

Article

From Responding to Co-Creating: Identifying Strategic CSR Processes Using Pre-Trained Language Models

Xia Fan, Dongxia Cheng  and Qianhua Lei *

School of Business Administration, South China University of Technology, Guangzhou 510645, China; fanxia@scut.edu.cn (X.F.); 202310188439@mail.scut.edu.cn (D.C.)

* Correspondence: bmlqh@scut.edu.cn

Highlights

Please indicate how your work links to systems science via your contributions to systems practice, theory, and/or methodology.

- This study integrates pre-trained language models (PLMs) with Latent Dirichlet Allocation (LDA) to develop a hybrid identification framework that enables paragraph-level classification and topic extraction of strategic corporate social responsibility (SCSR) from large-scale textual corpora. By providing a systematic statistical analysis of the prevalence and characteristics of SCSR engagement, this work contributes to the advancement of systems methodology.
- This work advances systems practice in CSR governance by empirically demonstrating how linguistic patterns in corporate disclosures reflect distinct organizational logics, offering a scalable, data-driven approach for regulators and stakeholders to assess the strategic orientation of CSR engagement across firms.

What are the main findings and/or the implications of the main findings?

- On average, only 25.6% of CSR activities exhibit the full set of strategic characteristics—strategic intent, business synergy, resource investment, and value creation. Among the four dimensions, business synergy (29.6%) emerges as the most critical bottleneck. Moreover, state-owned enterprises outperform in policy-aligned SCSR topics, while non-state-owned enterprises lead in market-oriented domains.
- These findings imply that enhancing the integration of CSR with core value chain operations—rather than merely increasing resource inputs—is essential for the strategic transformation of CSR and provides empirical tools for quality assessment, regulatory oversight, and firm-level strategic decision-making.

Abstract

Corporate social responsibility (CSR) is shifting from compliance-driven to strategically embedded practice, yet large-scale identification of strategic CSR (SCSR) remains methodologically challenging due to reliance on surveys or ratings. This study proposes a hybrid identification framework combining pre-trained language models (PLMs) and Latent Dirichlet Allocation to conduct paragraph-level semantic analysis on 3022 CSR reports from Chinese manufacturing firms. The approach effectively identifies and extracts the topics of both strategic and non-strategic CSR. Findings show that only 25.6% of CSR behaviors simultaneously possess strategic intent, business synergy, resource investment, and value output. SCSR practices are highly concentrated on value chain optimization and R&D collaboration, with significant variations across ownership types and industry environments. This study contributes a novel PLM-based methodology for SCSR identification



Academic Editor: Mitsuru Kodama

Received: 30 June 2026

Revised: 9 August 2026

Accepted: 11 August 2026

Published: 17 August 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

and offers empirical evidence and practical tools for CSR quality assessment, regulatory oversight, and strategic decision-making in manufacturing sectors.

Keywords: strategic corporate social responsibility; creating shared value; pre-trained language models; topic modeling; CSR reporting; manufacturing firms

1. Introduction

The global market is currently undergoing profound structural adjustments, and many companies are facing the dual pressures of maintaining short-term financial performance while fulfilling long-term social responsibilities [1]. Against this backdrop, more companies are exploring the integration of social and environmental issues into their business agendas and long-term strategies to realize both social value and sustainable economic value [2,3]. Unlike traditional corporate social responsibility (CSR) behaviors primarily driven by compliance, strategic CSR (SCSR) is manifested in clear strategic intent, close integration with value chain activities, dedicated resource investment [3,4], and continuous pursuit of measurable business value (such as cost savings, innovation, and market access) [5,6].

In practice, companies have explored various paths to integrate social responsibility with strategy. For example, Alibaba Group Holding Limited (headquartered in Hangzhou, China) developed the Ant Forest program to organically integrate environmental advocacy into Alipay's user scenarios, making low-carbon behavior part of the platform ecosystem, and Contemporary Amperex Technology Co. Limited (CATL; headquartered in Ningde, China) has deeply embedded its carbon neutrality goals into battery technology innovation and capacity planning, thus transforming environmental constraints into long-term technological advantages. These cases demonstrate that SCSR can not only enhance corporate reputation but also directly translate into operational efficiency, innovation capabilities, and market expansion to promote synergy between social and commercial value [7,8].

Although policy-driven initiatives and CSR reporting have surged, it remains difficult to determine the extent to which companies have integrated CSR into their business objectives, specifically which of their CSR efforts are strategic, truly focused on building long-term competitiveness and creating shared value, and which are non-strategic efforts aimed at complying with regulations, shaping an image, or following trends.

In theory, SCSR is centered around whether enterprises can internalize social needs as a source of business innovation to co-create social and economic value [3]. However, the literature still lacks an integrated and operational framework for identification and analysis [4,9,10]. Most related research is conducted at the macro level of whether SCSR is important and does not delve into its process mechanisms. Recent empirical evidence suggests that CSR activities do not create value in isolation. Several studies converge on the finding that CSR value creation depends on strategic integration [6,11]. Additionally, Lu et al. [12] find that prolonged poor performance drives firms to reduce irresponsible behaviors and enhance innovation, further underscoring the role of strategic responsiveness in CSR dynamics.

Methodologically, traditional approaches, including rating databases, questionnaires, and keyword-based text analysis, are constrained by limited generalizability, social desirability bias, and shallow semantic processing [13–15], making large-scale objective CSR assessment difficult. Recent studies have extended CSR inquiry to specific contexts through varied methodological approaches. These include organizational agility in energy firms via survey-based structural equation modeling [16], financial performance pathways through quantitative survey analysis [17], and policy-mediated effects in emerging economies via

qualitative case study design [18]. Yet these studies predominantly rely on single-case or qualitative designs and have yet to establish a replicable framework for large-scale SCSR identification.

To address these gaps, this study poses a central research question: How can strategic CSR be systematically identified, reliably differentiated from non-strategic CSR, and meaningfully characterized within the context of Chinese manufacturing firms? This overarching question is further decomposed into three interrelated sub-questions: (1) How can a hybrid text-mining framework be developed to effectively identify and classify strategic versus non-strategic CSR from unstructured corporate disclosures? (2) What are the prevailing patterns, structural features, and behavioral logics of strategic and non-strategic CSR practices among Chinese listed manufacturing firms? and (3) How does the strategic differentiation of CSR configurations systematically affect corporate economic and non-economic performance?

Pre-trained language models (PLMs), which have advantages in contextual understanding and implicit meaning inference, provide a unique approach to text analysis of CSR behaviors. In recent years, they have been applied to climate response analysis and ESG practice assessment [19,20], while large language models have also been employed for tasks such as greenwashing detection [21]. The core objective of this study is to construct a framework for identifying SCSR and to utilize advanced text mining techniques to achieve efficient, objective differentiation and in-depth analysis of strategic and non-strategic behaviors in CSR practices. This study adopts the process perspective of the creating shared value theory and constructs a four-dimensional strategic intent—business synergy—resource investment—value output identification framework. Furthermore, it combines PLMs and Latent Dirichlet Allocation (LDA) to analyze the CSR reports of Chinese A-share manufacturing companies between 2021 and 2023, thus revealing the structural and practical characteristics of strategic and non-strategic CSR.

This study uses a sample of manufacturing companies listed in China primarily for the following reasons. First, China is at a critical stage of transitioning from rapid economic growth to high-quality development; CSR practices such as ESG disclosure, green investment, and supply chain responsibility are developing rapidly, and the diverse social responsibility models enable us to identify SCSR. Second, Chinese manufacturing companies are deeply embedded in global value chains, and their social responsibility performance affects not only the domestic environment and society but also global supply chain standards and international market access. Using Chinese manufacturing companies as a sample provides empirical evidence from an emerging economy for theoretical research on SCSR.

This study makes three main contributions to the literature. First, at the theoretical level, the literature on CSR primarily considers the interests of stakeholders outside the company; few studies explore the common interests of all stakeholders, including the company itself. On the basis of the creating shared value perspective, this study deconstructs the connotation of SCSR and constructs an operable strategic intent—business synergy—resource investment—value output framework for identifying and analyzing SCSR processes. Importantly, this framework is specified as a system with explicit boundaries, a co-occurrence mechanism among its four dimensions, and a reinforcing feedback loop. As such, it advances systems theory in CSR research and provides a theoretically grounded analytical tool for understanding how firms transform CSR from fragmented activities into a self-reinforcing strategic capability. Second, at the methodological level, this study develops a hybrid text-mining framework to identify and compare strategic versus non-strategic CSR. PLMs first classify the two CSR text types, after which LDA topic modeling independently extracts the thematic distributions of each category. This design effectively circumvents the

fragmentation problem inherent in traditional LDA when applied to unlabeled raw texts, while preserving its strength in extracting macro-level thematic structures, providing a new technical pathway for objective, dynamic, and large-scale assessment of the quality of CSR strategies. Third, at the practical level, this study quantitatively reveals the lack of strategy in Chinese companies' CSR practices through quantitative characterization and contextual analysis of CSR practices in Chinese manufacturing enterprises. The findings can thus serve as evidence-based decision-making references for understanding the implementation mechanisms of CSR, for governments to optimize guidance mechanisms, for investors to identify high-quality CSR, and for companies to optimize strategic resource allocation.

The remainder of this paper is organized as follows. Section 2 reviews the theoretical literature on strategic CSR and text-mining methodologies, upon which we develop our analytical framework. Section 3 describes the research design, including data sources, variable measurement, and the hybrid text-mining approach. Section 4 presents the empirical results, covering the identification of SCSR, the LDA topic modeling analysis, the differentiated analysis of firm characteristics, and the economic and non-economic consequences of strategic CSR. Section 5 discusses the key findings, elaborates on their theoretical and practical implications, acknowledges the limitations, and outlines directions for future research.

2. Literature Review and Theoretical Framework

2.1. Theoretical Foundation of SCSR

SCSR is rooted in the theoretical framework of strategic management, with its core proposition concerning how companies can systematically transform CSR activities into sustainable competitive advantages. Companies do not passively respond to social demands but rather internalize them as a source of innovation, generating business returns while improving social welfare through developing new products, entering new markets, or optimizing operational processes [6,11]. Three theories offer complementary and progressive explanatory perspectives on SCSR.

First, stakeholder theory, which defines the relationship between corporations and diverse stakeholders, is a crucial theoretical foundation for CSR. This theory emphasizes that corporations should not only create value for shareholders but also assume responsibility for diverse stakeholders, including customers, employees, suppliers, governments, and communities [22,23]. However, traditional CSR concepts are often viewed as a one-way response or contribution, without considering the dynamic and evolving contractual relationship between corporations and society. Companies should proactively manage their relationships with key stakeholders and deeply integrate them into their mission and core business [2]. This means that CSR should not be viewed merely as a cost or charity but rather as a forward-looking strategic management tool that enables the co-creation of corporate and social value.

The resource-based view reveals the internal mechanisms by which SCSR creates value from the perspective of resource exchange and restructuring. The essence of SCSR lies in its ability to help companies accumulate scarce, difficult-to-imitate, and sustainable strategic resources, such as reputational capital, stakeholder trust, and green innovation capabilities [7,24]. Through these unique resources and capabilities, companies can build strong competitive barriers, thereby achieving long-term performance excellence [4,25].

The creating shared value framework integrates these perspectives to push SCSR to a more collaborative and interactive level of practice. Companies can move beyond unilateral actions and instead reshape CSR into strategic opportunities to create shared value through resource integration and interaction with key stakeholders such as customers and partners [9,26]. Companies can restructure their value chains by embedding social and environmental considerations into every stage from procurement and production to

marketing, thereby successfully integrating social and economic value in their core business areas [3,27]. This strategic integration not only reshapes the company's productivity base but also drives sustainable long-term performance growth, forming an evolutionary closed loop from responsible response to strategic co-creation.

2.2. Dimensional Identification of SCSR

SCSR is essentially a dynamic strategic management process aimed at creating shared value. Its operational logic is reflected in a complete chain from strategic awareness to value realization [4]. Companies proactively identify social needs, uncover potential business opportunities within them, and systematically invest financial, human, and innovative resources. Ultimately, this leads to the co-creation of economic and social value by providing solutions to social problems and developing emerging markets [11,28]. The key to the effective implementation of this process is the company's ability to deeply and systematically integrate CSR into the innovation process and operation of its core business, transforming it into a sustainable competitive advantage [9,25].

Scholars have gradually developed a multi-dimensional analysis framework to identify this dynamic process. Early studies, such as that by Burke and Logsdon [2], propose that SCSR can be identified through five key dimensions: centrality, specificity, proactiveness, voluntarism, and visibility. Husted and Allen [9] build upon this foundation, focusing on visibility, appropriability, and voluntarism and adding centrality and proactivity [10]. Camarena et al. [29] continue this framework, using cluster analysis to identify strong and weak SCSR. Centrality reflects the degree of alignment between a company's CSR activities and its core business, mission, and strategic objectives, and it is a key prerequisite for influencing corporate performance [3,25]. Appropriability refers to whether a company deeply integrates CSR with specific assets and obtains returns from its CSR investments [9]. Proactivity refers to a company's planning and implementation of CSR actions based on predictions of future social, market, or regulatory trends [30,31]. Voluntarism is the extent to which a company goes beyond legal, contractual, or direct business interest requirements [32]. Visibility is the degree to which a company's CSR activities can be effectively perceived and recognized by internal and external stakeholders, and it is an important link in achieving value creation [33].

Although these studies lay the foundation for an identification framework, prior research has two limitations in defining the dimensions. First, the boundaries between the definitions of each dimension are often unclear. For example, proactivity is often viewed as the opposite motivation of reactivity [34,35], which makes it difficult to distinguish it from voluntarism. In addition, the meaning of visibility varies depending on the research subject [29]. Second, studies often treat each characteristic as an independent variable and fail to deeply reveal the intrinsic connections and systemic synergistic mechanisms between them. For example, Husted and Allen [10] find that centrality, voluntarism, and visibility have significant effects, but appropriability and proactivity do not. Abdelkarim [36] points out that the effects of these characteristics may vary depending on the industry context. Third, the conceptual status of certain dimensions remains ambiguous. For instance, appropriability concerns whether a firm can capture returns from CSR investments, which is an outcome-side issue rather than an input-side perspective. Similarly, visibility concerns stakeholder awareness and disclosure quality, which is not equivalent to actual value creation. This indicates that research has not yet formed a unified and systematic identification framework, which prevents a comprehensive understanding of the dynamic mechanisms of SCSR. This provides the theoretical impetus for our proposed four-dimensional framework.

2.3. Theoretical Framework for SCSR Identification

Although the literature has not reached consensus on the definitions of dimensions, it systematically touches upon core logical aspects such as strategic intent (proactiveness, voluntarism), business synergy (centrality), resource investment (dedicated assets), and value output (competitive advantages). Based on a review of theoretical perspectives and empirical consensus, this study proposes a four-dimensional strategic intent—business synergy—resource investment—value output identification framework (as illustrated in Figure 1) to integrate these disparate characteristics, which are widely discussed but relatively fragmented in the literature. By highlighting the dynamic connection and synergistic effects of each dimension in the shared value creation process, this framework can provide a more coherent and explanatory theoretical basis for the identification and analysis of SCSR.

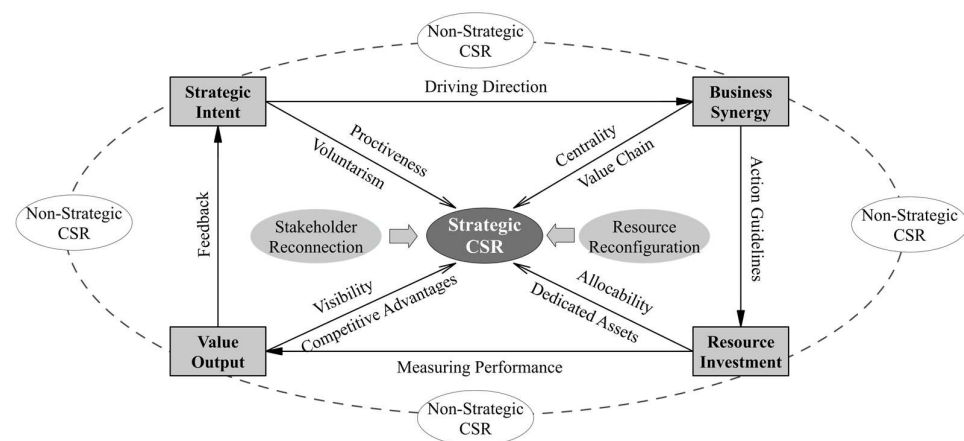


Figure 1. A four-dimensional framework for identifying strategic CSR.

The system boundary of this framework is defined as the firm's CSR decision-making and implementation process, encompassing strategic planning, operational integration, resource allocation, and performance feedback.

The strategic intent dimension emphasizes the voluntary and proactive nature of a company's fulfillment of CSR, relating to the strategic orientation and timing of actions. Voluntarism is reflected in proactive choices that go beyond compliance obligations [32], and proactiveness manifests in forward-looking planning based on future trends [30]. Together, these two aspects are strategic drivers of CSR practices. The business synergy dimension focuses on the deep integration of CSR with business operations, aiming to achieve strategic synergy and shared value creation [3]. Its core lies in the relevance of, and mutual reinforcement between, social responsibility issues and the company's main business.

The resource investment dimension reflects a company's substantive and sustained commitment of resources to CSR. By thoroughly integrating CSR with its proprietary assets (such as its technology, brand, and distribution channels), a company can transform CSR into a sustainable competitive advantage [4,28]. This is manifested in stable long-term financial, technological, and operational investments. This conceptualization extends the notion of appropriability from prior studies. In contrast to appropriability, which concerns capturing returns from CSR investments, the resource investment dimension captures the actual commitment of financial [9], human, and technological inputs, offering a more comprehensive basis for understanding resource commitment in CSR.

The value output dimension captures the dual value created by CSR activities. This includes both economic value (e.g., cost savings, revenue growth, operational efficiency) and social value (e.g., environmental improvement, community welfare, stakeholder well-being) [31]. Importantly, value output is operationalized through measurable and verifiable

evidence in the text, such as quantified emissions reductions, resource savings, or social impact metrics. This distinguishes value output from visibility: visibility concerns stakeholder awareness and recognition [2,9], which is an important precursor to value realization, whereas value output concerns the actual, demonstrable outcomes of CSR activities. The cumulative effect of sustained value output is the gradual development of competitive advantages, such as enhanced reputation, stakeholder trust, and innovation capacity [3].

The four dimensions play distinct logical roles in the SCSR system. Strategic intent serves as the processual driver, providing the motivational and directional force for CSR. Business synergy serves as the necessary condition, ensuring that CSR is integrated with core operations. Resource investment serves as the enabling condition, providing the material foundation for action. Value output serves as the outcome indicator, capturing measurable economic and social results.

The relationships among the four dimensions operate through two system mechanisms. First, the co-occurrence mechanism specifies that strategic intent, business synergy, resource investment, and value output must be jointly present within the same unit of analysis for a CSR practice to be identified as SCSR. The absence of any single condition means that the practice remains at the non-strategic level. Together, the four dimensions constitute a complete functional logic chain that progresses from motivational drivers to implementation and ultimately to outcomes. Crucially, this configuration reflects a logical complementarity among the dimensions, rather than a temporal sequence of causation.

Second, the reinforcing feedback mechanism enables SCSR to form a dynamic loop. Value output provides feedback through two pathways: reputational feedback, through which positive stakeholder recognition reinforces the firm's legitimacy to continue pursuing SCSR; and resource feedback, through which value creation attracts additional resources for reinvestment in further CSR activities. As this loop operates at the organizational level over time, responsive and fragmented CSR is transformed into a dynamic, system-level capability for sustainable value creation [33].

2.4. From Theoretical Framework to Empirical Operationalization

The four-dimensional framework provides a theoretical lens for deconstructing SCSR, yet its large-scale empirical application faces two challenges. First, the four dimensions are latent constructs embedded in textual disclosures and cannot be directly observed. Second, distinguishing strategic from non-strategic CSR requires interpreting deep strategic logic, which demands contextual semantic understanding beyond simple keyword matching.

Prior studies have relied on questionnaires, case studies, or dictionary-based matching [37,38], but these approaches are subjective and limited in external validity; dictionary methods are prone to false positives. Recently, pre-trained language models (PLMs) have been applied to extract key information from CSR texts [19,20], enabling the identification of SCSR components at the paragraph level and the separation of the two CSR types in an objective, replicable manner at scale.

PLMs alone, however, cannot capture the thematic structures that differentiate the two CSR types at the macro level. To address this, we introduce LDA topic modeling to extract and compare thematic distributions across the two corpora in an unsupervised, data-driven way [39]. The integration of PLMs and LDA enables examination of the SCSR system across multiple levels. PLMs perform theory-driven semantic classification at the paragraph level, from which firm-level strategic integration is aggregated at the report level. LDA uncovers thematic divergences at the behavioral level, revealing the evolutionary trajectory of the overall SCSR system at the firm level, thereby extending the methodological boundaries of strategic CSR research.

3. Materials and Methods

Based on the theoretical framework above, this study proposes a hybrid methodology that integrates pretrained transformer-based language models with LDA topic modeling. This method follows the three-stage analytical logic of theoretical framework construction, automatic semantic classification, and topic difference mining. The approach combines longitudinal strategic process tracing with horizontal content structure analysis to provide an integrated research perspective.

First, the texts of CSR reports were structured and manually annotated based on the four-dimensional framework to construct a training corpus. Second, pretrained transformer encoders (BERT, RoBERTa, and XLNet) with multi-task learning algorithms were used to recognize the performance of textual data across the four dimensions and vertically rebuild the entire CSR process from strategy conception to value realization. Building on this, LDA topic modeling was used to perform lateral clustering analysis of the classified strategic and non-strategic CSR texts by extracting high-frequency topics and keywords. This systematically revealed the content differences between the two types of CSR in dimensions such as operations and innovation and enabled their internal structure to be deconstructed from a semantic distribution perspective. Through the complementary functions of PLMs and LDA, this method achieved a dual analysis from strategic path to topic structure that avoided methodological overlap and enhanced the explanatory depth of the research, thus providing methodological support for subsequent empirical testing. The overall research process is shown in Figure 2.

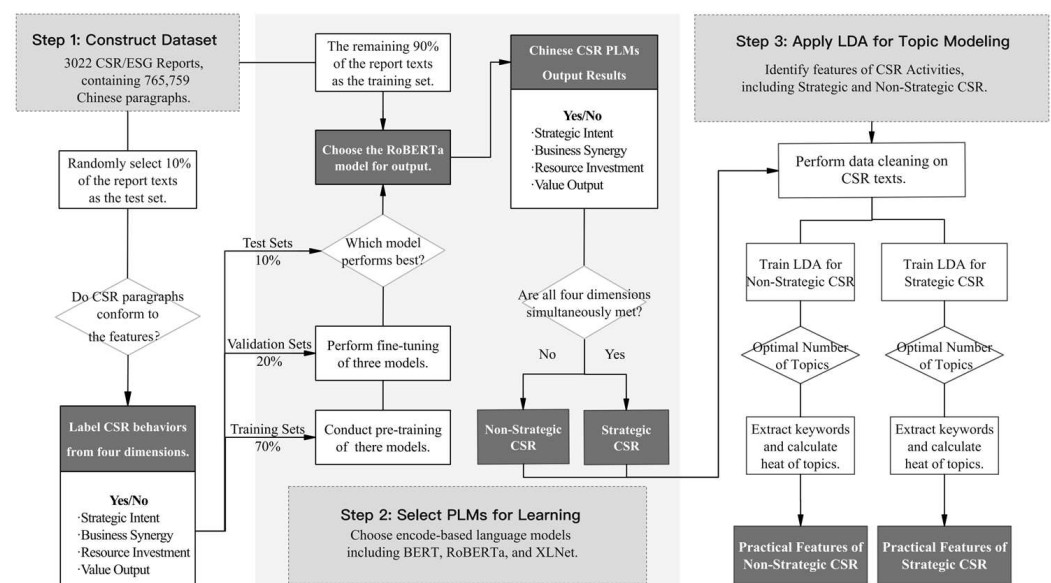


Figure 2. Research methodology and process.

3.1. Data Sources and Processing

The data used in this study for listed companies are from the China Stock Market & Accounting Research Database (CSMAR). In addition, CSR, ESG, and sustainability reports from A-share listed manufacturing companies in Shanghai and Shenzhen from 2021 to 2023 were manually collected from the CNINFO website as a text dataset, which covers a large amount of information on CSR practices and plans. To mitigate the influence of extreme samples, reports from firms labeled as ST or ST* (special treatment) were excluded. Reports from firms that were delisted or underwent backdoor listings during the research period were also removed.

We applied the following text preprocessing and cleaning to the dataset. First, the PDF-formatted CSR reports were converted into editable and processable plain text, retaining the original paragraphs and line breaks. Noise unrelated to the report text, including garbled characters, URL links, and uncommon symbols, was removed. Then, a two-stage segmentation method based on format and semantics was used. To preserve the sentence structure and semantic information integrity of the original documents, line breaks in the original text were identified as primary paragraph delimiters, and the text was initially segmented into multiple paragraphs. To ensure that each paragraph unit had an appropriate length, initial paragraphs exceeding 400 characters were optimized using a Chinese natural language processing toolkit in Python. Finally, reports from different companies and years were randomly sampled to observe the reasonableness of paragraph distribution after cleaning. This study thus obtained a dataset of CSR reports from 2021 to 2023, with a release period from 1 January 2022 to 31 December 2024, covering 3022 CSR or ESG reports and totaling 765,759 Chinese paragraphs.

3.2. Text Analysis Based on Pre-Trained Language Models

3.2.1. Annotation Rules and Examples

To enhance the representativeness and generalizability of the model, this study employed a stratified random sampling method to ensure that the selected sample covered manufacturing firms across various years and of different sizes. A subset comprising 10% of the total CSR reports (i.e., 302 reports) was extracted as the annotation pool, and the logical paragraphs generated during the preprocessing stage served as the fundamental coding and analytical units.

The annotation work was conducted in three sequential phases. In the calibration phase, three doctoral students from the research team jointly annotated 10 reports through multiple rounds of group discussion. This process iteratively refined the codebook and resolved definitional ambiguities. In the inter-annotator reliability phase, the same three annotators independently coded another 20 reports without discussion. For each of the four binary dimensions, we calculated Fleiss' Kappa coefficients, which were 0.647, 0.619, 0.680, and 0.711, respectively. According to the benchmarks of Landis and Koch [40], these results indicated substantial agreement. When disagreements arose among the three annotators in this phase, a majority voting rule was applied (two out of three agreement). In cases where all three annotators disagreed (three-way split), a fourth arbitrator made the final determination based on the codebook. In the scalable annotation phase, after confirming that reliability metrics met acceptable thresholds, the remaining reports were divided among the three annotators for single-annotation in a division-of-labor manner. To ensure sustained quality stability during this phase, a random audit mechanism was implemented, whereby a second annotator independently re-annotated 10% of the completed reports on a weekly basis. If the periodic audit agreement fell below 0.80, the entire batch was returned for group discussion and re-annotation.

The four binary label variables correspond respectively to strategic intent, business synergy, resource investment, and value output—dimensions that are conceptualized and operationalized within the theoretical framework established in Section 2.3, which synthesizes prior dimensional identification studies, including those of Burke and Logsdon [2] and Husted and Allen [9,10]. Integrating insights from stakeholder theory, the resource-based view, and the creating shared value perspective, this framework constitutes a complete analytical chain tracing the strategic process from motivational drivers through implementation to performance feedback.

Based on this framework, we establish a conjunctive rule for operational annotation. A paragraph must be coded “Yes” on all four dimensions to be classified as a “strategic CSR”

paragraph; if any dimension is coded “No,” the paragraph is excluded from the strategic CSR category. It should be noted that the “Yes” examples under each dimension in Table 1 are provided solely to illustrate the judgment criteria for that individual dimension and do not imply that the paragraph as a whole has been classified as strategic CSR. Specific annotation examples are shown in Table 1.

Table 1. Examples of strategic CSR annotation by dimension.

Dimension	Judgment Criteria	Examples Annotated “Yes”	Examples Annotated “No”
Strategic Intent	Whether the firm demonstrates proactive, voluntary, and forward-looking intent regarding CSR initiatives.	A semiconductor company: Leveraging core strengths in smart chips to promote the deployment of 5G-enabled telemedicine and smart education infrastructure in underserved rural areas, thereby bridging the urban-rural digital divide through technological innovation.	A company: Strictly complies with laws and regulations including the Company Law of the People’s Republic of China and the Law of the People’s Republic of China on the State-Owned Assets of Enterprises. . .
Business Synergy	Whether the chosen topics of CSR initiatives are highly relevant to the firm’s core business.	A photovoltaic manufacturer: Develops and produces clean, efficient PV modules and constructs integrated clean energy power generation and storage systems to provide sustainable electricity to areas with no or weak grid access.	A food production company: Donates books and sports equipment to children in remote areas. Donated 500,000 RMB to the pandemic relief fund for purchasing COVID-19 vaccines.
Resource Investment	Whether the firm commits specific, stable, and long-term resources to its CSR initiatives.	- A company: Has established a dedicated CSR foundation, annually allocating 3% of its net profit to targeted community education and ecological restoration programs, with permanent implementation teams in multiple regions ensuring sustained resource commitment. - A company: Launched a five-year “Green Supply Chain Partner Program,” investing 15 million RMB annually to help upstream suppliers adopt cleaner production technologies and reduce carbon emissions across the value chain.	- An enterprise: We always believe that social responsibility should be closely integrated with the company’s development strategy and daily operations. (Remains at the level of intent and planning without concrete action.) - An enterprise: Donated 500,000 RMB to the pandemic relief fund for purchasing COVID-19 vaccines. (A one-time, large cash donation lacking characteristics of sustained resource allocation.)
Value Output	Whether the CSR initiative is well-designed, measurable, disclosable, and capable of generating a virtuous cycle.	A food production company’s tracking and evaluation of a biogas recovery project: Recovered and utilized 17.1 million standard cubic meters of biogas throughout the year, equivalent to saving 12,000 tons of standard coal, reducing CO₂ emissions by 31,000 tons. Biogas consumption accounted for 20.03% of total energy use.	An enterprise: Establishes a comprehensive career development platform for employees, provides a safe working environment, fully safeguards their physical and mental health, and creates a happy workplace. (Uses only qualitative descriptions without any concrete data support.)

While Table 1 demonstrates how each dimension is coded individually, the sequential identification logic requires that all four dimensions be considered jointly. An example illustrating this joint application is as follows: “The company has integrated sustainable development into its core strategy and plans to invest 500 million RMB over the next three years in the R&D and production of new-energy batteries. This initiative builds directly on our existing automotive manufacturing expertise. A dedicated R&D center and permanent project team have been established, and the project is expected to reduce CO₂ emissions by approximately 80,000 tons annually.” This paragraph fulfills all four criteria and is consequently classified as strategic CSR.

After annotation, all of the labeled texts were randomly partitioned into training, validation, and test sets in a 7:2:1 ratio. To prevent data leakage and evaluation bias, the split was performed at the firm-report level, ensuring that all paragraphs from the same corporate report appeared exclusively in one dataset. This precaution avoided artificially inflating the model’s performance due to it learning the writing style of any single report. The complete codebook, including operational definitions, inclusion and exclusion criteria, disagreement resolution rules, and annotator qualifications, is provided in the Supplementary Materials.

3.2.2. Model Selection and Training

To ensure computational feasibility and timeliness, this study constructed a specialized classifier by fine-tuning three pretrained transformer encoders (BERT, RoBERTa, and XLNet) on a labeled corpus of CSR paragraphs. Influenced by international standards and the requirements of the China Securities Regulatory Commission, the wording and writing style of CSR reports differ significantly from those of other corporate disclosures, making generic pre-trained models less directly applicable without task-specific adaptation. To ensure optimal model suitability, these three widely used and representative encoder-based language models were selected for comparative experiments.

BERT (Bidirectional Encoder Representations from Transformers), which was trained through a masked language model and context prediction tasks, provides a strong foundation for bidirectional contextual understanding by capturing semantic information and syntactic features in the text [41]. RoBERTa enhances the training strategy by increasing the training data and time to fully exploit the potential of the original BERT model [42]. XLNet (Generalized Autoregressive Pretraining for Language Understanding) is a permutation language modeling mechanism that captures more flexible contextual dependencies [43]. It is important to clarify that these three models are encoder-based discriminative language models optimized for language understanding tasks, which are distinct from generative large language models such as GPT.

In terms of model architecture, this study adopted a multi-task learning framework, using a shared encoder to process the four classification tasks in parallel. This design was mainly based on two considerations. First, the four dimensions in the theoretical framework are logically closely related. Multi-task learning can capture the inherent connections between the dimensions through parameter sharing, which can greatly reduce the risk of overfitting and improve the efficiency of the pre-trained model [44]. Second, the unified framework ensures the comparability of performance between different base models. Specifically, in each independent experiment, the input text was encoded by the same pre-trained language model encoder (BERT, RoBERTa, or XLNet). The output shared feature vectors were input into four fixed independent linear classification layers corresponding to the four dimensions of strategic intent, business collaboration, resource investment, and value output. The final identification of SCSR was determined and integrated by a logical “AND” rule applied across these four outputs.

3.2.3. Model Implementation and Hyperparameter

We employed three pre-trained language models with their exact Hugging Face identifiers: BERT-base-chinese (bert-base-chinese), Chinese RoBERTa-wwm-ext (hfl/chinese-roberta-wwm-ext), and XLNet-base-cased (xlnet-base-cased). All models were obtained from the Hugging Face Transformers library (version 5.13.1) and implemented using PyTorch 2.11.0 with CUDA 12.1. The input sequence was truncated or padded to a maximum length of 128 tokens, with a per-device batch size of 32, and fine-tuned for 3 epochs. The optimizer was AdamW with a learning rate of 2×10^{-5} , weight decay of 0.01, and 1000 warmup steps. The four classification tasks used binary cross-entropy loss, summed with equal task weights (1.0 for each task) to form the total training objective. The classification threshold was set to the default argmax (equivalent to 0.5). We did not apply class weighting or resampling; instead, we report precision, recall, F1-score, and AUC alongside accuracy at the evaluation stage to ensure transparent assessment across both majority and minority classes. Mixed-precision (fp16) training was enabled to accelerate computation and reduce memory usage. All models were fine-tuned on a single NVIDIA A100-SXM4-40 GB GPU with 40 GB memory, running on Google Colab Pro+.

To ensure full reproducibility, we set random seeds at multiple levels: Python, NumPy, PyTorch, and CUDA were all set to 42, and the entire training procedure was repeated with five different random seeds (42, 123, 456, 789, 2024). We report the mean, standard deviation, and 95% bootstrap confidence intervals ($B = 1000$) for all metrics across these five runs.

As a robustness check, we additionally trained three conventional classifiers on TF-IDF features as baselines: logistic regression (LR), linear support vector machine (SVM), and random forest (RF). The TF-IDF vectorizer was configured with `max_features = 10,000` and `ngram_range = (1, 2)`. RoBERTa substantially outperformed all baselines in macro-average F1 (0.936 vs. 0.860 for LR, 0.859 for SVM, and 0.847 for RF), confirming the superior performance of pre-trained language models for this task. Detailed baseline results are provided in Appendix A Table A1.

3.2.4. Performance Evaluation and Comparison

After training, we conducted a comprehensive validity assessment of the model on a completely independent test set. This study plotted the receiver operating characteristic (ROC) curve and calculated the area under the curve (AUC). As shown in Figure 3a, the BERT and RoBERTa models exhibited excellent discriminative performance, with their ROC curves closely aligned with the upper left corner and their AUC values reaching 0.94. In contrast, the XLNet model had limited discriminative ability, with an AUC value of 0.63.

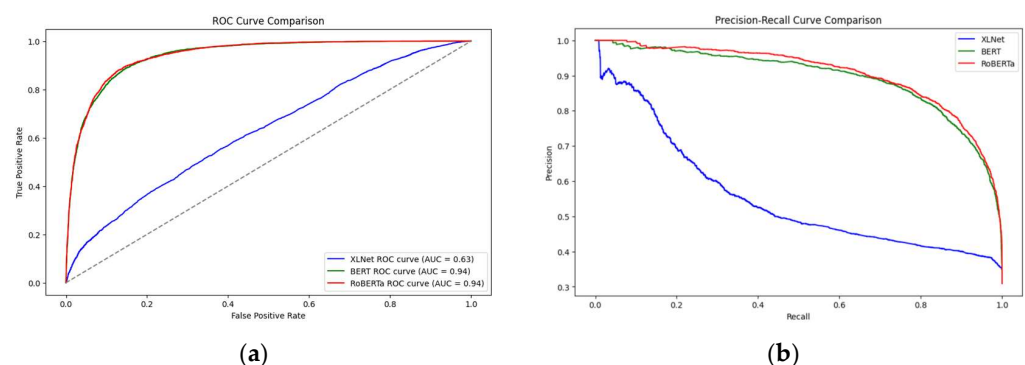


Figure 3. Performance evaluation of pre-trained models on the SCSR classification task. (a) Receiver operating characteristic (ROC) curves of the BERT, RoBERTa, and XLNet models, with the area under the curve (AUC) values annotated for each model; (b) Precision–recall (PR) curves of the three models, illustrating the trade-off between precision and recall across classification thresholds.

To further examine the model's ability to identify difficult samples and minority classes, this study plotted the precision–recall curve (PR curve). The closer the PR curve is to the upper right corner, the better the model's performance in balancing precision and recall. As shown in Figure 3b, RoBERTa's PR curve was closest to the upper right corner, indicating that it achieved higher recall while maintaining high precision and resulting in the optimal overall performance.

Table 2 reports the classification performance for each of the four dimensions across all three pre-trained language models. Among them, RoBERTa consistently achieves the highest performance across all dimensions, which justifies its selection as the primary model for subsequent analyses. The narrow 95% bootstrap confidence intervals for F1-scores (reported in Appendix A Table A2) confirm the stability of these performance estimates across different random seeds.

Table 2. Performance of pre-trained language models across four strategic CSR dimensions.

Models	Dimension	Accuracy	Precision	Recall	F1-Score
BERT	Strategic Intent	0.9371	0.9529	0.9172	0.9346
	Business Synergy	0.9466	0.9187	0.8842	0.9010
	Resource Investment	0.9582	0.9698	0.9782	0.9739
	Value Output	0.9337	0.9415	0.9085	0.9247
RoBERTa	Strategic Intent	0.9388	0.9649	0.9083	0.9357
	Business Synergy	0.9493	0.9319	0.8798	0.9051
	Resource Investment	0.9608	0.9724	0.9787	0.9756
	Value Output	0.9379	0.9549	0.9040	0.9287
XLNET	Strategic Intent	0.7457	0.7120	0.8101	0.7578
	Business Synergy	0.7750	0.6244	0.4563	0.5271
	Resource Investment	0.8760	0.8958	0.9561	0.9250
	Value Output	0.7597	0.7065	0.7938	0.7473

Based on the overall evaluation results for the ROC curves, PR curves, and classification metrics, the RoBERTa model demonstrated optimal and stable performance in overall discriminative performance, adaptability to imbalanced data, and key classification metrics. Therefore, this study selected RoBERTa as the base encoder for the multi-task learning framework, and it was used for subsequent text recognition and topic analysis tasks.

Disclosure on Generative AI Usage: No generative artificial intelligence (e.g., GPT series, ChatGPT, Claude, or similar models) was used for text generation, data analysis, data interpretation, or any other substantive aspect of this study. The pre-trained language models employed (BERT, RoBERTa, and XLNet) are discriminative models used exclusively for supervised text classification. All classification outputs were reviewed and verified by the authors, who take full responsibility for the accuracy and validity of the results presented in this manuscript.

3.3. Topic Analysis Based on Latent Dirichlet Allocation

Building on the classification results from the RoBERTa model, this study applied LDA topic modeling separately to each corpus to identify and compare the topics between strategic and non-strategic CSR, revealing their different value creation logics and practical characteristics. LDA is a generative probabilistic model that assumes each document is a mixture of latent topics, and each topic is a distribution over words [45]. This study used LDA to more efficiently identify prominent topics under different CSR categories. Following the approach of Li and Lemke [46], LDA was used to analyze different types of CSR behaviors.

The process involves several key steps. First, data cleaning and text preprocessing were performed, including Chinese character filtering, word segmentation using the jieba tokenizer, filtering of invalid characters and useless information, and removal of Chinese stop words. Open-source stop word lists such as “Chinese Stop Word Library,” “Harbin Institute of Technology Stop Word List,” “Sichuan University Stop Word List,” and “Baidu Stop Word List” were merged and processed. Over 100 commonly used stop words from CSR reports were added, resulting in a list of 2443 stop words for further filtering. Second, a bag-of-words model was constructed. To optimize term frequencies and improve modeling quality, we constructed document-term matrices with TF-IDF weighting as a preprocessing step to down-weight terms that are frequent across documents, thereby enhancing the discriminative power of topic identification [37]. The parameters were set to $\text{max_df} = 0.95$ (ignoring terms appearing in more than 95% of documents) and $\text{min_df} = 2$ (ignoring terms appearing in fewer than 2 documents). It is important to clarify that this TF-IDF weighting is a preprocessing step applied before model estimation; the underlying LDA algorithm itself remains a generative probabilistic model based on word-count distributions. The formula is as follows:

$$TF - IDF_{i,j} = TF_{i,j} \times IDF_i, \quad (1)$$

where $TF_{i,j} = \frac{f_{i,j}}{\sum_k f_{k,j}}$, is the frequency of word i in document j , $IDF_i = \log\left(\frac{N}{1+n_i}\right)$, is the inverse document frequency, N is the total number of documents, and n_i is the number of documents containing word i .

The LDA model was implemented using the LatentDirichletAllocation module from the scikit-learn library (version 1.6.1) in Python. For each corpus, the model was estimated using the online variational Bayes algorithm. Prior to model estimation, the optimal number of topics K for each corpus was determined via a systematic grid search over $K \in [2, 20]$ using PMI coherence and perplexity as evaluation metrics. The procedure yielded $K = 7$ for the strategic CSR corpus (peak coherence: 0.690) and $K = 12$ for the non-strategic CSR corpus (peak coherence: 0.747) (reported in Appendix A Table A3). With the optimal K values determined, model estimation was then conducted. The key hyperparameters were set as follows: $\text{learning_offset} = 10.0$ (controlling learning rate decay in early iterations), $\text{batch_size} = 128$ (number of documents per update), and $\text{max_iter} = 100$ (maximum iterations). To ensure stability, the random seed was set to $\text{random_state} = 42$ for reproducibility, and the model was run with five different random seeds to verify consistency across runs.

After model estimation, LDA was applied to extract keywords and identify topics. Since separate LDA models were fitted to the strategic and non-strategic corpora, the topic spaces of the two models are not directly aligned, and thus the document-topic posterior probabilities for each topic are not directly comparable across the two models. To address the non-comparability of topic spaces across the two models, this study adopted a three-stage approach to compute topic weights and determine topic labels. In the first stage, LDA automatically identified the keyword set for each topic through word co-occurrence patterns. In the second stage, based on the automatically discovered keyword sets, we counted their frequency of occurrence within each document and computed the topic weights. In the third stage, topic labels were determined by three independent researchers who reviewed the top 20 high-frequency keywords for each topic, discussed discrepancies, and reached consensus on topic names. This procedure ensured the reliability of manual topic naming.

The logic of this approach is that LDA automatically identifies semantically coherent thematic structures through word co-occurrence patterns, and keyword frequency counts

based on these structures provide a robust, interpretable measure of thematic prevalence that aligns with the theoretical framework [46]. The calculation formula is as follows:

$$R_{d,z} = \frac{\sum_{w \in K_z} f_{w,d}}{\sum_{w \in V} f_{w,d}}, \quad (2)$$

where $R_{d,z}$ represents the weight of topic z in document d , K_z is the set of keywords for z , $f_{w,d}$ is the frequency of word w in document d , and V is all words in the vocabulary. To eliminate the influence of redundant vocabulary, the topic prevalences for strategic and non-strategic CSR were normalized separately.

This metric focuses on reflecting the extent of explicit linguistic coverage of a given topic within the text. Through normalization, the topic weights across the two corpora are presented on the same scale, effectively capturing the relative importance of different topics within each corpus. It should be noted that the thematic contrasts reported in this study focus on qualitative comparisons, i.e., the dominant topic types within each corpus, rather than direct numerical comparisons across models, thereby avoiding potential bias while revealing the thematic structural differences between the two types of CSR behaviors.

4. Results

4.1. Identification of SCSR and Dimensional Characteristics Based on PLMs

Based on the identification results of the multi-task RoBERTa model, this study reveals the overall strategic level of CSR practices among Chinese firms. As shown in Figure 4, SCSR practices that satisfy all four dimensions account for only 25.6% of the entire sample, whereas non-strategic practices comprise a substantial 74.4%. This indicates that the CSR practices of listed companies in China are still generally focused on responsiveness and compliance and have not yet been widely transformed into strategic business activities that drive long-term competitive advantage.

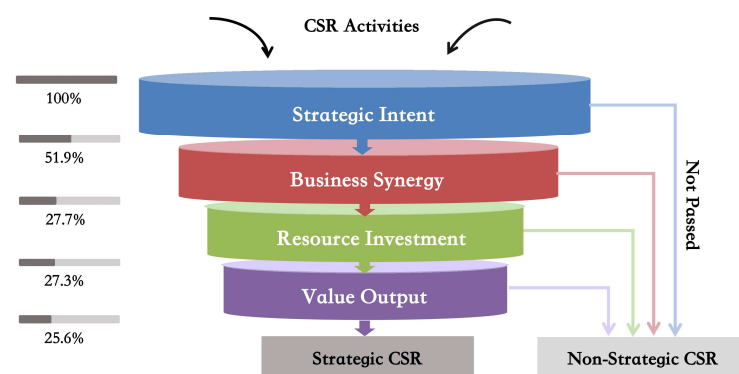


Figure 4. The proportion of strategic CSR classified by dimensions.

Further analysis reveals substantial imbalance across the four dimensions of SCSR. As shown in Figure 5a, practices with explicit resource investment account for the highest proportion (86.8%), while those achieving deep business synergy have the lowest proportion (only 29.6%), with strategic intent (51.9%) and value output (56.6%) falling in between. The considerably higher proportion of strategic intent than that of business synergy (51.9% vs. 29.6%) indicates that many strategic plans remain at the vision level and that implementation paths have not yet been deeply integrated with corporate operations. In addition, practices with resource investment and value output have a relatively high proportion (86.8% vs. 56.6%), further demonstrating that current CSR practices largely rely on resource-driven rather than strategy-integrated approaches.

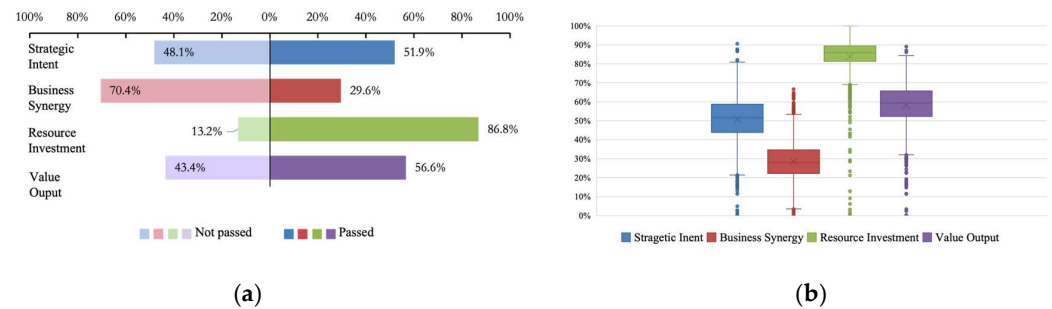


Figure 5. Identification of the four dimensions of strategic CSR. (a) Proportion of CSR practices identified across each of the four dimensions; (b) Boxplots of the distribution of paragraph-level identification ratios across the four strategic CSR dimensions.

Figure 5b presents boxplots illustrating the distribution of paragraph-level identification ratios across the four dimensions for each firm-year report. The distribution reveals that the median identification ratio for resource investment is the highest, at slightly above 85%, with a concentrated distribution and an evident upper-tail elongation. The median ratios for value output (approximately 60%) and strategic intent (slightly above 50%) are moderate, with both distributions exhibiting right-skewed patterns. By contrast, the median ratio for business synergy falls below 30%, with a left-skewed distribution.

A critical caveat should be noted regarding the interpretation of these classification results. Our classifier, trained on CSR report texts, may capture not only genuine strategic CSR attributes but also firms' reporting quality and impression management. Even though the annotation scheme explicitly directs annotators to assess the presence of specific strategic attributes rather than linguistic polish, the possibility of reporting-quality confounding remains. To mitigate this concern, we adopted a 50% threshold for aggregating paragraph-level classifications into report-level measures: a report was classified as exhibiting a given dimension if more than 50% of its paragraphs were positively identified for that dimension, following the majority-voting principle commonly used in text classification studies. This conservative criterion reduces the influence of isolated well-written passages that may reflect impression management rather than genuine strategic orientation.

Applying this rule, among the 3022 annual CSR reports, 98.74% (2984 reports) exceeded the threshold for resource investment, 80.09% (2420 reports) for value output, and 56.36% (1703 reports) for strategic intent. In stark contrast, only 2.3% (69 reports) exceeded the threshold for business synergy. These results clearly demonstrate that while most firms demonstrate willingness and resource commitment in fulfilling their social responsibilities, there remains a significant shortfall in their capacity to systematically and structurally integrate social issues with their core business strategies. To assess the sensitivity of our results to the choice of the unit of analysis, we replicated the SCSR identification procedure at three aggregation levels: paragraph, report, and firm-year.

Subsequently, this study conducted a cross-analysis of the four dimensions to explore their inherent connections and combination patterns. As shown in Figure 6, in practices lacking strategic intent, although most (74.4%) still involved resource investment, the proportion that generated value output was extremely low (24.4%). In contrast, in practices with strategic intent, the proportions of both resource investment and value output were significantly higher (98.4% and 86.5%, respectively). This indicates that proactive strategic planning is a key to transforming resources into actual value.

From the perspective of business synergy, there is little difference between synergistic and non-synergistic practices in terms of resource investment and value output, but a significant divergence exists in strategic intent. Specifically, 93.3% of synergistic practices had strategic intent, while this proportion was only 34.6% in non-synergistic practices.

This indicates that the realization of business synergy greatly depends on forward-looking strategic planning rather than spontaneous formation.

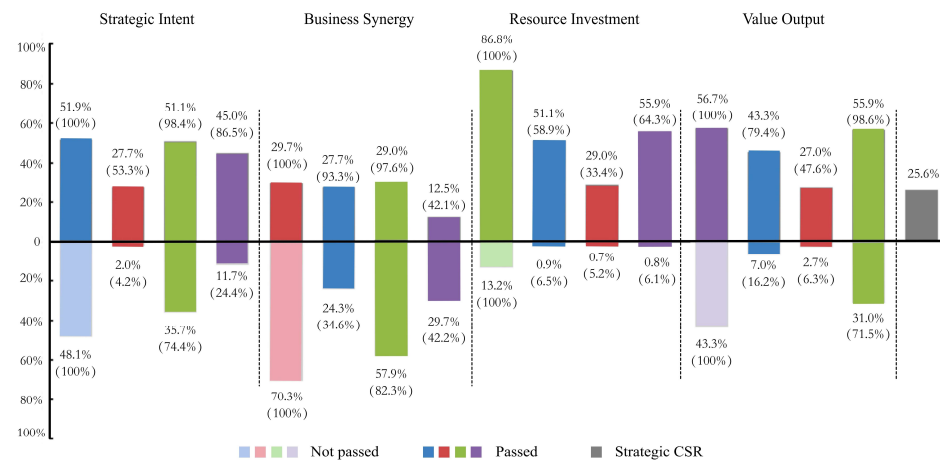


Figure 6. Intersection analysis for four dimensions of strategic CSR.

Regarding resource investment, the vast majority of practices (86.8%) included explicit resource commitments, and when resource investment was lacking, the realization rates of other dimensions dropped to extremely low levels (approximately 6%), indicating that resource investment is a necessary foundation for SCSR practices. However, where resources were invested, the proportion of business synergy remained low.

From the perspective of value output, the proportion of practices that failed to generate value with strategic intent was very low (16.2%), whereas this proportion was 79.4% for practices that did generate value, indicating that value creation depends on the guidance of strategic intent. A contradiction lies in the fact that for practices that generated value, the proportion of resource investment (98.58%) was far higher than the proportion of business synergy (47.6%), meaning that large amounts of resources were allocated to superficial actions with low relevance to the core strategy and corporate operations, such as achieving carbon reduction targets by purchasing carbon sinks rather than technological upgrades.

To assess the sensitivity of our results to the choice of the unit of analysis, we replicated the SCSR identification procedure at three aggregation levels: paragraph, report, and firm-year. At the paragraph level, the AND rule (all four dimensions present) and the OR rule (any single dimension present) serve as a conservative lower bound and a lenient upper bound, respectively. At the report level, paragraph-level classifications were aggregated by counting the number of dimensions present within each report (≥ 2 , ≥ 3 , or $=4$ dimensions). At the firm-year level, we computed the mean score across the four dimensions and applied thresholds of 0.3 and 0.5. The results, reported in Appendix A Table A4, confirm that the “AND” rule serves as a conservative lower bound, while alternative aggregation rules capture progressively broader manifestations of SCSR-related discourse.

4.2. Topic Modeling Analysis of CSR Practices Based on LDA

Following the identification of the two CSR behavior types, this study applied LDA to extract their thematic structures. Table 3 presents the identification results, showing seven topics categorized into four modules. This complementary LDA-based analysis extracts content rather than stylistic features, providing a more substantive basis for distinguishing strategic from non-strategic CSR.

Operation and Supply Chain Management (0.511) had a relatively dominant position, indicating that enterprises mainly achieve strategic synergy by deeply embedding the concept of responsibility into the value chain. Specifically, Supply Chain Relationship

Management (0.260) was frequently identified, with keywords such as “supplier,” “customer,” “quality,” and “supply chain” indicating that enterprises optimize upstream and downstream collaboration and enhance overall resilience through CSR practices. Green Operations and Energy Efficiency Improvement (0.197) concerns the sustainable transformation of production and operation links, with keywords such as “production,” “energy,” “energy saving,” and “reduction,” reflecting that enterprises promote the unity of environmental and economic benefits through technological transformation and process optimization. Technology Upgrading and Transformation (0.053), by contrast, captures a more specific aspect of this transformation, with keywords such as “water use,” “electricity use,” “pollutant,” “high temperature,” “welding,” and “reagent,” terms that reflect process-level retrofitting, heat recovery, and cleaner production measures in energy-intensive manufacturing sectors.

Table 3. LDA Topic Modeling results for strategic CSR.

Thematic Module	Weight	Topic	Keywords
Operation and Supply Chain Management (0.511)	0.260	Supply Chain Relationship Management	supplier, quality, customer, development, management, continuous, product, supply chain, procurement
	0.053	Technology Upgrading and Transformation	water use, sequencing, pollutant, lead to, electricity use, aviation, maritime, distributed, early warning, high temperature, welding, reagent
	0.197	Green Operations and Energy Efficiency	production, energy, green, emissions, energy saving, reduction, utilization, equipment, reduction, environmental protection, process
Innovation and Intellectual Property (0.258)	0.183	Technology Exploration and R&D	technology, R&D, intelligent, manufacturing, innovation, industry, new energy, center, global, automotive, materials
	0.075	IP Creation and Protection	patent, invention patent, licenses, authorization
Talent and Organizational Development (0.178)	0.178	Employee Skills and Career Development	employee, training, talent, development, cultivation, learning, skill, career, professional, competency, position
Social and Community Investment (0.054)	0.054	Rural Revitalization and Philanthropy	rural, revitalization, assistance, pandemic, poverty alleviation, poverty relief, agriculture, cultivation, green ecology, road

Innovation and Intellectual Property (0.258) emphasizes internalizing CSR into core technological competitiveness through innovation. Among these topics, Technology Exploration and R&D (0.183) was the most common, with keywords such as “technology,” “R&D,” “intelligence,” “new energy,” and “innovation,” showing a clear tendency to focus on cutting-edge fields and drive technological breakthroughs through open collaboration. Intellectual Property Creation and Protection (0.075), a supplementary topic, includes keywords such as “patents,” “invention patents,” and “licenses,” emphasizing the importance of intellectual property for building competitive barriers through corporate innovation achievements.

Talent and Organizational Development (0.178) is focused on employee skills and career development, with keywords including “employees,” “training,” “talent,” and “skills.” It goes beyond basic welfare guarantees and focuses on strategic investment in human capital to promote employee resources into a sustainable source of innovation and intangible assets.

Social and Community Investment (0.054) is manifested in rural revitalization and public welfare, with keywords including “assistance,” “poverty alleviation,” and “green ecology.” It seeks to combine corporate resources with major social needs by exploring ways to open up new markets or resource channels for business development while fulfilling social commitments and building a broader symbiotic ecosystem.

Table 4 shows the topic modeling identification results for non-strategic CSR. Eleven topics were identified, which were categorized into four modules. Compliance Disclosure and Governance (0.480) was the most prominent module, reflecting companies’ passive response to and compliance with regulatory requirements and general principles and exhibiting strong internal governance characteristics. The most frequent of these topics, CSR Philosophy Statement (0.154), broadly involves comprehensive concepts such as “environment” and “governance.” Risk and System Management (0.145) focuses on operational risks such as “production” and “occupational health.” Corporate Governance and Auditing (0.130) concentrates on governance structures such as “board of directors” and “disclosure.” Customer Communication and Feedback (0.051) had a much lower frequency.

Table 4. LDA Topic Modeling results for non-strategic CSR.

Thematic Module	Weight	Topic	Keywords
Compliance Disclosure and Governance (0.480)	0.154	CSR Philosophy Statement	society, development, responsibility, sustainable, environment, interest, issue, governance, green, innovation, management, philosophy
	0.145	Risk and System Management	risk, supplier, production, management, training, emergency response, quality, occupational, health, management system, environment, information security
	0.130	Corporate Governance and Auditing	board of directors, committee, disclosure, information, director, internal, governance, shareholders’ meeting, compliance, audit, board of supervisors
	0.051	Customer Communication and Feedback	customer, reporting, complaint, satisfaction, feedback, survey, confidentiality, consumer, service, user, whistleblower, channel
Environmental and Production Compliance (0.162)	0.092	Pollution Emission and Monitoring	emission, waste, environmental protection, exhaust gas, gas, wastewater, energy saving, energy, greenhouse, reduction, pollutant, monitoring
	0.035	Hazardous Substance Testing and Handling	water use, vehicle, hazardous waste, danger, supplier, local, air pollution, storage, nitrogen oxides, concentration
	0.035	Compliant Operations and Disclosure	electrical, calculation, general, mineral, conflict, emissions, poor, food safety, responsibility statement
Basic Safeguards and Philanthropic Response (0.173)	0.094	Employee Rights and Benefits	employee, training, compensation, staff, welfare, labor, provide, talent, safeguard, recruitment, female, health
	0.079	Social Care and Charity	pandemic, Party building, donation, education, employee, assistance, organization, culture, theme, public welfare, rural, charity, revitalization
Investor Communication and Information Disclosure (0.185)	0.073	Investor Relations and Performance Statements	investor, shareholder, performance, proposal, interaction, dividend, relationship, communication, website, cash, announcement, briefing session
	0.072	Routine Operations and Data Disclosure	carbon dioxide, time, annual, revenue, operation, authorization
	0.040	Compliance in Information Statements	false, statement, misleading, record, omission, authenticity, content, insurance, accuracy, completeness, stakeholder

Environmental and Production Compliance (0.162) focuses on the environment, with the main purpose of fulfilling basic environmental commitments and preventing penalty risks. Among these topics, Pollution Emission and Monitoring (0.092) was most common, with keywords such as “exhaust gas” and “wastewater” directly pointing to the post-discharge treatment and monitoring of emissions and waste. Hazardous Waste Treatment and Handling (0.035) involves deep-level risk management in the supply chain. Such practices exhibit a clear tendency toward post-event response and end-of-pipe treatment rather than proactive management and prevention.

Basic Safeguards and Philanthropic Response (0.173) involves responding to external demands and providing routine benefits. These practices are loosely linked to a company’s core strategy and business operations. Employee Rights and Benefits (0.094) primarily involve basic welfare and security, while Social Care and Charity (0.079) are highly tied to external policies and social hotspots such as “party building,” “revitalization,” and “public welfare.”

Investor Communication and Information Disclosure (0.185) focuses on meeting formal requirements and stakeholder expectations through information dissemination, using information as the primary tool for responsibilities. This includes unilateral responses to investor concerns via Investor Relations and Performance Statements (0.073), Routine Operations and Data Disclosure (0.072), and Compliance in Information Statements (0.040).

4.3. How Firm Characteristics Influence Variations in CSR Behavior

Building on the topic identification of strategic and non-strategic CSR, this section compares the differences in topic weights across firms operating under distinct institutional environments and market conditions to further investigate the shaping role of external circumstances on CSR behaviors.

A company’s CSR behavior is essentially an adaptive choice driven by both institutional and market logic [47]. In particular, the institutional motivation of a company to fulfill CSR is directly influenced by the nature of its ownership, which determines its position within the national governance structure and its access to resources [48]. Industry characteristics define the key external pressures that companies face (such as environmental regulations and consumer concerns) and profoundly influence their market-oriented path to integrating CSR issues with their core business [26,49]. Therefore, to systematically reveal how external circumstances shape the differentiation between these two types of CSR practices, the following analysis focuses on the two structural dimensions of ownership and industry characteristics.

Figure 7a illustrates the differences in topic weights of SCSR practices among enterprises with different ownership natures. The results show that state-owned enterprises (SOEs) show higher topic weights in areas highly correlated with policy guidance, such as *Technology Exploration and R&D* and *Rural Revitalization and Philanthropy*, while non-SOEs show higher topic weights in areas closer to market competition, such as *IP Creation and Protection*, *Supply Chain Relationship Management*, and *Employee Skills and Career Development*. These descriptive differences suggest that ownership nature affects an enterprise’s different focuses and implementation paths in fulfilling SCSR by influencing its institutional role and resource acquisition.

Figure 7b illustrates the differences in topic weights between enterprises with different ownership natures regarding non-strategic CSR practices. The results show that in areas subject to mandatory legal and regulatory oversight and standardized practices, such as *Pollution Emission and Monitoring*, *Hazardous Substance Testing and Handling*, and *Routine Operations and Data Disclosure*, the practices of SOEs and non-SOEs are largely similar, with both demonstrating a high degree of consensus on compliance bottom lines.

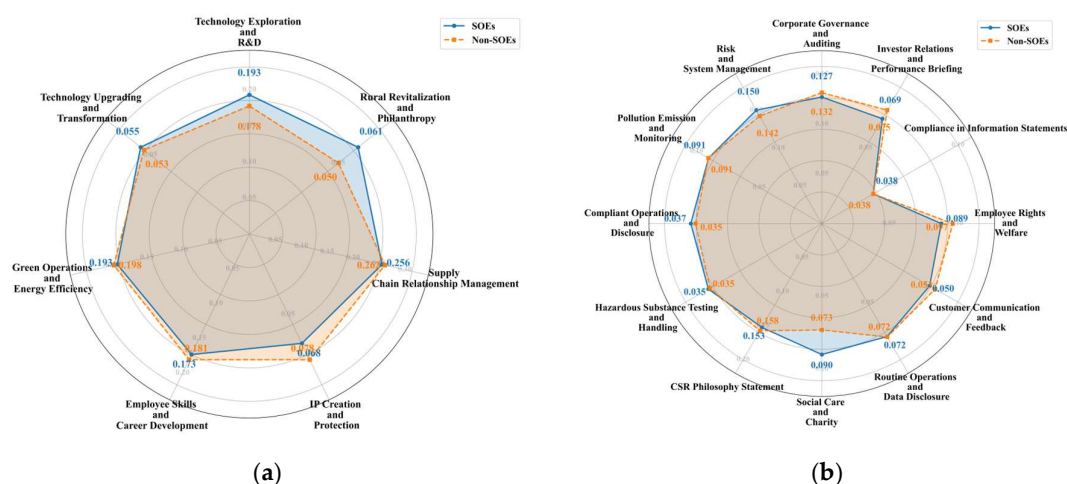


Figure 7. Topic weights of strategic and non-strategic CSR across ownership natures. (a) Comparison of SCSR topic weights between SOEs and non-SOEs; (b) Comparison of non-strategic CSR topic weights.

However, in other dimensions of non-strategic social responsibility, these enterprises' practices differ in emphasis. SOEs tend to focus more on external public issues and compliance governance, showing higher topic weights such as *Social Care and Charity*, *Risk and System Management*, and *Compliant Operations and Disclosure*. In contrast, non-SOEs concentrate more on internal organizational development and market communication, including *Employee Rights and Welfare* and *Investor Relations and Performance Briefing*. In general, SOEs demonstrate stronger policy rigidity and a social safety net in their non-strategic CSR practices, whereas non-SOEs tend to prioritize market adaptation and governance optimization in fulfilling their responsibilities to direct stakeholders.

Figure 8 illustrates the differences in the topic weights of strategic and non-strategic CSR practices between enterprises with different industry characteristics. The classification is first based on the degree of environmental responsibility constraints. Industries with high environmental responsibility encompass sectors such as energy and chemicals, metallurgy and mining, and study and printing. In the SCSR dimension, these industries show higher topic weights in *Green Operations and Energy Efficiency* but show lower topic weights in other industries in *Technology Exploration and R&D* and *Employee Skills and Career Development*. In the non-strategic CSR dimension, these industries show higher topic weights in compliance issues such as *Corporate Governance and Auditing* and *Hazardous Substance Testing and Handling*, while their investment in *Social Care and Charity* is relatively limited. These industries face strict emission regulations and the risk of fines for exceeding standards, and the pressure of environmental compliance directly drives their strategic and non-strategic CSR resources toward environmental technology improvements and cleaner production.

High-consumer-sensitivity industries encompass sectors such as food and pharmaceuticals, automobiles and home appliances, and internet and data services. These industries show higher topic weights in *IP Creation and Protection* and *Employee Skills and Career Development*, focusing on addressing consumers' core concerns about safety, quality, and privacy through technological protection and human resource investment. In contrast, low-consumer-sensitivity industries tend to build technological barriers through *Technology Exploration and R&D*. These descriptive patterns suggest that market choices, represented by consumers, significantly influence the allocation of CSR resources. Industries with high consumer sensitivity often embed SCSR into their service systems because their products and services directly involve health, safety, and data privacy.

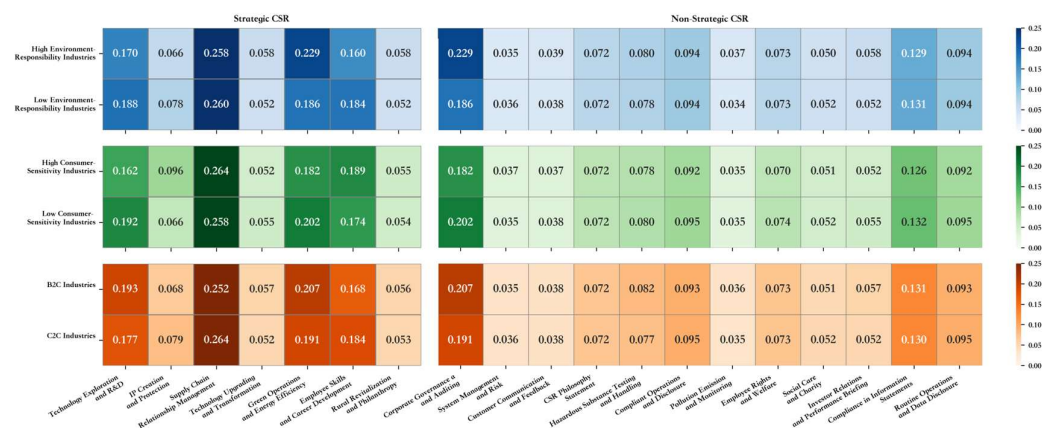


Figure 8. Topic weights of strategic and non-strategic CSR across industries.

Categorizing by proximity to consumers, terminal consumer goods manufacturing industries (B2C industries) encompass sectors such as fast-moving consumer goods, fashion and entertainment products, and durable consumer goods, which reach consumers through retail channels or e-commerce platforms. B2C industries show higher topic weights in *IP Creation and Protection* and *Employee Skills and Career Development*, while C2C industries show higher topic weights in areas such as *Technology Exploration and R&D*, *Technology Upgrading and Transformation*, and *Green Operations and Energy Efficiency*. These descriptive observations suggest that consumer pressure and demand significantly influence the allocation of CSR resources for these companies. B2C industries appear to address competitive pressure by increasing their brand value and enhancing employees' capabilities, while C2C industries improve the efficiency of synergy through technological upgrading and green innovation.

It should be noted that the comparisons of CSR topic weights across ownership types and industry contexts reported in this section are based on descriptive observations rather than formal statistical tests. These findings should therefore be interpreted as exploratory and hypothesis-generating, rather than confirmatory.

4.4. Regression Results for SCSR and Firm Performance

According to the above analysis, enterprises exhibit differentiated CSR practices under the influence of institutional and market logic. This leads to a proposition awaiting verification: Does differentiated allocation of CSR resources exert a systematic influence on firms' economic outcomes? To address this question, this section systematically examines the consequences of SCSR on financial performance and market valuation. The following benchmark regression model is constructed:

$$Performance_{i,t} = \beta_0 + \beta_1 StrategicCSR_{i,t-1} + \gamma Controls_{i,t} + \lambda_t + \mu_j + \epsilon_{i,t} \quad (3)$$

where i represents the firm and t represents the year. To control for time trends and industry heterogeneity, we include year fixed effects λ_t and industry fixed effects μ_j . All models are estimated using ordinary least squares (OLS). Given the three-year panel structure, which provides limited within-firm temporal variation for firm fixed effects estimation, we employ industry and year fixed effects as our main specification. Standard errors are clustered at the firm level to account for serial correlation within firms.

The dependent variable, $Performance_{i,t}$, comprises three variables: Return on Assets (ROA), Return on Equity (ROE), Price-to-Book Ratio (PB). The explanatory variable, $StrategicCSR_{i,t-1}$, represents the proportion of SCSR disclosed in the CSR report from the previous year.

Controls are a series of firm-level control variables, including firm size (Size), debt-to-equity ratio (Lev), growth (Growth), liquidity (Liquid), R&D intensity (R&D), board independence (Indep) and size (Board), and equity concentration (Top1). We also control for CSR disclosure quality using a composite index (0–8) measuring the breadth of CSR reporting across eight dimensions [50], to mitigate concerns that reporting quality may confound the SCSR–performance relationship.

Table 5 reports the empirical results. In columns 1 and 2, SCSR shows positive coefficients for ROA and ROE. In column 3, SCSR shows a positive coefficient for PB. Overall, these patterns suggest that firms with a higher proportion of SCSR tend to perform better in financial performance and market valuation. However, it is also possible that profitable and well-governed firms have greater capacity to conduct and disclose strategic CSR, which may drive the observed associations. These findings should therefore be interpreted as associations rather than causal effects.

Table 5. Regression results of strategic CSR on firm performance.

	(1) ROA	(2) ROE	(3) PB
StrategicCSR _{i,t-1}	0.0421 *** (0.0126)	0.0740 *** (0.0240)	1.4573 ** (0.6182)
Size	0.0118 *** (0.0013)	0.0247 *** (0.0024)	−0.5214 *** (0.0672)
Lev	−0.1799 *** (0.0114)	−0.2564 *** (0.0255)	3.2658 *** (0.7584)
Growth	0.0591 *** (0.0049)	0.1146 *** (0.0093)	1.9240 *** (0.2085)
Liquid	0.0016 (0.0011)	−0.0005 (0.0016)	0.2353 *** (0.0593)
R&D	−0.3251 *** (0.0406)	−0.5530 *** (0.0751)	6.5146 *** (1.5488)
Indep	0.0329 ** (0.0147)	0.0714 ** (0.0279)	−0.0140 (0.6313)
Board	−0.0141 *** (0.0052)	−0.0268 ** (0.0106)	−0.1834 (0.2453)
Top1	0.0192 ** (0.0087)	0.0402 ** (0.0161)	0.9153 ** (0.4500)
CSR Quality	−0.0007 (0.0020)	0.0023 (0.0039)	−0.1387 (0.0960)
Constant	−0.1306 *** (0.0346)	−0.3774 *** (0.0646)	13.2200 *** (1.6805)
Industry/Year FE	yes	yes	yes
N	3022	3022	3022
Adjusted R ²	0.4008	0.3353	0.2533

Notes: *** $p < 0.01$, ** $p < 0.05$.

5. Discussion

This study addresses the identification and process mechanisms of strategic CSR. A four-dimensional framework was developed, comprising strategic intent, business synergy, resource investment, and value output. The system boundary was explicitly defined, and the co-occurrence mechanism and reinforcing feedback mechanism were specified as the system’s operational logic. Building on this theoretical foundation, a hybrid text-mining approach was constructed by integrating pre-trained language models with LDA topic modeling. This PLM-LDA framework was applied to examine three research questions concerning the identification, differentiation, and economic performance implications of SCSR in the context of Chinese manufacturing firms.

Three interrelated findings emerged from the analysis.

First, a structural contradiction exists in the CSR practices of Chinese manufacturing firms. Firms have developed the willingness and resource capacity to engage in CSR, yet they have not translated these commitments into strategic integration with core operations. The cross-analysis revealed that strategic intent is a key driver of business synergy, corroborating Burke and Logsdon's [2] observation that the alignment between CSR and a firm's mission is a prerequisite for performance effects. Resource investment constitutes a necessary foundation, yet it fails to generate value without the complementarity of strategic intent and business synergy [9,51]. These findings validate the co-occurrence logic of the four-dimensional framework, indicating that all four dimensions must be jointly present for CSR to translate into sustainable competitive advantage.

However, this persistent gap calls for reflexive scrutiny beyond variable-level correlations. As Lovatte et al. [18] reveal through their qualitative analysis of a Brazilian airline, CSR practices are discursively constructed within socio-political and organisational contexts; crucially, the integration of CSR into strategy is mediated by governance arrangements, implying that the shortfall observed in our findings may be attributable to governance arrangements that have not yet fully matured. With this caveat in mind, we turn to the second finding concerning the thematic differentiation of CSR practices across ownership and industry contexts.

Second, strategic and non-strategic CSR practices occupy distinct thematic spaces. At the ownership level, state-owned enterprises show higher topic weights in policy-aligned domains, whereas non-state-owned enterprises show higher topic weights in market-aligned domains. This pattern is consistent with the institutional logic perspective [47,48,52]. At the industry level, environmentally regulated industries tend to prioritize green operations in response to regulatory pressures, while high-consumer-sensitivity industries emphasize intellectual property creation and employee skills in response to consumer demand. B2C industries show higher investment in these areas than C2C industries. These patterns suggest that firms adjust their strategic CSR strategies under the dual pressures of institutional mandates and market demands [26].

Third, SCSR shows positive associations with financial performance and market valuation. These findings are consistent with prior studies reporting positive relationships between CSR and firm performance [6,25] and with the meta-analytic evidence that strategic orientation strengthens the CSR-performance link [53]. They also align with the theoretical expectation that strategic CSR generates competitive advantage through the accumulation of intangible resources [53,54].

This study offers practical tools for firms and stakeholders to evaluate the strategic quality of CSR. Managers should articulate clear strategic intent for CSR initiatives and align them with the firm's core mission and capabilities. Policymakers should incentivize strategic integration rather than merely mandating disclosure. Investors and rating agencies should treat the strategic quality of CSR, rather than the volume of investment, as a signal of long-term firm value.

Three limitations warrant acknowledgment. First, the reliance on voluntarily disclosed CSR reports may introduce social desirability bias. Firms with stronger impression-management incentives may produce rhetorically sophisticated but substantively shallow reports, potentially inflating the identification of SCSR. The RoBERTa model may also learn to associate certain linguistic patterns with strategic CSR that do not necessarily reflect actual strategic integration. Future research could triangulate the findings with independent data sources such as external audits, third-party ESG ratings, or media reports. Second, the paragraph-level analysis captures a cross-sectional snapshot rather than a longitudinal strategic process, which prevents tracking of how individual firms' SCSR

profiles evolve over time. Future research could adopt panel data designs to examine the temporal dynamics of the four dimensions. Third, the comparisons of CSR topic weights across ownership types and industry contexts are based on descriptive observations rather than formal statistical tests with multiple-testing corrections. Future research could employ formal coefficient-difference tests with FDR adjustments to validate the observed patterns.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/systems14081007/s1>, Annotation Codebook, including operational definitions, inclusion and exclusion criteria, disagreement resolution rules, and annotator qualifications.

Author Contributions: Conceptualization, X.F. and Q.L.; methodology, X.F.; software, D.C.; validation, X.F. and D.C.; formal analysis, D.C.; investigation, X.F.; resources, D.C.; data curation, D.C.; writing—original draft preparation, X.F. and D.C.; writing—review and editing, Q.L.; visualization, D.C.; supervision, Q.L.; project administration, Q.L.; funding acquisition, X.F. and Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Major Program of the National Fund of Philosophy and Social Science of China, grant number 24ZDA031; the National Natural Science Foundation of China, grant number 72072060; and the National Natural Science Foundation of China, grant number 72374074.

Data Availability Statement: The CSR report data used in this study are publicly available from the CNINFO website at <http://www.cninfo.com.cn> and the other data from the CSMAR database at <https://www.csmar.com>.

Acknowledgments: The authors would like to express their sincere gratitude to the National Fund of Philosophy and Social Science of China (Grant No. 24ZDA031) and the National Natural Science Foundation of China (Grant Nos. 72072060 and 72374074) for their financial support. The authors also thank the editors and anonymous reviewers for their constructive comments and suggestions.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AUC	Area under the curve
BERT	Bidirectional Encoder Representations from Transformers
CSMAR	China Stock Market & Accounting Research Database
CSR	Corporate social responsibility
ESG	Environmental, Social, and Governance
GenAI	Generative artificial intelligence
LDA	Latent Dirichlet Allocation
PLM	Pre-trained language model
PR	Precision–recall
ROC	Receiver operating characteristic
SOE	State-owned enterprise
ST	Special treatment
SCSR	Strategic corporate social responsibility
TF-IDF	Term Frequency–Inverse Document Frequency

Appendix A

This appendix provides supplemental results that support the methodological choices in the main text. Table A1 reports the performance of conventional TF-IDF baselines. Table A2 presents the 95% bootstrap confidence intervals for F1-scores of the pre-trained

language models. Table A3 reports the LDA topic model diagnostics. Table A4 presents the robustness checks for alternative aggregation rules.

Table A1. Performance of TF-IDF-based conventional baselines across four strategic CSR dimensions.

Models	Dimension	Accuracy	Precision	Recall	F1-Score
RF (TF-IDF)	Strategic Intent	0.8623	0.8536	0.8674	0.8604
	Business Synergy	0.8785	0.8885	0.6410	0.7447
	Resource Investment	0.9131	0.9192	0.9773	0.9474
	Value Output	0.8574	0.8633	0.8105	0.8360
RF (TF-IDF)	Strategic Intent	0.8623	0.8536	0.8674	0.8604
	Business Synergy	0.8785	0.8885	0.6410	0.7447
	Resource Investment	0.9131	0.9192	0.9773	0.9474
	Value Output	0.8574	0.8633	0.8105	0.8360
RF (TF-IDF)	Strategic Intent	0.8623	0.8536	0.8674	0.8604
	Business Synergy	0.8785	0.8885	0.6410	0.7447
	Resource Investment	0.9131	0.9192	0.9773	0.9474
	Value Output	0.8574	0.8633	0.8105	0.8360

Table A2. 95% bootstrap confidence intervals for F1-scores of pre-trained language models across four strategic CSR dimensions.

Models	Dimension	F1 Mean	F1 Std	F1 95% CI (Lower)	F1 95% CI (Upper)
BERT	Strategic Intent	0.9347	0.0027	0.9302	0.9374
	Business Synergy	0.9011	0.0058	0.8921	0.9064
	Resource Investment	0.9739	0.0010	0.9724	0.9753
	Value Output	0.9247	0.0021	0.9222	0.9273
RoBERTa	Strategic Intent	0.9357	0.0021	0.9325	0.9385
	Business Synergy	0.9051	0.0060	0.8959	0.9120
	Resource Investment	0.9756	0.0021	0.9719	0.9776
	Value Output	0.9287	0.0034	0.9255	0.9342
XLNet	Strategic Intent	0.7578	0.0023	0.7558	0.7618
	Business Synergy	0.5271	0.0078	0.5164	0.5385
	Resource Investment	0.9250	0.0015	0.9224	0.9267
	Value Output	0.7473	0.0037	0.7422	0.7513

Table A3. LDA topic model diagnostics: coherence and perplexity across different numbers of topics.

Number of Topics (K)	Strategic CSR Corpus		Non-Strategic CSR Corpus	
	Coherence	Perplexity	Coherence	Perplexity
2	0.4443	−9.4946	0.6612	−9.4944
3	0.5176	−9.5997	0.6972	−9.5507
4	0.6352	−9.7129	0.7076	−9.8279
5	0.6587	−9.8517	0.6994	−10.0167
6	0.6800	−9.9930	0.7138	−10.3835
7	0.6903	−10.2777	0.7194	−10.6681
8	0.6765	−10.6175	0.7267	−11.2315
9	0.6581	−10.9721	0.7345	−11.7637
10	0.6650	−11.3479	0.7392	−12.4130
11	0.6573	−11.9448	0.7443	−12.9686
12	0.6469	−12.6588	0.7468	−13.3547
13	0.6692	−13.4376	0.7392	−13.6673
14	0.6539	−14.0140	0.7322	−13.9430

Table A3. Cont.

Number of Topics (K)	Strategic CSR Corpus Coherence	Strategic CSR Corpus Perplexity	Non-Strategic CSR Corpus Coherence	Non-Strategic CSR Corpus Perplexity
15	0.6670	−14.3830	0.7334	−14.2512
16	0.6231	−14.7026	0.7126	−14.5329
17	0.6327	−15.0428	0.7258	−14.8131
18	0.6375	−15.3768	0.7246	−15.1070
19	0.6738	−15.7212	0.7308	−15.4249
20	0.6616	−16.0293	0.7117	−15.7277

Table A4. Robustness checks for SCSR identification across aggregation rules and thresholds.

Aggregation Unit	Unit Size (N)	SCSR Identified	Proportion (%)
Paragraph-AND (baseline)	765,759	196,188	25.62
Paragraph-OR	765,759	671,800	87.73
Report-level (≥ 2 dimensions present)	3022	2603	86.14
Report-level (≥ 3 dimensions present)	3022	2572	85.12
Report-level (all 4 dimensions present)	3022	2532	83.79
Firm-year (mean SCSR score ≥ 0.3)	3022	2896	95.85
Firm-year (mean SCSR score ≥ 0.5)	3022	2206	73.01

References

1. Pache, A.-C.; Santos, F. Inside the Hybrid Organization: Selective Coupling as a Response to Competing Institutional Logics. *Acad. Manag. J.* **2013**, *56*, 972–1001. [\[CrossRef\]](#)
2. Burke, L.; Logsdon, J.M. How Corporate Social Responsibility Pays Off. *Long Range Plan.* **1996**, *29*, 495–502. [\[CrossRef\]](#)
3. Porter, M.E.; Kramer, M.R. The Link between Competitive Advantage and Corporate Social Responsibility. *Harv. Bus. Rev.* **2006**, *84*, 78–92.
4. Bhattacharyya, S.S. Exploring the Concept of Strategic Corporate Social Responsibility for an Integrated Perspective. *Eur. Bus. Rev.* **2010**, *22*, 82–101. [\[CrossRef\]](#)
5. Jiang, W.; Wang, A.X.; Zhou, K.Z.; Zhang, C. Stakeholder Relationship Capability and Firm Innovation: A Contingent Analysis. *J. Bus. Ethics* **2021**, *167*, 111–125. [\[CrossRef\]](#)
6. Maury, B. Strategic CSR and Firm Performance: The Role of Prospector and Growth Strategies. *J. Econ. Bus.* **2021**, *118*, 106031. [\[CrossRef\]](#)
7. Crotty, J.; Holt, D. Towards a Typology of Strategic Corporate Social Responsibility through Camouflage and Courtship Analogies. *Corp. Soc. Responsib. Environ. Manag.* **2021**, *28*, 980–991. [\[CrossRef\]](#)
8. Zhang, D.; Pan, X.; Ma, S. The Effects of Institutional Investors and Their Contestability on Firm Innovation: Evidence from Chinese Listed Firms. *Res. Policy* **2023**, *52*, 104872. [\[CrossRef\]](#)
9. Husted, B.W.; Allen, D.B. Strategic Corporate Social Responsibility and Value Creation among Large Firms. *Long Range Plan.* **2007**, *40*, 594–610. [\[CrossRef\]](#)
10. Husted, B.W.; Allen, D.B. Strategic Corporate Social Responsibility and Value Creation: A Study of Multinational Enterprises in Mexico. *Manag. Int. Rev.* **2009**, *49*, 781–799. [\[CrossRef\]](#)
11. Mochales, G.; Blanch, J. Unlocking the Potential of CSR: An Explanatory Model to Determine the Strategic Character of CSR Activities. *J. Bus. Res.* **2022**, *140*, 310–323. [\[CrossRef\]](#)
12. Lu, J.; Guo, Z.; Singh, S.K.; Yuan, X.; Song, M. Paving the Pathways Toward Sustainable Innovation: Underperformance Duration and Corporate Innovation Willingness. *Sustain. Dev.* **2025**, *33*, 7166–7184. [\[CrossRef\]](#)
13. Bochkay, K.; Brown, S.V.; Leone, A.J.; Tucker, J.W. Textual Analysis in Accounting: What's Next? *Contemp. Account. Res.* **2023**, *40*, 765–805. [\[CrossRef\]](#)
14. Galbreath, J. ESG in Focus: The Australian Evidence. *J. Bus. Ethics* **2013**, *118*, 529–541. [\[CrossRef\]](#)
15. Muller, A.; Whiteman, G. Corporate Philanthropic Responses to Emergent Human Needs: The Role of Organizational Attention Focus. *J. Bus. Ethics* **2016**, *137*, 299–314. [\[CrossRef\]](#)
16. Chwiłkowska-Kubala, A.; Cyfert, S.; Malewska, K.; Mierzejewska, K.; Szumowski, W.; Prause, G. What Drives Organizational Agility in Energy Sector Companies? The Role of Strategic CSR Initiatives and the Dimensions of Proactive CSR. *Sustain. Futures* **2023**, *6*, 100133. [\[CrossRef\]](#)

17. Xia, L.; Li, Z.; Wei, J.; Gao, S. Doing Well by Doing Good: Unpacking the Black Box of Corporate Social Responsibility. *Asia Pac. J. Manag.* **2024**, *41*, 1601–1631. [\[CrossRef\]](#)
18. Lovatte, J.; Everton, S.S.; Imoniana, J.O.; Alakija, F. CSR-Mediating CO2 Policy Effects on Environmental Kuznets Curve in Brazil: Case Study of an Airline Company. *Adm. Sci.* **2025**, *15*, 207. [\[CrossRef\]](#)
19. Fritsch, F.; Zhang, Q.; Zheng, X. Responding to Climate Change Crises: Firms' Trade-Offs. *J. Account. Res.* **2025**, *63*, 2137–2179. [\[CrossRef\]](#)
20. Gupta, A.; Chadha, A.; Tewari, V. A Natural Language Processing Model on BERT and YAKE Technique for Keyword Extraction on Sustainability Reports. *IEEE Access* **2024**, *12*, 7942–7951. [\[CrossRef\]](#)
21. Sun, X.; Yin, J. The Power of Retail Investors: The Effect of Online Interactions over Environmental Issues on Corporate Greenwashing. *J. Bus. Ethics* **2026**, *online ahead of print*. [\[CrossRef\]](#)
22. Clarkson, M.B.E. A Stakeholder Framework for Analyzing and Evaluating Corporate Social Performance. *Acad. Manag. Rev.* **1995**, *20*, 92–117. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Hannan, M.T.; Freeman, J. Structural Inertia and Organizational Change. *Am. Sociol. Rev.* **1984**, *49*, 149–164. [\[CrossRef\]](#)
24. Chang, C.-H. Proactive and Reactive Corporate Social Responsibility: Antecedent and Consequence. *Manag. Decis.* **2015**, *53*, 451–468. [\[CrossRef\]](#)
25. Tang, Z.; Hull, C.E.; Rothenberg, S. How Corporate Social Responsibility Engagement Strategy Moderates the CSR–Financial Performance Relationship. *J. Manag. Stud.* **2012**, *49*, 1274–1303. [\[CrossRef\]](#)
26. Groening, C.; Pelozo, J. Are We Paying Attention to the Right Metric? The Impact on Firm Value of Consumers' Perceptions versus Actual Corporate Social Responsibility Actions. *J. Clean. Prod.* **2023**, *417*, 137981. [\[CrossRef\]](#)
27. Barlow, M.A.; Verhaal, J.C.; Angus, R.W. Optimal Distinctiveness, Strategic Categorization, and Product Market Entry on the Google Play App Platform. *Strateg. Manag. J.* **2019**, *40*, 1219–1242. [\[CrossRef\]](#)
28. Torugsa, N.A.; O'Donohue, W.; Hecker, R. Proactive CSR: An Empirical Analysis of the Role of Its Economic, Social and Environmental Dimensions on the Association between Capabilities and Performance. *J. Bus. Ethics* **2013**, *115*, 383–402. [\[CrossRef\]](#)
29. Camarena, J.L.; Yepes López, G.A.; Cruz Pulido, J.M.; Camacho Ríos, M.T. Are Firms Strategically Engaging in CSR? An Exploratory Study of SCSR Dimensions in an Emerging Market. In *Integrating Environmental Sustainability and Corporate Social Responsibility: A Move Beyond Profit*; Singh, P., Daga, S., Yadav, K., Eds.; Springer Nature Switzerland: Cham, Switzerland, 2025; pp. 135–160. [\[CrossRef\]](#)
30. Chen, X.; Liang, X.; Wu, H. Cross-Border Mergers and Acquisitions and CSR Performance: Evidence from China. *J. Bus. Ethics* **2023**, *183*, 255–288. [\[CrossRef\]](#)
31. Yin, J.; Jamali, D. Strategic Corporate Social Responsibility of Multinational Companies Subsidiaries in Emerging Markets: Evidence from China. *Long Range Plan.* **2016**, *49*, 541–558. [\[CrossRef\]](#)
32. Wu, Y.; Huang, D.; Ye, J. Proactive and Reactive CSR and Sustainability Performance: Knowledge and Organizational Learning. *J. Innov. Knowl.* **2026**, *12*, 100887. [\[CrossRef\]](#)
33. Porter, M.E.; Hills, G.; Pfitzer, M. *Measuring Shared Value: How to Unlock Value by Linking Social and Business Results*; FSG: Boston, MA, USA, 2012.
34. Groza, M.D.; Pronschinske, M.R.; Walker, M. Perceived Organizational Motives and Consumer Responses to Proactive and Reactive CSR. *J. Bus. Ethics* **2011**, *102*, 639–652. [\[CrossRef\]](#)
35. Latapí Agudelo, M.A.; Jóhannsdóttir, L.; Davíðsdóttir, B. Drivers That Motivate Energy Companies to Be Responsible: A Systematic Literature Review of Corporate Social Responsibility in the Energy Sector. *J. Clean. Prod.* **2020**, *247*, 119094. [\[CrossRef\]](#)
36. Abdelkarim, N. Strategic Corporate Social Responsibility and Value Creation: Empirical Study in Palestine. Master's Thesis, An-Najah National University, Nablus, Palestine, 2021.
37. Bocquet, R.; Le Bas, C.; Mothe, C.; Poussing, N. Strategic CSR for Innovation in SMEs: Does Diversity Matter? *Long. Range Plan.* **2019**, *52*, 101913. [\[CrossRef\]](#)
38. Baier, P.; Berninger, M.; Kiesel, F. Environmental, Social and Governance Reporting in Annual Reports: A Textual Analysis. *Financ. Mark. Inst. Instrum.* **2020**, *29*, 93–118. [\[CrossRef\]](#)
39. Contala, T.; Gerk, A.-M.; Hoettler, J.; Buettner, R. Topic Trends in Sustainability Disclosure of German DAX 40 Companies—A Text Mining-Based Analysis. *IEEE Access* **2024**, *12*, 77300–77335. [\[CrossRef\]](#)
40. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159–174. [\[CrossRef\]](#)
41. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186. [\[CrossRef\]](#)
42. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv* **2019**, arXiv:1907.11692. [\[CrossRef\]](#)
43. Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R.R.; Le, Q.V. XLNet: Generalized Autoregressive Pretraining for Language Understanding. *arXiv* **2019**, arXiv:1906.08237. [\[CrossRef\]](#)

44. Montariol, S.; Martinc, M.; Pelicon, A.; Pollak, S.; Koloski, B.; Lončarski, I.; Valentinčič, A.; Sitar Šuštar, K.; Ichev, R.; Žnidaršič, M. Multi-Task Learning for Features Extraction in Financial Annual Reports. In *Communications in Computer and Information Science*; Springer Nature Switzerland: Cham, Switzerland, 2023; pp. 7–24. [[CrossRef](#)]
45. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022. [[CrossRef](#)]
46. Li, L.; Lemke, F. Shaping the Future of Business Sustainability: LDA Topic Modeling Insights, Definitions, and Research Agenda. *J. Bus. Ethics* **2025**, *201*, 391–456. [[CrossRef](#)]
47. Greenwood, R.; Raynard, M.; Kodeih, F.; Micelotta, E.R.; Lounsbury, M. Institutional Complexity and Organizational Responses. *Acad. Manag. Ann.* **2011**, *5*, 317–371. [[CrossRef](#)]
48. Dong, Y.; Ma, P.; Sun, L.; Chng, D.H.M. Goodwill Hunting: Why and When Ultimate Controlling Owners Affect Their Firms' Corporate Social Responsibility Performance. *J. Bus. Ethics* **2024**, *193*, 535–553. [[CrossRef](#)]
49. Wang, H.; Liu, W. Bridging the Gap, Building Trust: Structural Holes as Dual Forces in Corporate Philanthropy. *J. Bus. Ethics* **2026**, online ahead of print.
50. Yuan, T.; Wu, J.G.; Qin, N.; Xu, J. Being Nice to Stakeholders: The Effect of Economic Policy Uncertainty on Corporate Social Responsibility. *Econ. Model.* **2022**, *108*, 105737. [[CrossRef](#)]
51. Korendijk, M.; Drogendijk, R. Beyond the CSO: How Alternative Attention Carriers Influence the Role of CSOs on CSR. *Bus. Soc.* **2025**, *64*, 1010–1051. [[CrossRef](#)]
52. Duan, W.; Xue, X.; Chen, Z.; Liu, L. How Do Institutional Pressures Influence Corporate Social Responsibility?—A Perspective Based on State-Owned Enterprises. *Inf. Technol. Manag.* **2025**, *27*, 261–275. [[CrossRef](#)]
53. Vishwanathan, P.; Van Oosterhout, H.; Heugens, P.P.M.A.R.; Durán, P.; Van Essen, M. Strategic CSR: A Concept Building Meta-Analysis. *J. Manag. Stud.* **2020**, *57*, 314–350. [[CrossRef](#)]
54. McWilliams, A.; Siegel, D.S.; Wright, P.M. Corporate Social Responsibility: Strategic Implications. *J. Manag. Stud.* **2006**, *43*, 1–18. [[CrossRef](#)]

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.