reusable firm-quarter measure of industry-relative semantic differentiation. In an increasingly crowded social media environment, it enables researchers to quantify how firms' messages differ from those of their industry peers, complementing expenditure-, activity-, and existing content-based indicators. Second, building on the distinction between substantive and communicative routes to differentiation, our analysis extends the R&D–marketing debate over complementarity and resource-allocation trade-offs to the content of firms' messages. The observed negative association suggests that innovation-related commitments and communicative distinctiveness need not move together. This perspective highlights digital communication not merely as a means of promoting innovation, but as a dimension of differentiation worthy of study in its own right.

The remainder of this paper is structured as follows. Section 2 develops the theoretical framework and hypotheses. Section 3 details the methodology for constructing *ContentDiff*. Section 4 outlines the data collection and empirical research design. Section 5 presents the empirical findings and robustness checks. Finally, Section 6 concludes with implications and future research directions.

2. Theoretical Background and Hypotheses

2.1. Firm Differentiation Through Innovation and Communication

The strategic-emphasis view of the R&D–marketing interface considers how firms balance resources and managerial attention between creating value through innovation and capturing value through market-facing activities [10,35]. This perspective also provides a starting point for examining differentiation, which refers to the creation of meaningful differences that distinguish a firm's offerings from those of its competitors [7,36]. Firms may pursue differentiation through substantive routes grounded in products, technologies, innovation, and other resource-based attributes [7,37], as well as through communicative routes that concern differences in how firms present and frame their value propositions through market-facing messages [4,5,27]. Competitive differentiation is therefore not only a matter of what firms develop or invest in, but also of how they communicate their distinctiveness to the market. These routes may reinforce one another, but firms may also place different emphasis on them as they pursue competitive advantage.

Substantive differentiation is often pursued through R&D investment, which can support the development of distinctive products, features, and technological capabilities [7,37,38]. R&D expenditure represents an input to innovation rather than a direct measure of realized product or technological differentiation, which is not directly observed in this study. However, corporate communication is not limited to technological innovation or product features, as it can also emphasize brand identity, service attributes, customer value, and other aspects of the firm's value proposition [4,5,24,31]. Firms make choices about which attributes and value propositions to emphasize, and how to frame them in their market-facing communication. These choices make communication a distinct dimension of differentiation, rather than merely a reflection of firms' underlying technological or product advantages. Research has also examined how firms communicate their strategic emphasis in corporate disclosures [39]. Building on this perspective, we examine how firms' R&D investments relate to the distinctiveness of their social media content relative to industry peers.

2.2. Communication Content as a Dimension of Differentiation

Social media gives firms direct control over their market-facing messages [16,22,40–42]. Through this content, firms communicate their products, values, activities, and brand positions [14,24,25,31]. Because firms can shape what they publish, the content itself becomes a dimension of competitive positioning, and firms can communicate in ways that differ from competitors [14,22,24,32].

Firms' social media content strategies may vary according to their business models and sources of competitive advantage. Firms that compete primarily through technological innovation may emphasize product features, technical capabilities, and development outcomes, whereas firms that rely more on branding, customer experience, or promotion may focus on different aspects of their market offerings [5,20]. Such differences in emphasis can shape the topics and language used in firms' social media communication, even among firms operating in the same industry [31,43]. Consequently, semantic differences between firms' content may reflect not only deliberate efforts to establish a distinctive market position, but also variation in their business models and communication priorities.

Assessing these semantic differences as a form of competitive differentiation requires a reference point. Firms in the same industry may draw on similar themes as they respond to shared market conditions, industry practices, and audience interests. Yet they can also depart from these patterns by emphasizing different aspects of their value propositions or framing familiar topics in their own ways [21,43]. An industry-wide benchmark places a firm's content within this broader competitive conversation, revealing the extent to which

its communication differs from prevailing patterns across its peers. Comparing content within the same period further situates the firm's position in its contemporaneous communication environment. This industry-relative perspective complements comparisons between specific rivals and provides a basis for examining how communicative positioning relates to R&D investment.

2.3. R&D Investment and Social Media Content Differentiation

The connections between innovation and communication invite two perspectives on how R&D investment relates to social media content differentiation. Each extends the R&D–marketing debate to firms' communication relative to industry peers, leading to competing hypotheses.

Prior research on the R&D–marketing interface highlights a complementary relationship between internal innovation and marketing activities [2,44]. R&D can support the development of technological capabilities and novel product offerings, while marketing helps communicate their distinctive attributes to target audiences and generate market returns from innovation. From this perspective, R&D and marketing commitments often reinforce one another to build sustained competitive advantage [6,8,45]. In the context of digital channels, rich social media campaigns effectively support product launches and signal novelty to external audiences [46]. Firms with greater R&D investment may have more innovation-related developments to communicate. When these developments differ from those emphasized by industry peers, highlighting them through social media can generate distinctive topics and language, increasing semantic distance from the industry benchmark. This complementarity perspective motivates the following hypothesis:

Hypothesis 1(a). *R&D investment is positively associated with social media content differentiation.*

An alternative perspective emphasizes that innovation and marketing activities may compete for limited organizational resources [10,11,35]. Firms may allocate resources toward developing products and technological capabilities through R&D or toward market penetration and brand building through marketing [7,9,10,12]. This perspective suggests that firms with lower R&D commitments may place greater emphasis on differentiating themselves through market-facing activities. In their social media communication, they may highlight brand identity, service attributes, or customer experience as alternative sources of distinction.

However, relying more heavily on communication does not necessarily make social media content more differentiated. What matters for *ContentDiff* is whether a firm's messages differ from the communication patterns prevailing among its industry peers. Emphasizing brand, service, or customer experience can increase semantic distance when these themes or their framing are uncommon within the industry. When similar messages are widely used by competitors, such an emphasis may produce little differentiation. Thus, a negative association between R&D investment and social media content differentiation could arise when firms with lower R&D commitments emphasize non-technological attributes in ways that differ from industry peers. This reasoning motivates the following competing hypothesis:

Hypothesis 1(b). *R&D investment is negatively associated with social media content differentiation.*

A negative association would be consistent with compensatory positioning, an interpretation in which firms with lower innovation-related resource commitments compensate by placing greater relative emphasis on distinctive market-facing communication. Business-model differences and communication-topic composition, however, could generate a similar association.

3. Social Media Content Differentiation Measurement

Unlike expenditure- and activity-based indicators that focus on marketing inputs or posting intensity, *ContentDiff* directly analyzes firm-generated content to represent how firms position their communication relative to industry peers. This section presents the measurement approach used to transform raw textual posts into this firm-level measure. Specifically, the approach combines word embeddings with distance metrics to quantify the semantic distinctiveness of a firm's social media content relative to a contemporaneous industry benchmark.

3.1. Methodological Foundation: Quantifying Semantic Uniqueness

Prior research has increasingly turned to text-based methodologies to analyze firm differentiation and marketing communication. Existing approaches, including binary term-presence vectors, TF-IDF, and dictionary-based measures, often represent textual content through lexical features or predefined categories. Beyond segment-based strategy measures [47], Hoberg and Phillips [48] constructed textual representations from 10-K product descriptions to measure product-market similarity. In the social media context, studies have used content coding, sentiment tools, and embeddings to examine creative strategies, consumer mindset, and user engagement [13,30,32]. Recent work has also examined competitive similarity in firm-generated posts [20,21,31]. These approaches demonstrate the value of textual information for analyzing firm differentiation and communication.

These measures were developed to address different questions, and their outputs are not directly interchangeable. They differ in measurement focus, in data and text representation, and in the basis of comparison. Table 1 compares selected measures along these dimensions. The comparison is intended to clarify differences in design rather than to rank existing measures.

For the text representation, we employ Word2Vec-based embeddings [33,34]. Word2Vec learns word vectors from patterns of word co-occurrence, allowing terms used in similar contexts to have similar representations. By aggregating these word vectors, we construct representations of firms' social media content. *ContentDiff* then measures the distance between each firm's quarterly content representation and a contemporaneous industry-quarter benchmark.

Table 1. Comparison of selected measures of firm differentiation and social media content similarity.

Study	Measurement Focus	Data & Representation	Comparison Basis
Litov et al. [47]	Corporate strategy uniqueness	Business-segment sales contribution	Industry strategy centroid
Ashley and Tuten [13]	Creative strategies and consumer engagement	Social media posts from leading brands; content analysis and categorical code	Frequencies of predefined content categories
Hoberg and Phillips [48]	Product-market similarity	Product descriptions in 10-K filings; binary word-presence vectors	Other firms in product space

Table 1. *Cont.*

Study	Measurement Focus	Data & Representation	Comparison Basis
Kübler et al. [30]	Sentiment and consumer mindset	Likes, comments, and shares; LIWC and SVM sentiment measures	Comparison of sentiment measures across 48 brands
Yuan et al. [31]	Business events and competitive behavior through FGC	Weibo posts; expert-labeled events; trigger terms	Pairwise similarities in firms' event structures, event mixes, and time sequences, measured using Pearson correlations
Lysyakov et al. [20]	Social media content similarity	Retailer-generated Twitter content; text-based similarity	Identified competing retailers
Yang et al. [32]	FGC characteristics and engagement	Weibo posts and user-behavior data; Word2Vec word embeddings; vector-based semantic similarity	Cosine similarity with official product information
Wang et al. [21]	Semantic differentiation and linguistic style matching	WeChat posts; TF-IDF vectors; cosine similarity; SC-LIWC function-word proportions	Local sequential benchmark; focal post versus the three immediately preceding posts in the same shopping circle
ContentDiff	Industry-relative semantic distinctiveness	Firm-generated Weibo posts; Word2Vec-based representations aggregated at the firm-quarter level	Contemporaneous industry-quarter content benchmark

3.2. Constructing the Differentiation Measure

The quantification process is formalized in the equations below. The foundation for this process is a pre-trained Word2Vec model that transforms each word in our corpus of Weibo posts into a 300-dimensional embedding vector. Because content differentiation is inherently relative to competitors, *ContentDiff* is defined against the contemporaneous communication benchmark of firms in the same industry. The specific calculation of our measure then proceeds as follows:

First, to represent the semantic content of a single social media post, we average the vectors of its constituent words to create a single post vector, V_{post}.

$$V_{\text{post}} = \frac{V_{\text{word1}} + V_{\text{word2}} + \cdots + V_{\text{word}n}}{n} \quad (1)$$

Second, to create a stable firm-level representation of a firm's overall content strategy for a given quarter, we calculate the dimension-wise median of its post vectors within that quarter. The median is chosen for its robustness to outliers from atypical posts. This results in a firm document vector, V_{firm}.

$$V_{\text{firm}} = \text{median}(V_{\text{post1}}, V_{\text{post2}}, \ldots, V_{\text{post}k}) \quad (2)$$

Third, to establish a benchmark for typical communication within an industry-quarter, we calculate the dimension-wise median of the firm document vectors for all m firms in the same industry and quarter. This industry-period benchmark represents the semantic center

of prevailing communication within the focal firm's peer group at a given point in time. The resulting benchmark is defined as the industry document vector, V_{industry}.

$$V_{\text{industry}} = \text{median}(V_{\text{firm1}}, V_{\text{firm2}}, \ldots, V_{\text{firm}m}) \tag{3}$$

Finally, with these firm-level and industry-level vectors established, we define Social Media Content Differentiation ($ContentDiff$) as the Euclidean distance between a firm's content vector and its corresponding industry content vector. The final calculation is as follows:

$$ContentDiff_{it} = \sqrt{\sum_{j=1}^{300}(V_{\text{firm},it,j} - V_{\text{industry},it,j})^2} \tag{4}$$

where i denotes the firm, t the quarter, and j the vector dimension. $V_{\text{industry},it,j}$ is the jth dimension of the benchmark vector for firm i's industry in quarter t. $ContentDiff_{it}$ quantifies the semantic uniqueness of firm i's social media content in quarter t. A larger value indicates greater semantic deviation from the prevailing communication pattern among contemporaneous industry peers, capturing the distinctiveness of the firm's market-facing social media communication. The workflow of this quantification process is visually summarized in Figure 1.

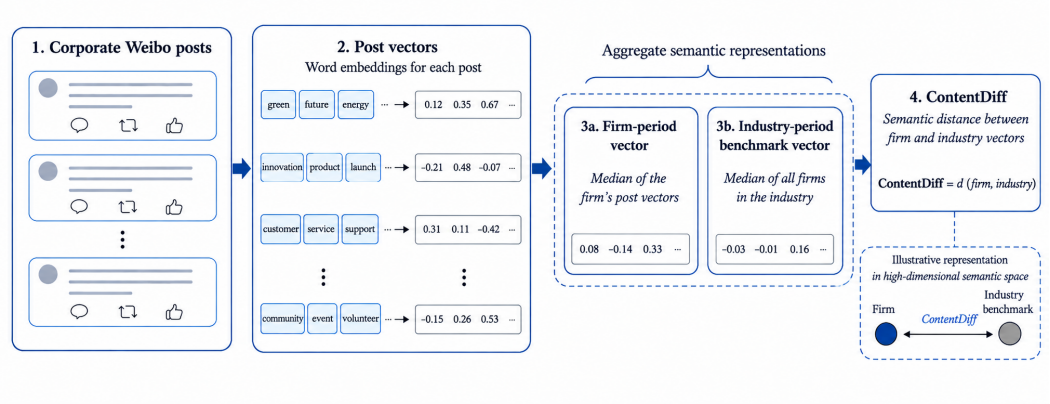


Figure 1. Measuring Social Media Content Differentiation. The figure summarizes the construction of $ContentDiff$, from firm-generated Weibo posts and post-level semantic representations to firm-period vectors, industry-period benchmark vectors, and the final semantic distance measure.

3.3. Measurement Illustration

To illustrate the semantic-distance logic underlying $ContentDiff$, we use three sample posts published by the same firm (stock code: 603660) in March 2020. Posts A and B describe the firm's business scope and a specific product, respectively, and therefore contain thematically similar content. In contrast, Post C concerns a festival-related corporate message and represents a substantively different topic. The original posts and their English translations are provided in the Appendix A (Figure A1).

After tokenization and embedding, we calculate the pairwise Euclidean distances among the three post vectors using the same distance logic employed in the construction of $ContentDiff$. As reported in Table 2, the distance between Posts A and B is relatively small (0.349583), whereas the distances between Post C and the other two posts are substantially larger (0.899569 and 0.819169, respectively). The numerical pattern is consistent with the qualitative differences among the three posts and provides an intuitive illustration of how semantic distance captures variation in firm-generated social media content. This post-level example illustrates the semantic-distance component underlying the firm-period measure rather than the complete construction of $ContentDiff$.

Table 2. Pairwise distances among three illustrative posts.

	Post A	Post B	Post C
Post A	0.000000		
Post B	0.349583	0.000000	
Post C	0.899569	0.819169	0.000000

Note: This table reports pairwise Euclidean distances among three sample social media posts published by a single firm in March 2020. Vector construction follows Equation (1), and distance computation follows Equation (4). Lower values indicate higher semantic similarity.

4. Data and Models

4.1. Data

This study utilizes a panel dataset constructed from multiple sources. First, we sourced financial metrics and firm-specific information for firms listed on the Shanghai Stock Exchange from the China Stock Market & Accounting Research (CSMAR) and Wind databases. From the Wind database, we also adopted the first-level Wind industry classification standards to serve as the basis for our industry benchmarks. (The Wind Industry Classification is a standardized system for Chinese listed firms. The first-level classification includes eleven sectors: information technology, utilities, health care, consumer discretionary, industrials, real estate, consumer staples, materials, telecommunication services, energy, and finance).

To capture firms' market-facing digital communication, we identified the official Weibo accounts of listed firms by checking their corporate websites and searching for firm names on Weibo. We retained verified accounts corresponding to the listed companies and excluded unverified accounts and accounts associated with specific brands or business units. We collected all posts published by these accounts between 2018 and 2020. As a major microblogging platform in China, Sina Weibo enables firms to communicate directly with consumers and other external audiences. Officially verified corporate accounts represent firm-controlled channels through which companies publish market-facing content. Unlike user-generated content, these posts reflect communication choices made by the firms themselves and therefore provide observable traces of firm-level digital marketing communication [20,25,31,32].

After excluding the finance sector, our data collection process yielded an initial sample of 365 non-financial firms. This 365-firm sample is crucial as it serves as the basis for constructing our industry-level content differentiation benchmarks (detailed in Section 3.2). The quarterly and industry distribution of posts from this sample is presented in Table 3. In total, 420,100 posts were collected. To construct the regression sample, we matched firm-quarter content differentiation measures with firm-level financial data. Observations with missing values in the required regression variables, including R&D expenditure and other financial indicators, were excluded. This resulted in a final sample of 2058 firm-quarter observations for the empirical analysis.

4.2. Models

We estimate panel regressions to examine the relationship between social media content differentiation and R&D investment, as specified in Model (5). The model also accounts for social media post volume because posting intensity may be associated with both firms' communication strategies and the observed distribution of their social media content [18,49].

$$ContentDiff_{it} = \alpha_0 + \beta_1 RD_{it} + \beta_2 Post_{it} + CV'_{it}\gamma + Ind_i + Qrt_t + \varepsilon_{it} \quad (5)$$

The dependent variable, $ContentDiff_{it}$, is the social media content differentiation for firm i in quarter t. The primary independent variable is RD_{it}, the natural logarithm of one plus R&D spending, and $Post_{it}$ is the total number of posts published by firm i in quarter t.

Table 3. Quarterly distribution of posts by industry.

Industry Quarter	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	Total
2018Q1	6101	1280	715	11,889	23,662	267	2737	190	478	751	48,070
2018Q2	6052	1159	1050	12,336	7570	286	2690	279	586	794	32,802
2018Q3	6302	756	1522	11,489	6885	324	2356	515	614	678	31,441
2018Q4	6140	869	1605	11,690	6982	301	2721	476	718	477	31,979
2019Q1	5656	623	1095	10,783	6717	305	2615	283	482	617	29,176
2019Q2	4605	544	1393	12,428	7253	342	2717	296	728	728	31,034
2019Q3	5314	579	1542	13,796	7677	299	2855	330	996	1032	34,420
2019Q4	5162	434	1575	11,996	7230	310	2701	275	952	1202	31,837
2020Q1	8508	301	1332	10,950	8091	457	3267	739	1400	2052	37,097
2020Q2	6268	582	1565	12,828	7678	456	3108	721	1315	3677	38,198
2020Q3	5739	864	1697	13,551	7697	440	3290	564	1351	3013	38,206
2020Q4	5115	1002	1810	12,053	7490	419	3058	502	1578	2813	35,840
Total	70,962	8993	16,901	145,789	104,932	4206	34,115	5170	11,198	17,834	420,100

Note: This table presents the quarterly distribution of posts from our initial sample of 365 non-financial firms, disaggregated by industry. The columns correspond to the first-level Wind industry classifications, which are: (1) Information Technology; (2) Utilities; (3) Health care; (4) Consumer Discretionary; (5) Industrials; (6) Real Estate; (7) Consumer Staples; (8) Materials; (9) Communication Services; (10) Energy.

To account for alternative explanations, the model includes a consistent set of control variables (CV'_{it}). The selling expense ratio (*SalesExp*) captures broader marketing-related expenditures that may be associated with firms' communication strategies. We also control for firm size (*Size*) to reflect differences in scale, visibility, and market power [22,38], financial leverage (*Lev*) to capture financing structure [7,28], and total asset growth (*Growth*) to reflect recent firm expansion [8,27]. In addition, we include ownership concentration (*Cent*3) to account for governance and control [50], ESG performance (*ESG*) to reflect sustainability performance and stakeholder orientation [37,51], and a dummy variable for state-owned enterprises (*SOE*, coded as 1 for SOEs and 0 otherwise) to account for ownership structure.

Because *ContentDiff* is constructed relative to the contemporaneous industry benchmark, industry (*Ind*) and quarter (*Qrt*) fixed effects are included to absorb persistent cross-industry differences in baseline communication patterns and common time shocks. In more stringent specifications, we further replace separate industry and quarter fixed effects with industry-by-quarter fixed effects to account for time-varying conditions specific to each industry. Robust standard errors clustered at the firm level are used throughout to address potential heteroscedasticity and within-firm serial correlation. A detailed description of all variables is provided in Table 4.

Furthermore, we conduct cross-sectional heterogeneity analyses to examine whether the relationship between R&D investment and content differentiation varies across firms with different characteristics. We focus on firm size (*Size*) and ESG performance (*ESG*), which capture differences in firms' public visibility and stakeholder exposure that may be associated with their communication strategies. The sample is divided at the median market value and the industry-median ESG rating, respectively, and the model is re-estimated separately for each subgroup with industry-by-quarter fixed effects.

Table 4. Measures of all variables.

Variable	Symbol	Measure
Social media content differentiation	*ContentDiff*	The Euclidean distance between the firm document vector and the industry document vector of posts.
Social media post volume	*Post*	The total number of posts published by a firm in a given quarter.
R&D investment	*RD*	Log of (R&D spending + 1).
Selling expense ratio	*SalesExp*	Selling expenses/operating revenue.
Firm size	*Size*	Log of firm market value.
Financial leverage	*Lev*	Total liabilities/total assets.
Firm growth	*Growth*	Total assets growth rate.
Ownership concentration	*Cent*3	Shareholding ratio of top three shareholders.
ESG performance	*ESG*	Quarterly ESG rating, provided by Huazheng ESG Rating.
State ownership	*SOE*	1 for state-owned enterprises and 0 otherwise.
Time effect	*Qrt*	Time dummy variable.
Industry effect	*Ind*	Industry dummy variable, based on the Wind industry classification standards.

5. Empirical Results

5.1. Descriptive Statistics

Table 5 summarizes the descriptive statistics for our sample. The average social media content differentiation ($ContentDiff$) was 0.446 ($SD = 0.254$), with a median of 0.393 and a range from 0.123 to 2.985. This indicates significant variation in firms' differentiated content strategies on the Weibo platform during the sample period. Similarly, the average social media post volume ($Post$) was 122.474 ($SD = 410.428$), with a median of 36, and values ranging from 1 to 7783. The substantial standard deviation and wide range highlight significant heterogeneity in firms' posting frequency. This diversity in both social media content differentiation and post volume suggests that firms employed varied and distinct social media content strategies.

Table 5. Summary of descriptive statistics.

Variable	*N*	Mean	Std. Dev.	Min	Median	Max
ContentDiff	2058	0.446	0.254	0.123	0.393	2.985
Post	2058	122.474	410.428	1	36	7783
RD	2058	17.748	2.332	0.000	17.721	23.963
SalesExp	2058	0.123	0.121	0.000	0.077	0.730
Size	2058	23.572	1.599	20.716	23.253	28.630
Lev	2058	0.443	0.203	0.039	0.437	1.522
Growth	2058	0.085	0.212	−0.470	0.038	2.844
*Cent*3	2058	56.818	16.690	12.296	58.774	94.985
ESG	2058	6.859	1.212	2	7	9
SOE	2058	0.348	0.477	0	0	1

Note: For the binary *SOE* indicator, the mean reports the proportion of firm-quarter observations associated with state-owned firms, and the standard deviation is approximately $\sqrt{p(1-p)}$, where p is that proportion.

In addition, the average value of R&D investment (RD) was 17.748 ($SD = 2.332$) in its natural logarithmic form (log of R&D spending plus one), with values ranging from

0 to 23.963. This variation indicates substantial differences in innovation-related resource commitment among the sample firms. The remaining control variables, including the selling expense ratio, firm size, financial leverage, firm growth, ownership concentration, ESG performance, and state-owned enterprise status, also exhibit considerable variation across the sample, as reported in Table 5.

The Pearson correlation coefficients among the main variables are presented in Table 6. Consistent with the subsequent regression results, social media content differentiation ($ContentDiff$) is negatively correlated with R&D investment (RD), although the magnitude of the correlation is relatively small ($r = -0.1396, p < 0.01$). $ContentDiff$ also shows generally weak correlations with the remaining firm-level variables. In addition, the variance inflation factor (VIF) values for Model (5) are low (mean = 1.37), indicating no severe multicollinearity concerns.

Table 6. Pearson correlation analysis.

	ContentDiff (1)	***Post*** (2)	***RD*** (3)	***SalesExp*** (4)	***Size*** (5)	***Lev*** (6)	***Growth*** (7)	***Cent3*** (8)	***ESG*** (9)	***SOE*** (10)
(1)	1.0000									
(2)	−0.1079 ***	1.0000								
(3)	−0.1396 ***	0.0205	1.0000							
(4)	−0.0894 ***	0.0204	−0.2428 ***	1.0000						
(5)	−0.0867 ***	0.0975 ***	0.5369 ***	−0.3795 ***	1.0000					
(6)	−0.0320	−0.0318	0.3211 ***	−0.3458 ***	0.4986 ***	1.0000				
(7)	−0.0366 *	−0.0196	0.0979 ***	−0.0791 ***	0.0103	0.0018	1.0000			
(8)	−0.1215 ***	0.0711 ***	0.0103	−0.0608 ***	0.1451 ***	−0.0068	−0.0031	1.0000		
(9)	−0.0188	0.1680 ***	0.1400 ***	−0.1517 ***	0.4353 ***	0.1202 ***	−0.0042	0.1044 ***	1.0000	
(10)	0.0528 **	0.1310 ***	0.1149 ***	−0.2563 ***	0.4326 ***	0.2466 ***	−0.1044 ***	0.0898 ***	0.3409 ***	1.0000

Note: ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

5.2. Regression Results

This section presents the empirical results of our analysis. We estimate panel regressions based on Model (5) to examine the relationship between R&D investment and social media content differentiation.

5.2.1. R&D Investment and Content Differentiation

The baseline regression, investigating the direct association between R&D investment and social media content differentiation, is presented in Column (1) of Table 7. In this initial specification, industry and quarter fixed effects, as well as firm-level controls, are deliberately excluded to highlight the raw, unadjusted relationship. The coefficient on R&D investment is negative and statistically significant ($\beta_{RD} = -0.0152, p = 0.003$), suggesting that firms with lower R&D investments tend to have higher levels of social media content differentiation.

To examine whether this relationship holds after accounting for firm characteristics, we introduce a set of controls alongside industry and quarter fixed effects. The coefficient for R&D remains negative in this specification (Column 2, Table 7; $\beta_{RD} = -0.0164, p = 0.065$).

We then account for communication volume by including the number of posts in the full model (Column 3, Table 7). The inverse relationship between R&D and differentiation remains statistically significant ($\beta_{RD} = -0.0177, p = 0.046$), while post volume itself shows no reliable association with differentiation ($\beta_{Post} = -0.000066, p > 0.10$). Thus, the observed relationship persists after accounting for posting intensity.

Table 7. Main regression results.

	(1) *ContentDiff*	(2) *ContentDiff*	(3) *ContentDiff*	(4) *ContentDiff*
RD	−0.0152 ***	−0.0164 *	−0.0177 **	−0.0178 *
	(0.00512)	(0.00886)	(0.00883)	(0.00913)
Post			−0.0000661	−0.0000662
			(0.0000411)	(0.0000419)
SalesExp		−0.184 *	−0.161	−0.157
		(0.109)	(0.108)	(0.111)
Size		−0.0133	−0.0114	−0.0114
		(0.0118)	(0.0118)	(0.0122)
Lev		−0.0270	−0.0326	−0.0306
		(0.0500)	(0.0498)	(0.0510)
Growth		−0.0747 ***	−0.0763 ***	−0.0771 ***
		(0.0263)	(0.0264)	(0.0283)
*Cent*3		−0.00171 ***	−0.00159 **	−0.00157 **
		(0.000653)	(0.000646)	(0.000652)
ESG		−0.00274	0.000287	0.000505
		(0.00958)	(0.00958)	(0.00982)
SOE		0.0299	0.0376	0.0390
		(0.0258)	(0.0259)	(0.0266)
Industry FE	NO	YES	YES	-
Quarter FE	NO	YES	YES	-
Industry × Quarter FE	NO	NO	NO	YES
Constant	0.715 ***	1.196 ***	1.154 ***	1.149 ***
	(0.093)	(0.186)	(0.183)	(0.188)
Observations	2058	2058	2058	2054
R^2	0.0195	0.0964	0.1068	0.1195
Adjusted R^2	0.0190	0.0849	0.0949	0.0744
F Statistic	8.80	5.412	5.125	4.82

Note: The dependent variable is *ContentDiff*. Industry and quarter fixed effects are included in Columns (2) and (3). Column (4) includes industry-by-quarter fixed effects, which subsume the separate industry and quarter fixed effects. The industry-by-quarter fixed-effects specification excludes four singleton observations, resulting in an estimation sample of 2054 firm-quarter observations. Robust standard errors clustered at the firm level are reported in parentheses. Statistical significance is based on *p*-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Finally, to account for time-varying industry-specific factors, we replace the separate industry and quarter fixed effects in the baseline model with industry-by-quarter fixed effects (Column 4, Table 7). This specification absorbs shocks and communication conditions shared by firms within the same industry and quarter. The coefficient on R&D investment remains negative and is statistically significant at the 10% level ($\beta_{RD} = -0.0178, p = 0.052$). The estimated coefficient is also similar in magnitude to that in the full model (Column 3, Table 7), in line with the negative association proposed in Hypothesis 1(b). The positive association proposed in Hypothesis 1(a) is not observed. The negative association is consistent with a compensatory-positioning interpretation, although alternative explanations involving firm characteristics, industry-specific communication practices, and unobserved strategic differences cannot be ruled out.

The estimated coefficient is modest in magnitude. A one-standard-deviation increase in *RD* is associated with a 0.164-standard-deviation decrease in *ContentDiff*. The R^2 of 0.1195 in Column (4) indicates that the model explains a limited share of the variation in content differentiation. Inference on β_{RD} rests on a firm-clustered *t*-test rather than on the overall *F* statistic. Detailed results for this specification are reported in Appendix B Table A1,

and corresponding results for the heterogeneity and robustness analyses are reported in Appendix B Table A2.

5.2.2. Cross-Sectional Heterogeneity

This section examines how the inverse relationship between social media content differentiation and R&D investment varies across different firm characteristics. Heterogeneity analyses are conducted from two perspectives: firm size (proxied by market value) and ESG performance. These factors are explored because larger firms often possess distinct resource endowments and face increased market attention [22,52], while firms with stronger ESG performance typically prioritize long-term value creation and sustainability [53,54]. These differences may lead to varying strategic deployments of social media content.

To examine heterogeneity based on firm size, the Model (5) specification was estimated separately for firms differentiated by market value (below or above the median). The analysis indicates that the negative relationship between social media content differentiation and R&D investment is statistically significant for larger firms ($\beta_{RD} = -0.02999$, $p = 0.047$), but not for smaller firms ($\beta_{RD} = -0.00204, p > 0.1$) (See Columns (1) and (2) of Table 8). A Fisher-type permutation test (bootstrap = 1000) confirms that this difference between the two size-based groups is statistically significant (difference between two coefficients $= 0.028, p < 0.001$). Firm size therefore appears to be an important context for the relationship between innovation investment and communication differentiation.

Table 8. Cross-sectional heterogeneity.

Variable	*ContentDiff*		*ContentDiff*	
	(1) Large Size	(2) Small Size	(3) High ESG	(4) Low ESG
RD	−0.02999 **	−0.00204	−0.02259 **	0.00615
	(0.01497)	(0.00554)	(0.01044)	(0.01266)
Constant	1.176 ***	1.517 ***	1.041 ***	0.911 **
	(0.407)	(0.562)	(0.224)	(0.364)
Controls	YES	YES	YES	YES
Industry × Quarter FE	YES	YES	YES	YES
Observations	1088	958	1365	676
Adjusted R^2	0.0881	0.0640	0.0924	0.0693
Difference in coefficients	0.028 ($p < 0.001$)		0.029 ($p < 0.001$)	

Note: The dependent variable is *ContentDiff*. Columns (1)–(2) split firms by median market value, and Columns (3)–(4) split firms by industry-median Huazheng ESG rating. All specifications include industry-by-quarter fixed effects. Difference-test *p*-values are based on Fisher-type permutation tests with 1000 bootstrap replications. Robust standard errors clustered at the firm level are reported in parentheses. Statistical significance is based on *p*-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Similarly, heterogeneity based on firms' ESG performance was examined using Model (5), with high-ESG firms defined as those above the industry median ESG score. These results are shown in Columns (3) and (4) of Table 8. For firms with high ESG scores, the negative relationship between social media content differentiation and R&D investment is statistically significant ($\beta_{RD} = -0.02259, p = 0.031$). Conversely, for firms with low ESG scores, this relationship is not statistically significant ($\beta_{RD} = 0.00615, p > 0.1$). The Fisher-type permutation test (bootstrap = 1000) confirms a statistically significant difference between these ESG-based groups (difference between two coefficients = 0.029,

$p < 0.001$). The negative association is therefore more pronounced among firms with higher ESG performance.

Collectively, these heterogeneity analyses show that the inverse relationship between R&D investment and social media content differentiation varies across firms. Its greater strength among larger and higher-ESG firms suggests that organizational characteristics may shape how innovation investment relates to digital communication positioning.

5.3. Robustness Tests

A series of additional analyses was conducted to assess the robustness of the main findings. First, we replace the primary content differentiation measure with an alternative embedding-based measure and re-estimate the models. Second, we perform matched-sample analyses based on posting volume. Third, we re-estimate the model after imposing minimum posting volume thresholds. Overall, the results remain broadly consistent with the main regressions.

5.3.1. Alternative Measure of Social Media Content Differentiation

As an alternative semantic representation, we employ Moka Massive Mixed Embedding (M3E), a text embedding model that maps text inputs into dense vectors and supports sentence-level semantic similarity computation [55]. We use this model to examine whether the main results are sensitive to the choice of text representation.

The two approaches operate at different levels of representation. Word2Vec learns word-level vectors from local contexts [33,34], and our baseline procedure averages these vectors to represent each post. M3E instead encodes text inputs directly using a RoBERTa-based model trained with contrastive learning and in-batch negative sampling on more than 22 million Chinese sentence pairs [55]. This distinction allows us to assess the robustness of the findings using text-level embeddings as an alternative to averaged word embeddings.

We construct $ContentDiff_{\mathrm{M3E}}$ using the same aggregation and benchmarking procedure as the baseline measure, replacing the Word2Vec representation with M3E embeddings. The correlation coefficient between the two indicators ($ContentDiff$ and $ContentDiff_{\mathrm{M3E}}$) is 0.68, and statistically significant at the 5% significance level. The positive correlation between the Word2Vec- and M3E-based measures provides evidence that $ContentDiff$ captures a common underlying dimension of semantic differentiation across alternative text representations.

Subsequently, $ContentDiff$ was replaced with $ContentDiff_{\mathrm{M3E}}$, and all previous regression specifications were re-estimated. Table 9 demonstrates that the key results across all specifications hold true under this alternative text embedding approach. Specifically, the findings concerning the relationship between social media content differentiation and R&D investment (Model 5) remained consistent. The consistency of the results across the Word2Vec- and M3E-based measures suggests that the main findings are not highly sensitive to the embedding approach used.

5.3.2. Propensity Score Matching (PSM)

As an additional robustness check, we employ propensity score matching (PSM) based on firms' posting volume to account for selection on observables related to posting intensity. We then re-estimate the regressions within the matched sample to examine whether the observed relationship persists after accounting for such differences.

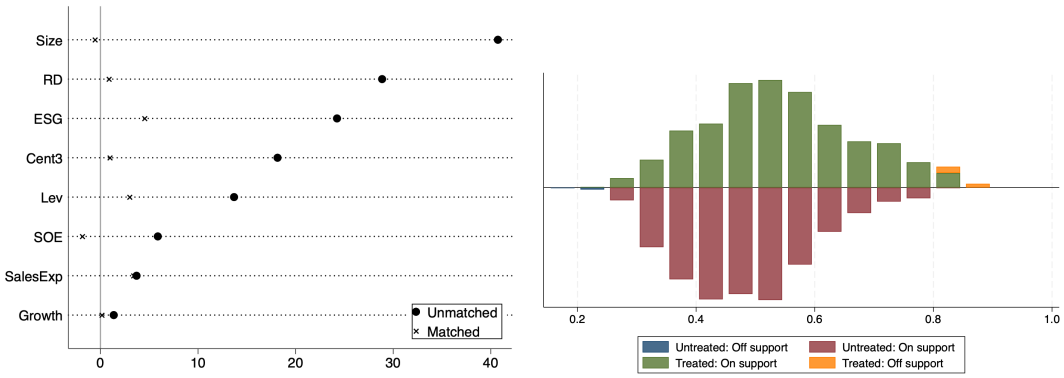
Table 9. Robustness tests using the M3E measure.

	(1) $ContentDiff_{M3E}$	(2) $ContentDiff_{M3E}$	(3) $ContentDiff_{M3E}$	(4) $ContentDiff_{M3E}$
RD	−0.1154 ***	−0.0770 *	−0.0924 **	−0.0924 **
	(0.0342)	(0.0442)	(0.0422)	(0.0430)
Post			−0.000763 **	−0.000771 **
			(0.000352)	(0.000363)
SalesExp		−2.249 ***	−1.980 **	−1.938 **
		(0.845)	(0.824)	(0.839)
Size		−0.220 ***	−0.199 ***	−0.196 ***
		(0.0734)	(0.0709)	(0.0721)
Lev		−0.0747	−0.140	−0.163
		(0.423)	(0.412)	(0.421)
Growth		−0.401 *	−0.419 *	−0.373
		(0.223)	(0.222)	(0.235)
*Cent*3		−0.00962 *	−0.00815	−0.00814
		(0.00539)	(0.00531)	(0.00536)
ESG		−0.0370	−0.00208	0.00242
		(0.0733)	(0.0719)	(0.0738)
SOE		0.0422	0.130	0.137
		(0.206)	(0.204)	(0.207)
Industry FE	NO	YES	YES	-
Quarter FE	NO	YES	YES	-
Industry × Quarter FE	NO	NO	NO	YES
Constant	7.84 ***	13.48 ***	13.00 ***	12.89 ***
	(0.613)	(1.390)	(1.356)	(1.379)
Observations	2058	2058	2058	2054
R^2	0.0225	0.1551	0.1827	0.2009
Adjusted R^2	0.0221	0.1443	0.1718	0.1600
F Statistic	11.41	6.249	6.422	6.07

Note: This table reports Model (5) regression results using the M3E-based content differentiation measure as the dependent variable. The industry-by-quarter fixed-effects specification excludes four singleton observations, resulting in an estimation sample of 2054 firm-quarter observations. Robust standard errors clustered at the firm level are reported in parentheses. Statistical significance is based on *p*-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Firms were divided into two groups based on the median of their social media post volume (*Post*), with those above the median assigned to the treatment group (1) and those below to the control group (0). All control variables served as matching variables, utilizing a kernel matching approach to estimate the average treatment effect on the treated.

After matching, the estimated average treatment effect on the treated (ATT) for *ContentDiff* is −0.1840 ($t = -15.69$, $SE = 0.0117$, $p < 0.001$). The effectiveness of the PSM procedure is illustrated in Figure 2, which displays the standardized percentage bias across covariates (Figure 2a) and the distribution of propensity scores (Figure 2b). Figure 2 clearly shows a substantial reduction in bias for each covariate after matching, indicating improved covariate balance.



(**a**) (**b**)

Figure 2. Covariate Balance and Propensity Score Distribution. (**a**) Standardized % bias across covariates. It reports the standardized percentage bias across covariates, showing a substantial reduction in imbalance after matching. (**b**) Distribution of the propensity scores. It depicts the distribution of the propensity scores, confirming sufficient overlap and common support between the treatment and control groups.

Further consistent with these patterns, Table 10 reports standardized percentage bias values below the commonly used 10% threshold for all covariates, while the corresponding post-matching *t*-tests are statistically insignificant. These results indicate improved covariate balance between the matched groups. The matched sample is then used to re-estimate the baseline regression model.

Table 10. Covariate balance before and after propensity score matching.

Variable	Sample U/M	Mean		Bias (%)	Bias Reduction (%)	*t*-Test	
		Treated	Control			*t*	*p*
RD	U	18.07200	17.40600	28.9		6.55	<0.001
	M	17.99900	17.97900	0.9	97.0	0.21	0.836
SalesExp	U	0.12517	0.12071	3.7		0.84	0.401
	M	0.12666	0.12257	3.4	8.5	0.75	0.452
Size	U	23.88100	23.24400	40.7		9.21	<0.001
	M	23.81200	23.82100	−0.6	98.6	−0.12	0.904
Lev	U	0.45640	0.42857	13.7		3.11	0.002
	M	0.45620	0.45013	3.0	78.2	0.68	0.498
Growth	U	0.08603	0.08313	1.4		0.31	0.757
	M	0.08681	0.08651	0.1	89.7	0.03	0.974
Cent3	U	58.28200	55.26900	18.1		4.11	<0.001
	M	57.74400	57.58400	1.0	94.7	0.22	0.825
ESG	U	7.00090	6.70900	24.2		5.50	<0.001
	M	6.99140	6.93690	4.5	81.3	1.03	0.303
SOE	U	0.36200	0.33400	5.9		1.33	0.183
	M	0.35221	0.36106	−1.9	68.4	−0.42	0.673

Note: U = Unmatched sample; M = Matched sample. This table presents the balance diagnostics for covariates before and after PSM. %bias indicates the standardized percentage bias. The bias reduction and non-significant *p*-values for the *t*-tests (for matched samples) confirm improved covariate balance.

The regression results for the PSM-matched sample are reported in Table 11. The coefficient on R&D investment remains negative and statistically significant. This finding indicates that the negative association between R&D investment and social media content differentiation persists after matching firms on observed covariates related to posting intensity, providing supplementary evidence of the robustness of the main findings.

Table 11. Robustness tests using the PSM-matched sample.

	(1) *ContentDiff*	(2) *ContentDiff*	(3) *ContentDiff*	(4) *ContentDiff*
RD	−0.0141 **	−0.0179 *	−0.0196 *	−0.0195 *
	(0.0056)	(0.0103)	(0.0103)	(0.0106)
Post			−0.0000650	−0.0000642
			(0.0000413)	(0.0000413)
SalesExp		−0.178	−0.157	−0.153
		(0.109)	(0.108)	(0.110)
Size		−0.0088	−0.00734	−0.00718
		(0.0126)	(0.0125)	(0.0129)
Lev		−0.0348	−0.0382	−0.0372
		(0.0508)	(0.0508)	(0.052)
Growth		−0.0743 ***	−0.0754 ***	−0.0752 ***
		(0.0263)	(0.0264)	(0.028)
*Cent*3		−0.0016 **	−0.00153 **	−0.00151 **
		(0.000649)	(0.000643)	(0.000649)
ESG		−0.003938	−0.000734	−0.000630
		(0.00965)	(0.00970)	(0.00994)
SOE		0.0284	0.0358	0.0371
		(0.0257)	(0.0259)	(0.0265)
Industry FE	NO	YES	YES	-
Quarter FE	NO	YES	YES	-
Industry × Quarter FE	NO	NO	NO	YES
Constant	0.6970 ***	1.127 ***	1.098 ***	1.088 ***
	(0.101)	(0.184)	(0.182)	(0.185)
Observations	2038	2038	2038	2033
R^2	0.0148	0.0887	0.0985	0.1130
Adjusted R^2	0.0143	0.0769	0.0864	0.0675
F Statistic	6.38	5.01	4.722	4.40

Note: This table reports Model (5) regression results from using the PSM-matched sample. The dependent variable is *ContentDiff*. The industry-by-quarter fixed-effects specification excludes five singleton observations, resulting in an estimation sample of 2033 firm-quarter observations. Robust standard errors clustered at the firm level are reported in parentheses. Statistical significance is based on *p*-values. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

5.3.3. Sensitivity to Minimum Quarterly Posting Volume

Quarterly posting volume varies considerably across the 2058 firm-quarter observations. Although most observations contain at least ten posts, 3.40% contain fewer than two, 11.13% contain fewer than five, and 21.23% contain fewer than ten. Firm-quarter content vectors derived from a small number of posts may be less representative of firms' quarterly communication and more sensitive to individual messages. Including posting volume as a regression control does not directly address this potential measurement concern. Consequently, we re-estimate the model using restricted samples that require a minimum of two, five, and ten posts per firm-quarter, respectively.

As reported in Table 12, the coefficient on R&D investment remains negative and statistically significant across all three restricted samples. The association therefore persists when firm-quarter observations with sparse social media activity are excluded.

Table 12. Robustness tests with minimum quarterly posting requirements.

Variable	$Post \geq 2$		$Post \geq 5$		$Post \geq 10$	
	(1)	(2)	(3)	(4)	(5)	(6)
RD	−0.0185 **	−0.0189 **	−0.0103 **	−0.0101 **	−0.0106 **	−0.0106 **
	(0.0089)	(0.0092)	(0.0045)	(0.0045)	(0.0050)	(0.0052)
Constant	0.9546 ***	0.9457 ***	0.7852 ***	0.7701 ***	0.7282 ***	0.7159 ***
	(0.1735)	(0.1768)	(0.1569)	(0.1599)	(0.1583)	(0.1589)
Controls	YES	YES	YES	YES	YES	YES
Industry FE	YES	-	YES	-	YES	-
Quarter FE	YES	-	YES	-	YES	-
Industry × Quarter FE	NO	YES	NO	YES	NO	YES
Observations	1988	1984	1829	1824	1621	1614
R^2	0.1146	0.1288	0.1172	0.1338	0.1282	0.1474
Adjusted R^2	0.1024	0.0826	0.1040	0.0841	0.1134	0.0947
F Statistic	4.32	4.05	3.50	3.23	2.52	2.40

Note: This table reports the regression results of Model (5) across subsamples subject to varying minimum posting thresholds per firm-quarter. Columns (1)–(2), (3)–(4), and (5)–(6) correspond to thresholds of at least two, five, and ten posts, respectively. The dependent variable is *ContentDiff*. The reduction in observations reflects the exclusion of singleton observations within industry–quarter groups under the industry-by-quarter fixed-effects specification. Robust standard errors clustered at the firm level are reported in parentheses. Statistical significance is based on *p*-values. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

6. Discussion and Conclusions

6.1. Discussion

In digital markets, firms seek to distinguish themselves not only through what they develop, but also through what they communicate. Our findings bring this communicative dimension into focus by examining the content firms share relative to their industry peers. Measures of marketing expenditure, posting activity, and audience engagement offer valuable insights into firms' digital presence, while *ContentDiff* adds a perspective on how their messages differ in meaning. This perspective highlights social media content as part of firms' market positioning and provides a basis for considering how that positioning relates to their innovation investments.

The complementarity perspective suggests that innovation can give firms distinctive developments to communicate. Our findings point to a different pattern: the negative association between R&D investment and social media content differentiation is consistent with Hypothesis 1(b), rather than the positive association proposed in Hypothesis 1(a). This finding suggests that innovation-related investment and communicative differentiation do not necessarily move together. Firms with lower R&D investment tend to communicate in ways that differ more from prevailing industry patterns. Because *ContentDiff* is defined relative to contemporaneous industry peers and the preferred model includes industry-by-quarter fixed effects, this association is not simply a reflection of differences in typical communication themes across industries. One interpretation is compensatory positioning, whereby firms draw on their communication to express distinctive value propositions when innovation-related activities are less prominent [10,12]. Differences in business models and communication priorities among firms within the same industry may also contribute to this pattern. Together, these possibilities invite a broader view of differentiation, in which digital communication is a dimension of market positioning alongside firms' innovation-related commitments.

The relationship also varies across firms, as the negative association is more pronounced among larger firms and those with stronger ESG performance. Their greater visibility

and exposure to stakeholder expectations may make the way they communicate their value propositions particularly consequential. These findings suggest that the relationship between innovation investment and digital communication should be understood within the firm's organizational and market context. They also point to the value of examining how firms with different characteristics use social media content to position themselves among their peers.

Taken together, these findings highlight the value of considering firms' innovation investments alongside the content they share in digital markets. Bringing these perspectives together opens up new opportunities to understand how firms distinguish themselves within their competitive environment.

6.2. Theoretical and Methodological Contributions

This study contributes to research on digital marketing communication by introducing *ContentDiff* as a reusable measure of social media content differentiation. Beyond indicators of communication inputs, activity volume, and audience response, examining message content offers a direct window into how firms position themselves in digital markets. By capturing semantic distance from a contemporaneous industry benchmark, *ContentDiff* enables researchers to quantify how a firm's communication departs from prevailing patterns among its peers. It complements existing content-based approaches and offers a scalable way to study this dimension of digital market positioning across firms and periods.

The study also extends the strategic-emphasis perspective on the R&D–marketing interface to the content of firms' digital communication. Prior research has examined both firms' innovation and marketing investments and how they communicate strategic emphasis in corporate disclosures [10,35,39]. Our analysis brings attention to how firms' social media messages differ from those of their industry peers. The observed negative association suggests that firms' innovation-related commitments and communicative distinctiveness need not move together. By bringing these substantive and communicative dimensions into the same discussion, our analysis highlights social media content as a distinct dimension of market positioning alongside firms' commitments to innovation.

These contributions draw on evidence from Chinese listed firms on Sina Weibo, adding an emerging-economy perspective to research on digital marketing communication.

6.3. Managerial Implications

Managers can use an industry-relative view of content to better understand their firm's position in the social media landscape. *ContentDiff* offers a starting point for identifying whether a firm's communication broadly follows industry patterns or differs from them. Managers can then examine their posts alongside those of industry peers to understand which topics and messages contribute to the difference. This comparison can make content strategy discussions more concrete and reveal opportunities to communicate the firm's value in ways that resonate with its intended audience.

With a clearer view of their communication position, managers can consider which aspects of the firm's value proposition deserve greater emphasis. The relationship observed between R&D investment and content differentiation highlights the value of looking beyond innovation alone, as firms may also communicate through product applications, services, brand identity, or other attributes relevant to their audiences. The choice of emphasis should reflect the firm's business model, capabilities, and market positioning, so that its social media content presents a coherent and credible account of what it offers.

Finally, firms can evaluate these communication choices against their intended objectives. Content differentiation shows how a firm's messages compare with industry patterns, while measures such as audience engagement and brand perception can help managers as-

sess how those messages are received. Considering both perspectives can inform decisions about how to develop future content. Industry-relative comparisons can also help investors and analysts identify firms whose communication warrants closer examination and better understand the value propositions they present to the market.

6.4. Limitations and Future Research

These findings should be considered in light of several limitations, which also suggest directions for future research. First, *ContentDiff* focuses on text and does not show how audiences perceive differences in firms' communication. Future studies could combine semantic measures with audience responses and incorporate features such as tone, images, and video to develop a richer understanding of content differentiation.

Second, the observed association between R&D investment and content differentiation may reflect compensatory positioning, business models, or communication topics. Future research could examine these explanations more directly and explore how innovation investment and communication choices evolve over time.

Third, the empirical setting is limited to verified Sina Weibo accounts of Chinese listed firms during 2018–2020. Applying this approach across platforms, countries, and types of firms would offer further insight into how firms communicate their value and distinguish themselves in digital markets.

In conclusion, this study introduces an industry-relative semantic measure of social media content differentiation and connects it to the strategic-emphasis debate on the R&D–marketing interface. By examining how substantive innovation commitments relate to communicative distinctiveness in firm-generated content, it offers a foundation for future research on how firms position themselves in digital markets.

Author Contributions: Conceptualization: J.Y. and L.L.; Methodology: J.Y., Y.H. and A.L.; Formal analysis and investigation: J.Y.; Writing—original draft preparation: J.Y.; Writing—review and editing: J.Y. and L.L.; Funding acquisition: L.L.; Data curation: Y.H. and A.L.; Supervision: L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China [71401139] and Sichuan Science and Technology Program [2019YJ0251].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study come from the third-party sources described in Section 4. The underlying source data are not publicly redistributed by the authors, as the authors do not hold unrestricted redistribution rights. Enquiries regarding data access may be directed to the corresponding author. Any sharing is subject to applicable permissions and licensing conditions.

Acknowledgments: This research is supported by the National Natural Science Foundation of China [71401139] and Sichuan Science and Technology Program [2019YJ0251]. During the preparation of this work the authors used AI tools to correct spelling and grammar errors. After using this tool, the authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: Author Yao He was employed by Meituan, and author Aoxuan Liu was employed by Chengdu Xgate Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

R&D	Research and Development
NLP	Natural Language Processing

Appendix A. Examples of Firm-Generated Weibo Posts

Figure A1 presents three original Weibo posts from the same firm (stock code: 603660) in March 2020. Posts A and B describe the firm's business scope and a specific product, respectively, and therefore contain thematically similar content. In contrast, Post C presents a festival-related corporate message and represents a different content theme. These examples illustrate the semantic variation in the firm-generated content underlying the construction of *ContentDiff*.

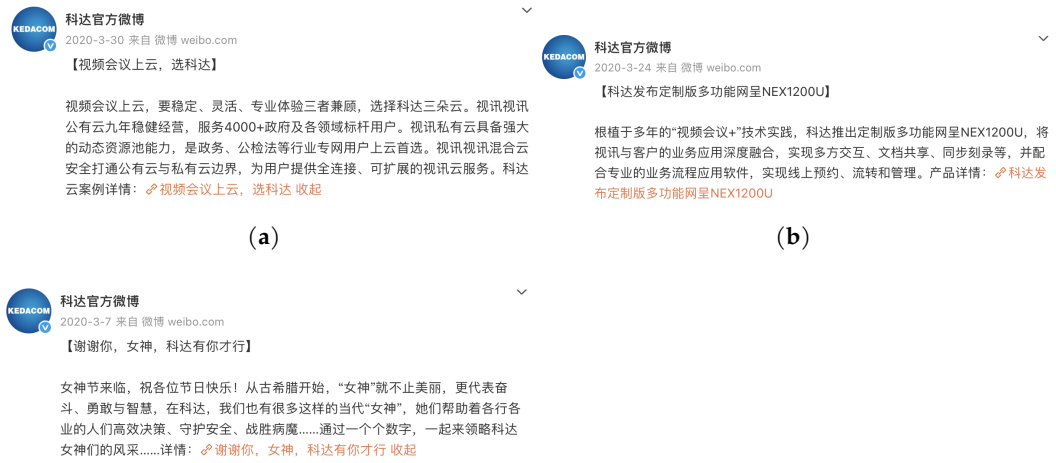


Figure A1. Examples of original firm-generated Weibo posts: (**a**) Post A, cloud video conferencing services; (**b**) Post B, the NEX1200U product; and (**c**) Post C, a Goddess Day message about women at Kedacom.

English translations of the posts are provided below.

(**a**) **Screenshot of Post A.** "Cloud Video Conferencing: Choose Kedacom." When moving video conferencing to the cloud, it is essential to balance stability, flexibility, and professional experience. Choose Kedacom's "three clouds." The public cloud for video conferencing has operated steadily for nine years, serving over 4000 government and benchmark users in various fields. The private cloud for video conferencing offers dynamic resource pooling capabilities and serves private-network users in industries including government, public security, procuratorates, and courts. The hybrid cloud for video conferencing connects public and private cloud environments and provides scalable video cloud services.

(**b**) **Screenshot ofPost B.** "Kedacom Releases Customized Multi-functional Network Presenter NEX1200U." Building on years of "Video Conferencing+" technical practice, Kedacom launched the customized multi-functional network presenter NEX1200U. It integrates video communication with customer business applications, enabling multi-party interaction, document sharing, synchronous recording, and other functions. Together with business-process application software, it also supports online appointments, circulation, and management.

(**c**) **Screenshot of Post C.** "Thank You, Goddesses; Kedacom Thrives Because of You." As Goddess Day approaches, we wish everyone a happy holiday. Since ancient Greece, "goddesses" have represented not only beauty, but also struggle, courage, and wisdom. Kedacom also has many such contemporary "goddesses." They help people across different

industries make efficient decisions, safeguard security, and overcome illness. The post then introduces several female Kedacom employees and their work.

Appendix B. Detailed Regression Results

Table A1 reports the detailed regression results for Column (4) of Table 7, which includes industry-by-quarter fixed effects. Table A2 reports the R&D coefficient estimates for the industry-by-quarter fixed-effects specifications in the heterogeneity and robustness analyses.

Table A1. Detailed results of the main regression.

ContentDiff	Coefficient	Std. Coefficient	Robust Std. Err.	*t*	*p*	[95% Conf. Interval]
RD	−0.0177955	−0.1635136	0.0091289	−1.95	0.052	[−0.0357637, 0.0001726]
Post	−0.0000662	−0.1070773	0.0000419	−1.58	0.115	[−0.0001487, 0.0000163]
SalesExp	−0.156876	−0.0745424	0.1106935	−1.42	0.158	[−0.37475, 0.0609979]
Size	−0.0113949	−0.0716931	0.0121695	−0.94	0.350	[−0.0353476, 0.0125579]
Lev	−0.0305646	−0.0244913	0.0509515	−0.60	0.549	[−0.1308505, 0.0697214]
Growth	−0.0771174	−0.064463	0.0283137	−2.72	0.007	[−0.1328462, −0.0213887]
*Cent*3	−0.0015669	−0.1030736	0.000652	−2.40	0.017	[−0.0028502, −0.0002836]
ESG	0.000505	0.0024099	0.009816	0.05	0.959	[−0.0188155, 0.0198256]
SOE	0.0389732	0.1534898	0.0265822	1.47	0.144	[−0.0133476, 0.091294]
Constant	1.149404	−0.0532805	0.1876632	6.12	<0.001	[0.7800336, 1.518775]

Note: This table provides supplementary regression results corresponding to Column (4) of Table 7. The dependent variable is *ContentDiff*. The coefficient estimates, *t*-statistics, *p*-values, and 95% confidence intervals are based on unstandardized variables. The standardized-coefficient column reports estimates using a standardized dependent variable and standardized continuous explanatory variables. The model includes industry-by-quarter fixed effects, which absorb the separate industry and quarter fixed effects. Robust standard errors are clustered at the firm level.

Table A2. Detailed results of the heterogeneity and robustness analyses.

Analysis	β_{RD}	Std. β_{RD}	Robust Std. Err.	*t*	*p*	[95% Conf. Interval]
Cross-sectional heterogeneity						
Large Firm	−0.0299912	−0.2830131	0.0149667	−2.00	0.047	[−0.0595357, −0.0004468]
Small Firm	−0.0020391	−0.0151552	0.0055377	−0.37	0.713	[−0.0129715, 0.0088933]
High ESG	−0.0225902	−0.220698	0.0104411	−2.16	0.031	[−0.0431577, −0.0020226]
Low ESG	0.0061544	0.0449879	0.0126551	0.49	0.627	[−0.0188469, 0.0311556]
Robustness tests						
M3E Measure	−0.0923642	−0.1202564	0.0429891	−2.15	0.033	[−0.1769781, −0.0077504]
PSM-matched	−0.0194722	−0.1683959	0.0106353	−1.83	0.068	[−0.0404062, 0.0014619]
Minimum 2 Posts	−0.0188628	−0.1733202	0.0091799	−2.05	0.041	[−0.0369332, −0.0007925]
Minimum 5 Posts	−0.0100677	−0.0925069	0.0045122	−2.23	0.026	[−0.0189518, −0.0011837]
Minimum 10 Posts	−0.0105766	−0.0971824	0.0051718	−2.05	0.042	[−0.0207614, −0.0003918]

Note: This table provides supplementary regression results corresponding to Tables 8, 9, 11 and 12, including detailed statistical information not reported in the tables. The dependent variable is *ContentDiff*, except in the M3E specification, where it is $ContentDiff_{M3E}$. β_{RD}, *t*-statistics, *p*-values, and 95% confidence intervals are based on unstandardized regression coefficients reported in the main text. Std. β_{RD} denotes the standardized coefficient on R&D investment, estimated using a standardized dependent variable and standardized continuous explanatory variables. The model includes industry-by-quarter fixed effects, which absorb the separate industry and quarter fixed effects. Robust standard errors are clustered at the firm level.

References

1. Lauga, D.O.; Ofek, E.; Katona, Z. When and how should firms differentiate? Quality and advertising decisions in a duopoly. *J. Mark. Res.* **2022**, *59*, 1252–1265. [CrossRef]
2. Koshksaray, A.A.; Quach, S.; Trinh, G.; Keivani, S.B.; Thaichon, P. Brand competitiveness antecedents: The interaction effects of marketing and R&D expenditure. *J. Retail. Consum. Serv.* **2023**, *75*, 103532. [CrossRef]
3. Miller, D. Relating Porter's business strategies to environment and structure: Analysis and performance implications. *Acad. Manag. J.* **1988**, *31*, 280–308. [CrossRef] [PubMed]
4. McAlister, L.; Srinivasan, R.; Jindal, N.; Cannella, A.A. Advertising effectiveness: The moderating effect of firm strategy. *J. Mark. Res.* **2016**, *53*, 207–224. [CrossRef]
5. Soberman, D.A.; Xiang, Y. Designing the content of advertising in a differentiated market. *Int. J. Res. Mark.* **2022**, *39*, 190–211. [CrossRef]
6. Slotegraaf, R.J.; Pauwels, K. The impact of brand equity and innovation on the long-term effectiveness of promotions. *J. Mark. Res.* **2008**, *45*, 293–306. [CrossRef]
7. Guo, B.; Wang, J.; Wei, S.X. R&D spending, strategic position and firm performance. *Front. Bus. Res. China* **2018**, *12*, 14. [CrossRef]
8. Peng, C.L.; Wei, A.P.; Chen, M.L.; Huang, W.T. Synergy Between R&D and Advertising on Shareholder Value: Does Firm Size Matter? *Can. J. Adm. Sci.* **2018**, *35*, 47–64. [CrossRef]
9. Kaiser, U. A microeconometric note on product innovation and product innovation advertising. *Econ. Innov. New Technol.* **2005**, *14*, 573–582. [CrossRef]
10. Sridhar, S.; Narayanan, S.; Srinivasan, R. Dynamic relationships among R&D, advertising, inventory and firm performance. *J. Acad. Mark. Sci.* **2014**, *42*, 277–290. [CrossRef]
11. Askenazy, P.; Breda, T.; Irac, D. Advertising and R&D: Theory and Evidence from France. *Econ. Innov. New Technol.* **2016**, *25*, 33–56. [CrossRef]
12. Cavenaile, L.; Celik, M.A.; Roldan-Blanco, P.; Tian, X. Style over substance? Advertising, innovation, and endogenous market structure. *J. Monet. Econ.* **2025**, *149*, 103683. [CrossRef]
13. Ashley, C.; Tuten, T. Creative strategies in social media marketing: An exploratory study of branded social content and consumer engagement. *Psychol. Mark.* **2015**, *32*, 15–27. [CrossRef]
14. Rathore, A.K.; Ilavarasan, P.V.; Dwivedi, Y.K. Social media content and product co-creation: An emerging paradigm. *J. Enterp. Inf. Manag.* **2016**, *29*, 7–18. [CrossRef]
15. Shankar, V.; Grewal, D.; Sunder, S.; Fossen, B.; Peters, K.; Agarwal, A. Digital marketing communication in global marketplaces: A review of extant research, future directions, and potential approaches. *Int. J. Res. Mark.* **2022**, *39*, 541–565. [CrossRef]
16. Li, Y.; Yin, Y.; Wang, Z. How Does the Relevance of Firm-Generated Content to Products Affect Consumer Brand Attitudes? *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 19. [CrossRef]
17. Tafesse, W.; Wien, A. Implementing social media marketing strategically: An empirical assessment. *J. Mark. Manag.* **2018**, *34*, 732–749. [CrossRef]
18. Lopes, A.R.; Casais, B. Digital content marketing: Conceptual review and recommendations for practitioners. *Acad. Strateg. Manag. J.* **2022**, *21*, 1–17.
19. Lee, D.; Hosanagar, K.; Nair, H.S. Advertising content and consumer engagement on social media: Evidence from Facebook. *Manag. Sci.* **2018**, *64*, 5105–5131. [CrossRef]
20. Lysyakov, M.; Kannan, P.; Viswanathan, S.; Zhang, K. Retailer Differentiation in Social Media: An Investigation of Firm-Generated Content on Twitter. *J. Mark.* **2025**, *89*, 39–58. [CrossRef]
21. Wang, F.; Yang, L.; Wang, S.; Wang, F.; Zhang, M. Standing out while fitting in: Content strategy and performance in social media marketing. *J. Bus. Res.* **2026**, *216*, 116360. . Corrigendum in *J. Bus. Res.* **2026**, *217*, 116479. [CrossRef]
22. Corejova, T.; Rostasova, M.; Rovnanova, A.; Valica, M. Competitive differentiation in the digital environment and social networks. In *Proceedings of the Advances in Cross-Section Data Methods in Applied Economic Research*; Springer: Cham, Switzerland, 2020; pp. 705–716. [CrossRef]
23. Dolan, R.; Conduit, J.; Frethey-Bentham, C.; Fahy, J.; Goodman, S. Social media engagement behavior: A framework for engaging customers through social media content. *Eur. J. Mark.* **2019**, *53*, 2213–2243. [CrossRef]
24. Li, F.; Larimo, J.; Leonidou, L.C. Social media marketing strategy: Definition, conceptualization, taxonomy, validation, and future agenda. *J. Acad. Mark. Sci.* **2021**, *49*, 51–70. [CrossRef]
25. Ahmad, B.L.; Alrwashdeh, M. Digital Content Marketing, Consumer Trust, and Digital Engagement on Social Commerce Platforms: Effects on Brand Attitude in Iraq's Hospitality Sector. *J. Theor. Appl. Electron. Commer. Res.* **2026**, *21*, 257. [CrossRef]
26. Lewis, R.A.; Reiley, D.H. Online ads and offline sales: Measuring the effect of retail advertising via a controlled experiment on Yahoo! *Quant. Mark. Econ.* **2014**, *12*, 235–266. [CrossRef]
27. Bayer, E.; Srinivasan, S.; Riedl, E.J.; Skiera, B. The impact of online display advertising and Paid Search Advertising Relative to Offline Advertising on Firm Performance and Firm Value. *Int. J. Res. Mark.* **2020**, *37*, 789–804. [CrossRef]

28. Jindal, N. The Impact of Advertising and R&D on Bankruptcy survival: A Double-Edged Sword. *J. Mark.* **2020**, *84*, 22–40. [CrossRef]
29. Valos, M.J.; Maplestone, V.L.; Polonsky, M.J.; Ewing, M. Integrating social media within an integrated marketing communication decision-making framework. *J. Mark. Manag.* **2017**, *33*, 1522–1558. [CrossRef]
30. Kübler, R.V.; Colicev, A.; Pauwels, K.H. Social Media's Impact on the Consumer Mindset: When to Use Which Sentiment Extraction Tool? *J. Interact. Mark.* **2020**, *50*, 136–155. [CrossRef]
31. Yuan, H.; Deng, W.; Ma, B.; Qian, Y. Monitoring events of market competitors: A text mining method for analyzing massive firm-generated social media. *J. Theor. Appl. Electron. Commer. Res.* **2023**, *18*, 908–927. [CrossRef]
32. Yang, X.; Su, X.; Lin, J.; Liu, J.; Miao, M.; Jiang, Y. How do we engage customers in official activity in the virtual brand community? The effects of enterprise-generated contents on customer engagement behaviors. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 3520–3546. [CrossRef]
33. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781. [CrossRef]
34. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2013; Volume 26.
35. Mizik, N.; Jacobson, R. Trading Off Between Value Creation and Value Appropriation: The Financial Implications of Shifts in Strategic Emphasis. *J. Mark.* **2003**, *67*, 63–76. [CrossRef]
36. Kotler, P.; Keller, K.L. *Marketing Management*, 13th ed.; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2009.
37. Coluccia, D.; Dabić, M.; Del Giudice, M.; Fontana, S.; Solimene, S. R&D innovation indicator and its effects on the market. An empirical assessment from a financial perspective. *J. Bus. Res.* **2020**, *119*, 259–271. [CrossRef]
38. Pourkarimi, P.; Kam, E. The impact of R&D and advertising on firm performance in high-tech industries—evidence from the us information and communications technology industry. *J. Quant. Econ.* **2022**, *20*, 723–753. [CrossRef]
39. Gensler, S.; Oehring, K.; Wiesel, T. Reported and Communicated Shifts in Strategic Emphasis and Firm Performance. *Int. J. Res. Mark.* **2024**, *41*, 220–240. [CrossRef]
40. Raji, R.A.; Rashid, S.; Ishak, S. The mediating effect of brand image on the relationships between social media advertising content, sales promotion content and behaviuoral intention. *J. Res. Interact. Mark.* **2019**, *13*, 302–330. [CrossRef]
41. Colicev, A.; Kumar, A.; O'Connor, P. Modeling the relationship between firm and user generated content and the stages of the marketing funnel. *Int. J. Res. Mark.* **2019**, *36*, 100–116. [CrossRef]
42. Homburg, C.; Wielgos, D.M. The value relevance of digital marketing capabilities to firm performance. *J. Acad. Mark. Sci.* **2022**, *50*, 666–688. [CrossRef] [PubMed]
43. Schanke, S.; Ray, G.; Adomavicius, G.; Wagle, M. Content and style of firm-generated posts on social media: A study of user engagement on hedonic and utilitarian product pages on Facebook. *J. Assoc. Inf. Syst.* **2023**, *25*, 341–361. [CrossRef]
44. Griffin, A.; Hauser, J.R. Integrating R&D and marketing: A review and analysis of the literature. *J. Prod. Innov. Manag.* **1996**, *13*, 191–215. [CrossRef]
45. Srinivasan, S.; Pauwels, K.; Silva-Risso, J.; Hanssens, D.M. Product innovations, advertising, and stock returns. *J. Mark.* **2009**, *73*, 24–43. [CrossRef]
46. Baum, D.; Spann, M.; Füller, J.; Thürridl, C. The impact of social media campaigns on the success of new product introductions. *J. Retail. Consum. Serv.* **2019**, *50*, 289–297. [CrossRef]
47. Litov, L.P.; Moreton, P.; Zenger, T.R. Corporate strategy, analyst coverage, and the uniqueness paradox. *Manag. Sci.* **2012**, *58*, 1797–1815. [CrossRef]
48. Hoberg, G.; Phillips, G. Text-based network industries and endogenous product differentiation. *J. Polit. Econ.* **2016**, *124*, 1423–1465. [CrossRef]
49. Peng, M.; Xu, Z.; Huang, H. How does information overload affect consumers' online decision process? An event-related potentials study. *Front. Neurosci.* **2021**, *15*, 695852. [CrossRef] [PubMed]
50. Zhang, Y.; Wang, H.; Zhou, X. Dare to be different? Conformity versus differentiation in corporate social activities of Chinese firms and market responses. *Acad. Manag. J.* **2020**, *63*, 717–742. [CrossRef]
51. Hsiao, C.Y.; Wang, S.L.; Li, Z.L.; Liu, H. ESG, R&D and Brand Value: Taking the Chinese Listed Companies of 3C Equipment Manufacturing Industry. *Adv. Econ. Manag. Res.* **2023**, *5*, 106. [CrossRef]
52. Healy, P.M.; Palepu, K.G. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *J. Account. Econ.* **2001**, *31*, 405–440. [CrossRef]
53. Eccles, R.G.; Ioannou, I.; Serafeim, G. The impact of corporate sustainability on organizational processes and performance. *Manag. Sci.* **2014**, *60*, 2835–2857. [CrossRef]

54. Hao, J.; He, F. Corporate social responsibility (CSR) performance and green innovation: Evidence from China. *Finance Res. Lett.* **2022**, *48*, 102889. [CrossRef]
55. Wang, Y.; Sun, Q.; He, S. M3E: Moka Massive Mixed Embedding Model. 2023. Available online: https://huggingface.co/moka-ai/m3e-base (accessed on 1 November 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.