Journal of
Theoretical and Applied
Electronic Commerce Research

MDPI

Article

Regulatory Innovation for Digital Platforms in the Data-Intelligence Era and Its Implications for E-Commerce

Danyang He [1,2,3], Yilin Cai [4], Hong Zhao [1,2,5] and Zongshui Wang [6,7,*]

[1] Sino-Danish College, University of Chinese Academy of Sciences, Beijing 100190, China; hedanyang22@mails.ucas.ac.cn (D.H.); zhaohong@ucas.ac.cn (H.Z.)
[2] Sino-Danish Center for Education and Research, University of Chinese Academy of Sciences, Beijing 100190, China
[3] Department of Organization, Copenhagen Business School, 2000 Copenhagen, Denmark
[4] School of International Trade and Economics, University of International Business and Economics, Beijing 100029, China; 202200180037@uibe.edu.cn
[5] School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China
[6] Business School, Beijing Information Science and Technology University, Beijing 100192, China
[7] Institute of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
* Correspondence: wangzongshui8@163.com

Abstract

The rapid diffusion of digital technologies, including big data, blockchain, and artificial intelligence, unlocks significant potential for marketing innovation in e-commerce while simultaneously raising fresh governance challenges. Digital platforms, as core infrastructures for online transactions and marketing interactions, have therefore come under increasing regulatory scrutiny amid tensions between technological progress and social stability. This study compiles a comprehensive Chinese Digital Platform Policy dataset consisting of national-level policy documents issued from 2000 through July 2025. We introduce a time-dimension topic clustering approach using density-based LDA algorithm to construct a policy corpus with reduced thematic overlap and develop a document-level policy intensity index by quantifying and aggregating the salience of domain-specific terms across documents. Validation exercises confirm the intensity measure strongly correlates with e-commerce transaction value and with digital innovation, with statistically significant lags consistent with policy implementation and firm adaptation. Beyond offering an empirically grounded metric, our analysis traces the dynamic co-evolution of regulation and technology adoption and identify composition effects—the joint influences of enabling and disciplining policy elements—on market outcomes. We argue that such effects also reconfigure the mix of marketing innovations. Collectively, the corpus and measurement framework provide a foundation for analyzing how regulatory innovation shapes the trajectory of marketing innovation and e-commerce development.

Keywords: digital platforms; emerging technologies; digital marketing innovation; policy intensity; Latent Dirichlet Allocation (LDA)

Check for updates

Academic Editor: Hua Pang

Received: 4 October 2025
Revised: 27 November 2025
Accepted: 3 December 2025
Published: 24 December 2025

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Emerging technologies are novel or rapidly advancing innovations that are either already generating or expected to generate substantial economic and social change. Increasing scholarly attention is directed to the pathway by which such technology progresses from invention to application and, ultimately, to value creation. In the e-commerce domain, technologies such as blockchain, the Internet of Things (IoT), artificial intelligence

(AI), and augmented reality (AR) are being integrated into core operations and are reshaping the competitive landscape [1,2]. By deploying these tools, firms can monitor and anticipate dynamic consumer needs [3], deliver personalized recommendations [4], enable more precise pre-transaction matching between supply and demand [5], and accelerate the digital and intelligent transformation of e-commerce processes and supply chains [6]. Collectively, these capabilities enhance efficiency, trust, and responsiveness, deepening customer engagement and unlocking new sources of value.

Although emerging technologies are often characterized by radical novelty, rapid growth, internal coherence, and significant societal impact, they are equally marked by uncertainty and ambiguity [7]. Such duality renders their governance particularly challenging. Policymakers face the enduring dilemma of striking an equilibrium: on the one hand, fostering technological innovation so that its economic and social benefits can be realized; on the other hand, instituting governance frameworks that anticipate potential risks, safeguard against monopolistic behaviors, and preserve human agency. In the context of e-commerce, this tension is especially evident. Regulators must simultaneously facilitate the deployment of new technologies to accommodate the continuously evolving technological and market environment. Meanwhile, they also need to consider mitigating the risks these technologies may introduce, such as threats to data privacy and the reinforcement of market concentration.

In regulating e-commerce, digital platforms constitute a pivotal actor. By definition, digital platforms are technologically mediated structures, built upon digital infrastructures, that facilitate interactions among multiple stakeholder groups [8–10]. In contrast, e-commerce refers to the entire set of processes underpinned by digital technologies that enable commercial transactions [11]. The growth of e-commerce has profoundly reshaped how both supply- and demand-side activities are organized, with digital platforms serving as the central coordinating mechanism. Consequently, the expansion of e-commerce is highly contingent upon the digital technologies and infrastructures provided by these platforms. Moreover, the rules of exchange, reputation systems, and governance frameworks embedded in platforms directly shape the order of e-commerce markets. This includes areas such as antitrust enforcement, data and algorithm governance, and the correction of information asymmetries and externalities. By safeguarding the principles of security and fairness while simultaneously leveraging the economies of scale and innovative potential inherent in platformization can e-commerce achieve high-quality and sustainable development.

Consequently, digital platform policies should not be framed merely in terms of regulation versus non-regulation. When designing such policies, it is essential to consider the interplay between regulatory control and innovation incentives. Digital platforms play a central role in constructing interconnected ecosystems, within which multiple stakeholders—platform operators, complementors, and users—jointly generate value. Excessive regulatory intervention, however, risks undermining the integrity of these ecosystems and may generate broad negative consequences. For this reason, a balanced approach that carefully reconciles regulatory oversight with the promotion of innovation is indispensable [12]. Based on a comparative analysis of platform policy measures across different countries, Kim and Ahn observe that Chinese authorities have adopted an approach that integrates innovation with regulation, thereby exemplifying a model of "regulatory innovation" [12]. This policy orientation offers a unique opportunity to investigate how regulation reshapes platform innovation dynamics and reconfigures the competitive order of digital commerce.

In the development of e-commerce, governments face the persistent challenge of trading off the economic, technological, and social benefits of emerging technologies against

the potential risks arising from their integration into digital platforms. As such, the influence of regulation is inherently ambivalent. On one side, regulation is frequently perceived as a constraint, raising firms' compliance costs or curtailing certain business activities. On the other side, regulation can also generate positive incentives, such as stimulating adaptive innovations or creating new market opportunities through regulatory requirements [13]. Recent empirical studies underscore this duality. Chen et al. (2024) find that data-protection measures enhance the quantity of innovation but reduce its quality [14]. Likewise, Blind et al. (2023) show that the EU's General Data Protection Regulation (GDPR) triggered a shift from radical to incremental innovations [13]. These findings highlight the asymmetric effects of enabling policies that lower barriers and expand opportunities and disciplining policies that safeguard security and fairness. Such evidence illustrates the composition effects of regulation: the aggregate impact on e-commerce does not arise from a single intervention but from the interaction of multiple policies over time, combining both enabling and disciplining elements. However, the extant literature largely investigates individual regulations in isolation, overlooking the cumulative nature of regulatory frameworks and their temporal continuity. This gap calls for a systematic and quantitative approach to measure the overall intensity of digital-platform regulation, capturing both the evolving policy agenda and its potential lagged effects on e-commerce outcomes.

To fill the identified research gap, this study undertakes a comprehensive quantitative assessment of China's digital platform policies and addresses three core questions:

RQ1: How has the intensity of national digital platform policies in China evolved between 2000 and 2025?

RQ2: In what ways is policy implementation associated with observable outcomes in e-commerce performance?

RQ3: How do the combined enabling and restrictive components of digital platform policies influence the trajectory of marketing innovation?

This study conducts a quantitative assessment of governmental support and intervention in digital platform development by analyzing official policy documents through content analysis and topic modeling. Prior work conceptualizes policy intensity as the substantive content embodied in policy instruments [15] or as a reflection of policy salience [16–18], underscoring the central role of policy texts in capturing policy outputs and measuring intensity. Text-based metrics thus enable the identification of shifting governance priorities and their temporal dynamics. Latent Dirichlet Allocation (LDA), a widely used topic-modeling technique, uncovers latent semantic structures through a three-layer Bayesian "document–topic–word" framework [19]. However, policy documents are inherently time-sensitive, reflecting the government's priorities at specific stages. Traditional LDA treats the corpus as static and overlooks the temporal characteristics embedded in policy issuance. To overcome this limitation, we extend conventional LDA by incorporating a temporal dimension, dividing policy texts into yearly time slices. We further apply DBSCAN, a density-based clustering algorithm, to perform post-clustering of topics, reducing thematic overlap and enabling the tracing of annual policy-topic evolution. Based on this workflow, we construct a digital platform policy theme corpus and, by mapping across topic and lexical layers, develop a time-integrated policy intensity index that dynamically quantifies governmental support and intervention over time.

By doing so, we not only extend the application of topic modeling techniques in policy analysis but also contribute to the underdeveloped field of quantitative studies on digital platform regulation. Moreover, by examining the relationship between the evolving regulatory landscape, the development of e-commerce, and technological progress, this study demonstrates the positive guiding role of policy intensity in fostering both the expansion

of e-commerce and digital innovation, thereby offering insights for the optimization of digital platform governance.

The remainder of the paper is organized as follows. Section 2 reviews the literature on digital platform regulation and its implications for marketing innovation and digital innovation, and surveys existing approaches to measuring policy intensity. Section 3 describes the research design in detail, including data sources as well as the procedures for text clustering and the construction of the policy-intensity measures. Section 4 examines the relationship between policy intensity, e-commerce market outcomes, and digital innovation. Finally, Section 5 discusses the empirical findings and outlines directions for future research.

2. Digital Platform Policy and Marketing Innovation: A Brief Review

2.1. Digital Platform Regulation and Marketing Innovation

The rapid development and application of digital technologies—including the internet, big data, mobile devices, applications, social media, and more recently, AI and the Internet of Things—have profoundly reshaped marketing practice [20,21]. These technologies influence firm strategy and consumer behavior, while new modes of data collection, storage, analysis, and use have transformed how marketers create and deliver value. Together, they generate a data-rich environment that Wedel and Kannan characterize as a new form of marketing infrastructure [22], and that Huang et al. view as a set of "general-purpose capabilities" embedded within service and marketing systems [23].

To operate at scale and foster marketing innovation, emerging digital technologies rely on digital platforms that aggregate users and transactions, accumulate behavioral data, and run algorithms. These technologies provide platforms with computational power and the software–hardware infrastructure for algorithms and data analytics, while platforms institutionalize and productize these capabilities, embedding them in concrete marketing tools and practices. Because such technologies depend heavily on platforms' data concentration, algorithmic control, and cross-context integration, platform regulation exerts direct and indirect influences on marketing activities. Appel et al. (2020), in their discussion of the future of social media marketing, explicitly identify algorithmic recommendation, content moderation, and regulatory pressure as key exogenous forces shaping marketing innovation [24]. Davis et al. likewise argue that privacy, fairness, and related concerns raised by digital technologies—including issues endogenous to the digital arena such as social media addiction, fake news, cyberbullying, and risks associated with the sharing economy—affect marketing behavior, and call for cross-discipline research between marketing analytics and public policy to address the global well-being of consumers and society at large using novel, large-scale datasets [25].

2.2. Impacts of Digital Platform Regulation on Markets

The rapid rise in digital platforms, particularly large players such as Amazon and Facebook, has been fueled by their ability to leverage consumer data to deliver personalized products and services. However, this data-driven model has also intensified public and governmental concerns regarding the power and influence of digital platforms. Consequently, an increasing body of research has drawn on regulatory theory to examine platform regulation and its broader implications [26].

Existing studies have approached this issue from two main perspectives. One line of research focuses on the nature, form, and conceptualization of regulatory interventions [12,27]. Another examines the socioeconomic consequences [28] and economic dynamics [29] associated with regulation, seeking to understand how policy shapes market structure and outcomes. For example, Blind et al. show that the GDPR exhibits a posi-

tive effect on incremental innovation, while exerting certain adverse impacts on radical innovation [13]. Aridor et al. demonstrate that GDPR have complex effects from a data-management perspective: by altering the structure of consumer privacy behavior, these regulations trigger cascading changes in data quality and market pricing, partially offsetting their direct negative consequences [30]. In the Chinese context, Ke et al. find that the implementation of the Anti-Monopoly Guidelines for the Platform Economy did not enhance competition in the targeted markets. Instead, by increasing regulatory uncertainty, it significantly reduced venture capital investment and startup entry in affected industries, thereby weakening entrepreneurial activity and producing unintended negative effects [31].

Taken together, research on regulatory market-shaping reveals divergent views on the relationship between regulation and market outcomes. Some studies argue that regulation can directly or indirectly stimulate market development, while others show that regulation suppresses market activity. Yet much of this work concentrates on the static effects of a single policy, assessing market impacts before and after a particular regulatory intervention, especially its influence on innovation. What is often missing is an appreciation of the long-term, continuous, and dynamically evolving nature of digital-platform regulation. To address this gap, the present study jointly encodes a broad set of supportive (enabling) and restrictive (disciplining) policy instruments into a longitudinal policy-intensity index. This index captures the evolving overall strength of China's digital-platform regulatory environment and is used to explain the dynamic patterns of e-commerce development and digital innovation.

2.3. Policy Intensity and Measurement

Knill et al. (2012) conceptualize policy intensity as a separate dimension of regulatory output, capturing adjustments in the scope and calibration of policy instruments [15]. Conceptually, policy intensity extends the notion of policy instruments by referring to the strength and calibration of a single instrument or a policy mix—that is, the degree of strictness or generosity, the breadth of coverage, and the extent of resources, enforcement powers, and monitoring mechanisms mobilized for its implementation. In other words, the higher the policy intensity embedded in a given document, the greater the level of stringency faced by policy stakeholders [32]. Within comparative policy research, policy intensity is regarded as a substantive dimension of policy output, analytically distinct from simple counts of instruments or measures [33]. Accordingly, it provides a valuable tool for examining the dynamic evolution of policies over time.

Policy intensity has become an influential concept in quantitative policy analysis. Knill et al. suggest that it can be assessed through six dimensions—objectives, scope, integration, budgetary commitment, implementation, and monitoring [15]. These dimensions provide the basis for a content-oriented coding framework, enabling systematic evaluations of policy intensity across time, space, and policy domains [34]. Conducting such quantitative research typically requires extensive manual work, including reviewing, measuring, and annotating policy texts to construct comprehensive indices [35,36]. More recently, within the field of industrial policy, scholars have increasingly employed the administrative rank of issuing bodies as a more direct proxy for policy influence [37]. In the context of e-commerce, however, regulatory policy intensity has remained highly dynamic, shaped by the continual emergence and application of new technologies and the disruptions they generate. To date, no research has systematically compiled or disclosed a comprehensive dataset on China's digital platform policies in parallel with the evolving trajectory of emerging technologies.

To address this gap, our study takes into account both technological advancement and the dynamic evolution of policies by collecting national-level regulatory documents on digital platforms in China. On this basis, we construct a digital platform policy corpus, from which we derive a policy intensity index through text analysis and investigate its impact on the development of e-commerce.

3. Methodology and Data

3.1. Methodological Framework

This study aims to conduct a quantitative analysis of 239 regulatory documents on China's digital platform governance by employing an improved LDA topic modeling approach. Figure 1 illustrates the methodological framework of the study, which consists of five stages.

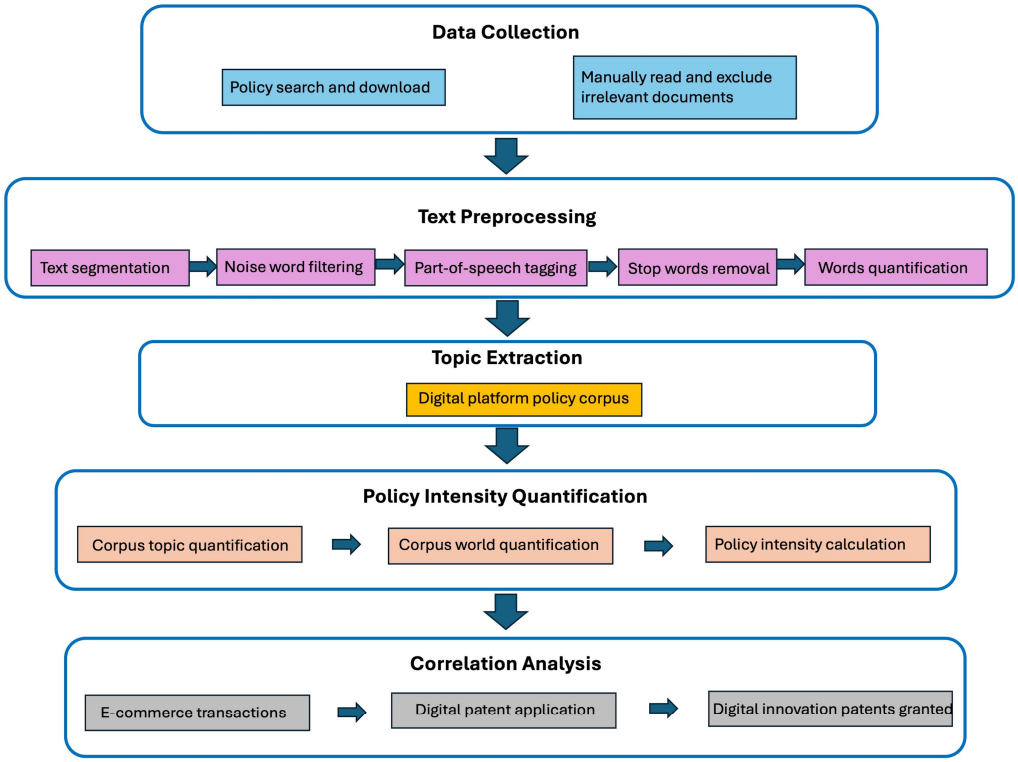


Figure 1. Research framework of the study.

Step 1: Data collection and screening. Relevant policy documents were retrieved from official policy databases through keyword searches. A manual review was then conducted to ensure accuracy, and policies unrelated to digital platforms were excluded.

Step 2: Text preprocessing. This stage involved multiple steps, including text segmentation, noise-word filtering, part-of-speech tagging, stop-word removal, and word quantification, to prepare the documents for analysis.

Step 3: Topic extraction. Using machine-learning-based topic modeling, we constructed a corpus of Chinese digital platform regulatory policies. The extracted terms form the basis for subsequent measurement of policy intensity.

Step 4: Policy intensity quantification. Building upon the terms identified in Step 3, we assigned values to each policy document to derive a systematic measure of policy intensity.

Step 5: Correlation analysis. We applied the discrete correlation function (DCF) to explore the relationship between policy intensity, e-commerce transaction volume, and indicators of digital innovation.

Text preprocessing, feature extraction, topic clustering, policy-intensity computation, and subsequent correlation analysis were all conducted using Python 3.13.7.

3.2. Data Collection

All policy documents analyzed in this study were drawn from publicly available sources. Following the policy document retrieval strategy of Huang et al. [38], we first used "digital platform" and "platform economy" as core search terms to retrieve titles from the State Council Policy Document Database (www.gov.cn) (accessed on 12 March 2025). After reviewing the initial results, we further screened documents directly related to digital platform governance to clarify the definition of "digital platform" and refine the subsequent keyword system. Building on this, we systematically collected the full texts of relevant policies from two standardized sources—the PKULAW database and the State Council Policy Document Database—to construct the policy text dataset used in this study. We conducted keyword searches in the Smart Law Retrieval database (www.pkulaw.cn) (accessed on 15 April 2025), developed by the Law Department of Peking University. As the earliest and largest legal information service platform in China, this database provides comprehensive access to statutes, regulations, and scholarly resources. Using its advanced search function, we searched policy titles with the keywords "e-commerce," "internet," "digital platform," "platform economy," and "platform enterprises." To ensure broader coverage, the same search strategy was applied to the State Council's official policy document repository (www.gov.cn) (accessed on 5 July 2025). The results from both databases were then consolidated, with duplicate documents removed. Subsequently, following the approach of Schaffrin et al. [39], we cross-validated policy documents obtained from official websites and standardized databases with non-standardized sources—such as the digital platform economy reports issued by the China Academy of Information and Communications Technology and the e-commerce reports of the Ministry of Commerce—and incorporated all verified documents into the final policy corpus.

Subsequently, we carried out a screening process guided by two main inclusion criteria: (1) documents must be issued at the national level; and (2) they must either explicitly regulate or substantially affect the behavior, obligations, or supervision of digital platforms and multi-sided markets, or stipulate operational requirements concerning core platform functions such as transaction intermediation, search and recommendation, advertising, payment and settlement, consumer and merchant protection, antitrust and competition rules, or platform labor issues. All documents were carefully reviewed, and only those explicitly pertaining to the above criteria were retained after multiple rounds of manual screening. To further enhance completeness, we cross-checked and supplemented the dataset with the annual compilations of relevant policies included in the China E-Commerce Reports published by the Ministry of Commerce. Following this procedure, we compiled an initial dataset of 239 policy documents issued between 2000 and July 2025.

In examining the impact of policy intensity on the development of e-commerce, data on the annual transaction value of e-commerce were obtained from the China E-Commerce Report and the China Statistical Yearbook compiled by the National Bureau of Statistics of China. To capture digital innovation, we employ two indicators: the number of digital-economy-related invention patent applications and the number of such patents granted in each year. These data were collected from the Chinese Research Data Services Platform (CNRDS), a comprehensive research database covering economics, finance, and business in China. The reliability of CNRDS has been well established, as its datasets have been widely used in prior academic studies.

3.3. Data Preprocessing

Prior to quantifying the policy texts, we implemented a comprehensive preprocessing pipeline to refine the dataset. The initial step involved segmenting Chinese sentences into lexical units using the jieba Python package, which is extensively applied in Chinese processing due to its robust segmentation accuracy. Next, we eliminated irrelevant or misleading tokens by filtering out terms with frequencies lower than 0.001 and higher than 99.999, and by discarding non-essential items such as personal names, locations, or institutional titles. We then carried out part-of-speech tagging to prioritize content-bearing categories while removing grammatical particles. This was followed by the removal of stop words, including high-frequency function words (e.g., prepositions, pronouns, conjunctions) that add little semantic value. As a final step, the cleaned corpus was transformed into TF–IDF (Term Frequency-Inverse Document Frequency) weighted vectors, which provide a standardized numerical representation of text and serve as the foundation for subsequent machine learning and topic modeling tasks [40].

3.4. Topic Modeling and Construction of Corpus

Policy documents typically encompass multiple thematic dimensions, such as the background of policy formulation, stated objectives, regulatory content, target groups, mechanisms of evaluation, resource allocation, and risk management. These dimensions constitute the core substance of policy communication. To systematically uncover the central themes of digital-platform regulation, we employ Latent Dirichlet Allocation (LDA) for topic modeling. LDA is an unsupervised Bayesian learning approach that identifies latent thematic structures within large text corpora without imposing strong prior assumptions about the distribution or composition of topics. By automating much of the analytic process and minimizing the need for manual intervention, LDA has become a widely used tool in computational text analysis [41,42].

While the LDA model performs well in extracting latent topics from text, the conventional formulation treats documents as static entities and largely ignores the fact that political content often evolves over time, exhibiting temporal continuity. Under this limitation, LDA assumes that each word may simultaneously belong to multiple topics, which can generate substantial topic overlap and result in ambiguous word-to-topic assignments. To better capture the dynamic nature of regulatory texts, we integrate a density-based clustering method (DBSCAN) into the stage of topic-word allocation. This study employs DBSCAN because it can automatically identify "topic clusters" and "noise points" based on local point density—without requiring a predefined number of topics, assuming regular semantic distributions, or imposing a fixed temporal structure. By strengthening topic boundaries, filtering out noise, and supporting dynamic clustering, DBSCAN directly addresses the core limitations of traditional LDA in analyzing policy corpora [43–45]. By leveraging DBSCAN's capacity to identify dense clusters and filter out noise, this enhancement improves the clarity of topic boundaries and mitigates the shortcomings of standard LDA in handling temporal variation. Consequently, the modified approach produces more coherent and accurate topic delineations in dynamic text environments.

DBSCAN is a density-based spatial clustering algorithm designed to identify clusters of arbitrary shapes while effectively handling noise. Two parameters are central to its operation: E_{ps}-neighborhood and MinPts Threshold. The E_{ps}-neighborhood of a given object p is defined as the set of points that lie within a radius ε of p. Formally, for any point of q in the dataset, if the distance between p and q satisfies $\mathrm{dist}(p, q) \leq \varepsilon$, then q is considered to belong to the E_{ps}-neighborhood of p. The MinPts Threshold specifies the minimum number of points required within a E_{ps}-neighborhood for that region to qualify as cluster. In other words, only when the number of neighboring points around a given

object is greater than or equal to MinPts Threshold can the point be regarded as part of a dense region and potentially designated as a core point.

If a point q lies within the E_{ps}-neighborhood of a core point p, then q is said to be directly density-reachable from p. More generally, if there exists a sequence of points $(p_1, p_2, \ldots, p_n)$ such that each p_{i+1} is directly density-reachable from p_i, then p_0 through p_n are considered density-reachable. Within a dataset, when all objects can be connected through chains of density-reachability, they are regarded as density-connected. The essential mechanism of the DBSCAN algorithm is to aggregate the largest possible groups of density-connected points within a dataset into clusters, while labeling the remaining unconnected points as noise.

In this study, our primary focus is on the regulatory policy texts concerning digital platforms. The documents are divided into time slices by year to capture temporal dynamics. For each slice, we first apply LDA to generate document–topic and topic–word distributions. High-weight keywords are then used to compute semantic distances, and DBSCAN is employed to perform density-based clustering and noise reduction within that time slice. Finally, clusters that are density-reachable across consecutive slices are linked to form coherent evolutionary topics. For the construction of the corpus using the EvoLDA-DB, we fine-tuned the model and determined the optimal set of hyperparameters, which are reported in Table 1.

Table 1. Hyperparameters of EvoLDA-DB model training.

Hyperparameters	Value
K	25
Alpha	0.2
Beta	0.01
Num_iters	1000

The resulting EvoLDA-DB model demonstrates notable feasibility and robustness, particularly in contexts where the number of topics fluctuates significantly over time and the vocabulary is highly dynamic. Under such conditions, its denoising capacity and topic consistency outperform standard approaches. In addition, we calculated topic coherence scores for multiple models using different calculation methods. The results indicate that EvoLDA-DB achieves the highest coherence value with different ways of calculation, confirming its superiority in capturing coherent thematic structures (see Appendix A for detailed comparisons).

3.5. Quantification of Digital Platform Policy Intensity

The quantification of terms in the digital platform policy corpus is operationalized through term frequency (Tf), which measures the relative importance of a word within the corpus. The formula is expressed as:

$$\mathrm{Tf}_w = \frac{n(w_i)}{\sum_{j=1}^{J} n(w_j)} \quad (1)$$

where $n(w_i)$ represents the number of times and the word w_i appears in the text matrix, and $\sum_{j=1}^{J} n(w_j)$ denotes the total number of word occurrences across all J terms in the matrix. This normalization ensures that the frequency of each word is considered in proportion to the entire corpus rather than in isolation. By applying this metric, the analysis avoids biases from document length and provides a standardized foundation for subsequent steps, such as topic modeling and the construction of the policy intensity index.

Next, the term frequencies of all words belonging to the same topic are aggregated to derive the topic quantization value (T_p). This procedure is expressed as:

$$T_p = \sum_{j=1}^{j} Tf_j \quad (2)$$

where Tf_j denotes the frequency of the j-th word within the topic and j represents the total number of words associated with that topic. The resulting T_p provides a unified measure that reflects the overall weight or intensity of a given topic, based on the cumulative contribution of its constituent terms. This value serves as a critical intermediate step for constructing the policy intensity index, as it links word-level quantification to broader thematic dimensions of the policy corpus.

To further quantify the intensity of each policy document, the original texts were first segmented into individual lexical units. These segmented terms were then matched with the vocabulary generated through the EvoLDA-DB model. For a given policy document, let {i} denote the set of corpus terms that appear within it. The policy intensity is calculated by multiplying the quantization value of each word (T_{pi}) by its frequency of occurrence in the document (n_i), and summing across all terms:

$$PI = \sum_{i=1}^{i} Tp_i * n_i \quad (3)$$

This procedure anchors the measurement directly in the textual content of the policy. Appendix B provides an illustrative example of how the intensity of a policy document is quantified. In this study, we measure the textual features of digital-platform policies to capture the degree of governmental support and intervention in platform development. Policy documents are used as the analytical data source because they constitute the government's formal articulation of objectives, regulatory tools, and resource allocations for a given domain. The frequency of domain-specific terms, the density of regulatory language, and the strength of normative constraints embedded in these documents serve as explicit signals of governmental intent. This approach has been adopted in prior research [46]. Methodologically, we construct a domain-specific corpus, quantify salient keywords, and compute a weighted sum of their occurrences within each policy document to derive a document-level "textual intensity" measure. By systematically extracting the regulatory density encoded in policy texts, this method translates the government's abstract policy design and intentions into a computable intensity indicator. It therefore provides a replicable and transparent basis for analyzing the strength of policy interventions.

3.6. Correlation Analysis

To investigate the impact of digital-platform policy intensity on the development of e-commerce—particularly with respect to transaction volume and digital innovation—this study applies the discrete correlation function (DCF). The method was originally introduced by Edelson and Krolik [47] and later refined by Hufnagel and Bregman [48]. DCF is especially well suited for analyzing time series with irregular sampling intervals or missing observations and is widely used to detect both correlations and time lags between two datasets. From a data perspective, our policy intensity series is derived from digital-platform regulatory texts; from a theoretical perspective, the relationship between policy issuance and e-commerce development is expected to involve temporal delays.

To compute the DCF, we first calculate the correlation for every possible pair of observations without binning, known as the unbinned discrete correlation function (UDCF). The UDCF is defined as:

$$UDCF_{ij} = \frac{(x_i - \overline{x})\left(y_j - \overline{y}\right)}{\sigma_x \sigma_y} \quad (4)$$

where x_i and y_j are individual observations from the two time series, $\bar{x}$ and $\bar{y}$ are their means, and σ_x and σ_y demote their standard deviations.

When measurement errors cannot be ignored, the denominator is modified to incorporate the error terms as:

$$UDCF_{ij} = \frac{(x_i - \bar{x})\left(y_j - \bar{y}\right)}{\sqrt{(\sigma_x^2 - e_x^2)(\sigma_y^2 - e_y^2)}} \quad (5)$$

where e_x and e_y represent the standard errors of the two datasets.

To obtain the DCF, these UDCF values are then grouped according to their time differences, $\Delta t_{ij} = t_j - t_i$ (t_i and t_j represent the respective time points of the datasets), falling within the interval:

$$\tau - \Delta\tau/2 \leq \Delta t_{ij} < \tau + \Delta\tau/2 \quad (6)$$

where τ is the time lag of interest and $\Delta\tau$ is the chosen bin width. Each group contains $M(\tau)$ pairs. The DCF at lag τ is then computed as the average of the UDCF values in that group:

$$DCF(\tau) = \frac{1}{M(\tau)} \sum_{ij=1}^{M(\tau)} UDCF_{ij} \quad (7)$$

And its variance is given by:

$$Var(\tau) = \frac{1}{(M(\tau) - 1)^2} \sum_{ij=1}^{M(\tau)} \left[UDCF_{ij} - DCF(\tau)\right]^2 \quad (8)$$

Table 2 reports the calibrated settings for the key hyperparameters of DCF following the fine-tuning process.

Table 2. Hyperparameters of DCF.

Hyperparameters	Value
Lag Lower Bound	−10
Lag Upper Bound	10
Step Size	0.25
Confidence Level	90%, 95%
Monte Carlo Simulations	1000

The aim of this study is to characterize the temporal correlation structure and dynamic co-movement between policy intensity, digital innovation, and the scale of e-commerce. Specifically, we are interested in whether policy intensity and outcome variables exhibit systematic co-movement across different temporal lags, and whether identifiable lag peaks exist. To capture these dynamic linkages, we employ the DCF method, which estimates the correlation between the unevenly spaced policy-intensity series and the corresponding digital-innovation and e-commerce series across multiple lag windows. This approach provides unconditional measures of correlation and co-movement between the variables. The resulting correlation patterns offer an empirical foundation for future causal analysis and serve as a reference for model specification in subsequent research.

4. Results

4.1. Keywords of Digital Platform Policy Corpus

Figure 2 illustrates the top 30 most frequently occurring keywords in the corpus. In conjunction with the word cloud (seen in Appendix C), the results reveal that terms such as "industry" and other industry-related expressions—"electronic commerce," "industrial,"

"markets," and so forth—feature prominently in China's digital platform regulatory policies. This pattern underscores that platforms function not merely as intermediaries facilitating interactions and transactions across multiple sides [49], but also as providers of foundational digital infrastructures, including data services, computing power, payment systems, logistics, and advertising. These infrastructural capacities enable platforms to embed themselves within diverse value chains—ranging from retail and manufacturing to logistics and cross-border trade. Consequently, platforms have evolved into both industrial infrastructure and ecosystem orchestrators, whose regulatory frameworks directly shape industrial organization, competitive dynamics, and pathways of innovation. In addition, electronic commerce is a very important industry that attracts the application and regulation of digital platforms.

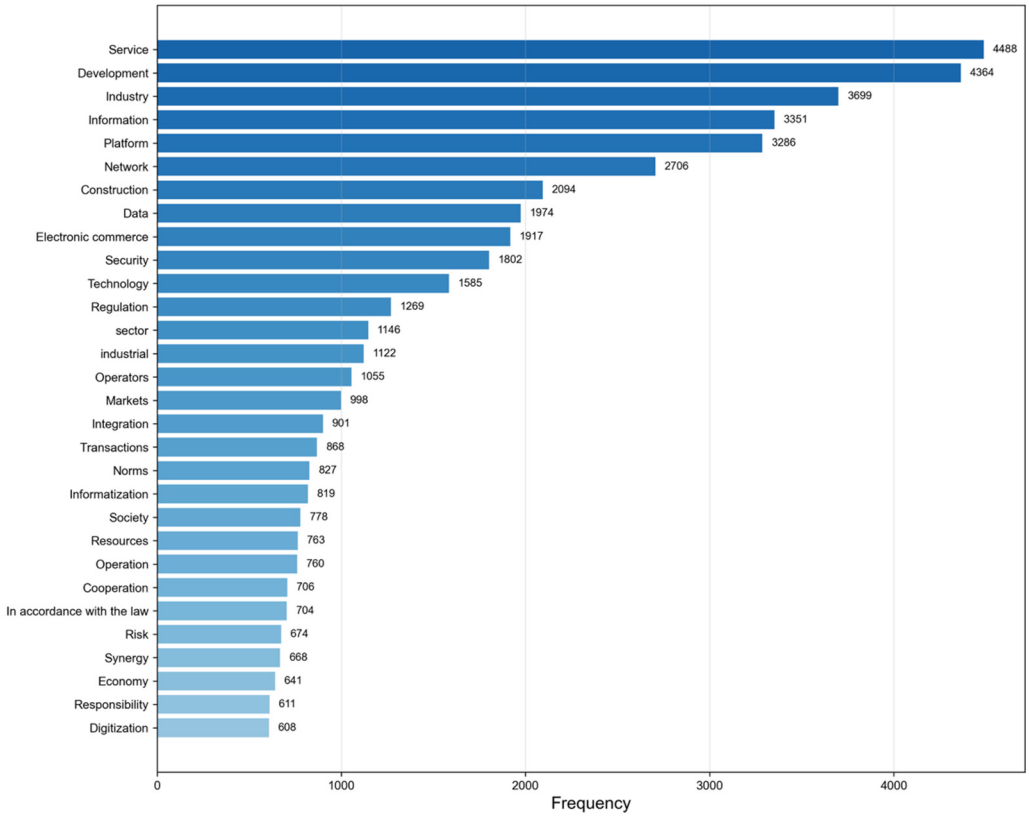


Figure 2. Top 30 high frequency words of the corpus.

Building on the preceding analysis, the terms extracted from the policy corpus are grouped into 22 thematic clusters. Table 3 reports the clustering results generated by the topic-modeling procedure, along with the average quantitative scores for each thematic cluster and the mean lexical-level intensity scores. The representative terms within each topic are listed in descending order of their intensity values.

Based on the thematic clusters and their average intensity scores reported in Table 3, four salient characteristics of China's digital-platform regulatory policies can be identified. First, the policies display a pronounced scenario-orientation, concentrating on high-frequency transactional contexts, such as e-commerce and consumption (Topics 1 and 9) and live-streaming (Topic 8). These clusters exhibit notably higher average intensity scores, indicating strong regulatory attention to concrete transactional modes and operational practices. Second, the policies emphasize industrialization and engineering-driven development. Themes centered on terms such as "industry," "construction," "manufacturing," "digitalization," "integration," and "informatization" (Topics 7, 10, and 17) rank among the highest in intensity, reflecting the policy objective of deeply embedding plat-

form applications within the real economy and industrial systems. Third, security and data governance constitute the foundational regulatory baseline. Keywords related to "data security," "cyberspace," and "risk" dominate Topics 11, 12, 14, 15, and 21, underscoring a bottom-line institutional framework. Fourth, the policies seek a dynamic balance between development and regulation. Regulatory terms such as "standardization," "orderly," "antitrust," "intellectual property," and "guidance" coexist with development-oriented vocabulary such as "development," and "collaboration," in Topics 2, 4, 13, and 16, revealing a policy philosophy that combines enabling and disciplining elements.

Table 3. Topic clusters of policy corpus based on topic model.

Topic	Topic Terms	Average Score of Cluster	Average Score of Words
1	E-commerce; Development; Service; Cooperation; Transaction; Cross-Border; Regulation; Information; Platform; Economy	1.52	198.44
2	Development; Platforms; Regulation; Standardization; Data Elements; Rules; Orderly; Technology; Collaboration; Competition	4.94	226.32
3	Government Affairs; Service Platforms; Nationwide; Operations; Processes; Security; Platforms; Data; Pilot Projects; Collaboration	2.88	103.87
4	Applications; Network Services; Transactions; Platforms; Regulation; Standardization; Information; Market Regulation; Transaction Platforms; Business Operations	1.59	37.69
5	Platforms; Services; Ride-hailing; Regulation; Drivers; In Accordance with the Law; Data; Security; Operations; Passengers	2.26	109.89
6	Algorithms; Services; Technology; Security; Information; Regulation; Cyberspace; Development; Ecosystem; Risk	7.80	55.71
7	Development; Digitalization; Services; Industry; Information; Technology; Integration; Informatization; Economy; Smart	4.52	267.30
8	Operators; Network; Services; Live Streaming; Consumers; Goods; Platforms; Transactions; In Accordance with the Law; Regulation	2.76	209.39
9	Development; Consumption; Market; Services; Economy; Goods; Platforms; Resources; Brands; E-commerce	5.74	150.14
10	Industry; Platforms; Construction; Development; Industrial; Security; Technology; Informatization; Data; Equipment	11.05	482.40
11	Security; Technology; Information; Functions; Data; Incidents; Public Opinion; Risks; Permissions; Information Security	1.66	67.30
12	Data; Data Security; Network; Data Processing; Information Processing; Risk; Law; Information; Services; Fines	4.31	155.56
13	Operators; Transactions; Platforms; Markets; Antitrust; Goods; Competition; Impact; Technology; Scrutiny	1.90	91.90
14	Network; Security; Technology; Infrastructure; Cyberspace; Informatization; Information Security; Telecommunications; Guidance; Cooperation	4.33	157.81
15	Compliance; Information; Network; Scrutiny; Risk; Reporting; Platforms; Guidance; Operations; Services	2.39	50.80
16	Intellectual Property; Pilot Programs; Commerce; Rights Protection; Inspection; Database; Rules; Platforms; Coordination; Efficiency	0.94	34.51
17	Intelligence; Technology; Resources; Industry; Manufacturing; Engineering; Security; Data; Sectors; Reform	7.24	271.22
18	Education; Society; Schools; Training; Standards; Information; Services; Development; Responsibilities; Oversight	1.22	30.66
19	Culture; Programs; Licenses; Filing; Operations; Services; Online; Telecommunications; Dissemination; Fines	1.83	80.20
20	Services; Healthcare; Information; Hospitals; Patients; Pricing; Regulation; Platforms; Data; Guidance	1.67	129.28
21	Rectification; Finance; Violations; Regulation; Risk; Financial Institutions; Assets; Regulatory Authorities; Society; Liability	1.66	56.70
22	Energy; Projects; Construction; Transactions; Development; Network; Market; Technology; Facilities; Inspection	1.48	55.13

Among the extracted keywords, another salient category relates to technology, represented by terms such as "information," "data," "informatization," and "digitalization." This emphasis reflects a defining attribute of digital platforms—namely, that they are technologically mediated entities [50]. As a result, platforms must continually adapt to the emergence and diffusion of digital emerging technologies [49,51]. Such technological advancements not only expand the opportunities and applications of platforms across different industries but also introduce new risks and governance challenges. This dynamic underscores the necessity of policy and regulatory frameworks that evolve in tandem with technological development.

A third group of keywords that plays a prominent role in the policy corpus concerns the impacts of digital platforms. On the economic dimension, particular attention has been given to the network effects of platforms, whereby the value of participation increases as the number of users and entities expands [52]. In addition, the policies emphasize the social consequences and security implications associated with platform growth, reflecting the broader range of concerns that regulators must address when designing governance frameworks for digital platforms.

4.2. Dynamic Changes in Digital Platform Intensity

Using the EvoLDA-DB model, we derived an intensity score for each policy document. To provide a clearer picture of how regulatory attention to digital platforms has evolved over time, we visualized the results with scatterplot-based boxplots (see Figure 3). The figure illustrates both the number of policy documents included in the dataset and the distribution of their intensity across different years. In the boxplot, the blue rectangle marks the interquartile range (the middle 50% of observations), while the horizontal bar inside the box indicates the median value. The upper and lower dashed lines extend to cover the top and bottom quartiles, and individual dots highlight outliers that deviate from the main distribution. The red markers denote the annual mean values of policy intensity, allowing comparison between typical yearly patterns and extreme cases.

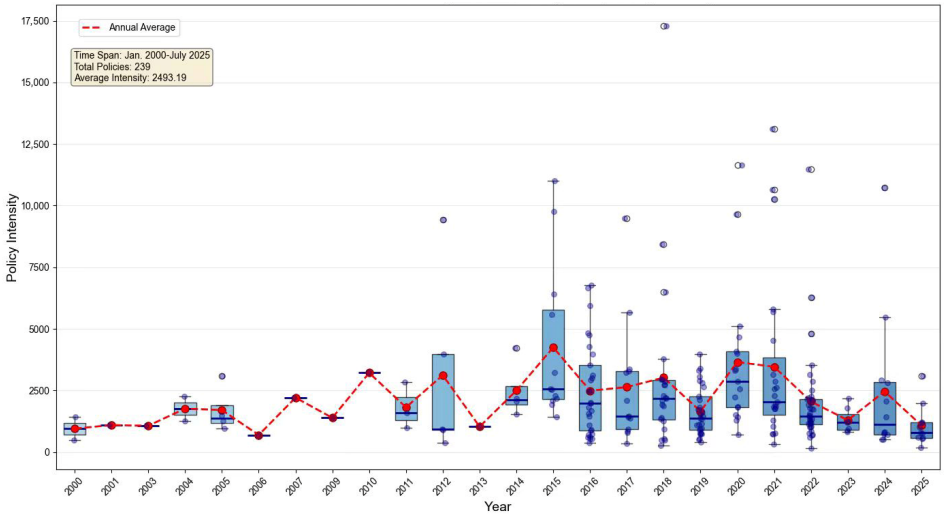


Figure 3. Boxplots of digital platform policy intensity distribution.

From 2000 to 2010, policy intensity remained at a low and relatively stable level. This period corresponds to the PC Internet era [53], when China had only recently gained Internet access. Digital platforms were still in their formative stage, with both technologies and related industries in their infancy. Regulatory actions therefore emphasized guidance and encouragement, resulting in low intensity values and limited year-to-year variation. Between 2011 and 2015, intensity began to rise steadily, reaching a notable peak in 2015.

This phase marks the mobile Internet era, during which platform business models proliferated and platform enterprises experienced explosive growth. The rapid spread of mobile Internet services accelerated e-commerce penetration, prompting not only an increase in the number of policies but also a shift in their substantive focus. Regulation began to move from promotion toward rule-setting. A landmark example is the State Council's 2015 Guiding Opinions on Actively Promoting the "Internet+" Initiative, which sought to expand Internet applications—including e-commerce—while simultaneously establishing norms for their governance. Another spike in 2018, accompanied by multiple outliers, coincided with the rapid diffusion of big data and cloud computing. To regulate the deployment of these emerging technologies, the government issued comprehensive and foundational legal instruments, most notably the E-Commerce Law 2018. As a result, the annual distribution of policy intensity during this stage displays both elevated values and strong dispersion. In 2019, the General Office of the State Council released the Guiding Opinions on Promoting the Well-Regulated and Healthy Development of the Platform Economy. Against the backdrop of powerful network and scale effects [49,52], several large platforms expanded rapidly and captured dominant market shares. This policy signaled a turning point toward stricter regulation, and in 2020 numerous ministries issued responsive measures, producing another high point in policy intensity. By 2021, China entered the intelligent Internet era. With emerging technologies increasingly embedded in platform operations, policymakers introduced topic-specific regulations on recommender systems, generative AI, and related applications, followed by a growing number of implementation guidelines and technical specifications. Taken together, these phases illustrate how regulatory intensity adapts dynamically to waves of technological adoption. Each technological breakthrough—whether mobile Internet, big data, or AI—has been accompanied by a corresponding regulatory response that both reflects and shapes the trajectory of digital platform development.

Viewed as a whole, the annual average line (in red) consistently lies above the median of the box, indicating a right-skewed distribution of policy intensity. In other words, a small number of high-intensity documents disproportionately raise the yearly average while exerting limited influence on the typical level captured by the median. In addition, beginning in 2015, both the interquartile range (IQR) and the whiskers of the boxplot expand considerably, reflecting a marked increase in intra-year variation in policy intensity. At the same time, the scatter of individual points grows denser after 2015, signifying an increase in the number of policy documents issued each year. However, the rise in intensity is not proportional to document volume; instead, it displays a pattern of larger policy density combined with greater stratification of intensity. This divergence underscores the distinction between policy density (the number of documents) and policy intensity.

4.3. Correlation Between Digital Platform Policy Intensity and E-Commerce Development

The development and consolidation of digital infrastructure create opportunities for the expansion of e-commerce. In addition, the market environment, trading rules, and reputation systems that govern platform operations are shaped and constrained by public regulation [54]. Changes in digital-platform regulatory policy affect platforms through two principal channels. First, on the market-exchange side, regulation can alter matching efficiency, transaction costs, and trust, thereby influencing the scale of e-commerce transactions. Second, viewed through the lens of the digital innovation ecosystem, government—an essential actor in that ecosystem—can steer firms' R&D investment, the direction of technological effort, and overall ecosystem dynamism via its regulatory choices.

Accordingly, examining the relationship between policy intensity and e-commerce development along two outcome dimensions—transaction scale and digital innovation—

allows us to capture both growth in scope and improvements in quality. This dual perspective provides a more comprehensive assessment of how policy intensity shapes the trajectory of e-commerce.

Drawing on regulatory theory, regulation is commonly differentiated into enabling regulation, which provides incentives and safeguards, and restrictive regulation, which imposes constraints and increases compliance costs. These two types of regulatory interventions often exert divergent influences on digital innovation and the development of e-commerce. Their combined real-world effects constitute both a central point of theoretical debate and a practical concern for policymakers. Against this backdrop, this study employs the DCF method to analyze the correlation structure between regulatory intensity, digital innovation, and e-commerce scale across multiple temporal lags, and introduces the concept of a "composite regulatory effect." The composite regulatory effect refers to the net co-movement pattern that emerges from the aggregate of regulatory enabling and restrictive forces under the prevailing institutional environment.

4.3.1. Correlation Between Digital Platform Regulation and Transaction Volume

Figure 4 depicts the relationship between the intensity of China's digital-platform regulatory policies and the annual transaction volume of e-commerce. The DCF curve reaches its peak at a lag of $\tau = 3$ years, which lies above the 90% and 95% Monte Carlo confidence intervals. This indicates that when policy intensity leads to e-commerce transaction volume by approximately three years, the correlation is strongest and statistically significant. At this point, the DCF value is 0.522, suggesting a clear positive correlation between regulatory intensity and subsequent growth in e-commerce transactions. In other words, increases in policy intensity are followed by a significant rise in e-commerce transaction volume after about three years.

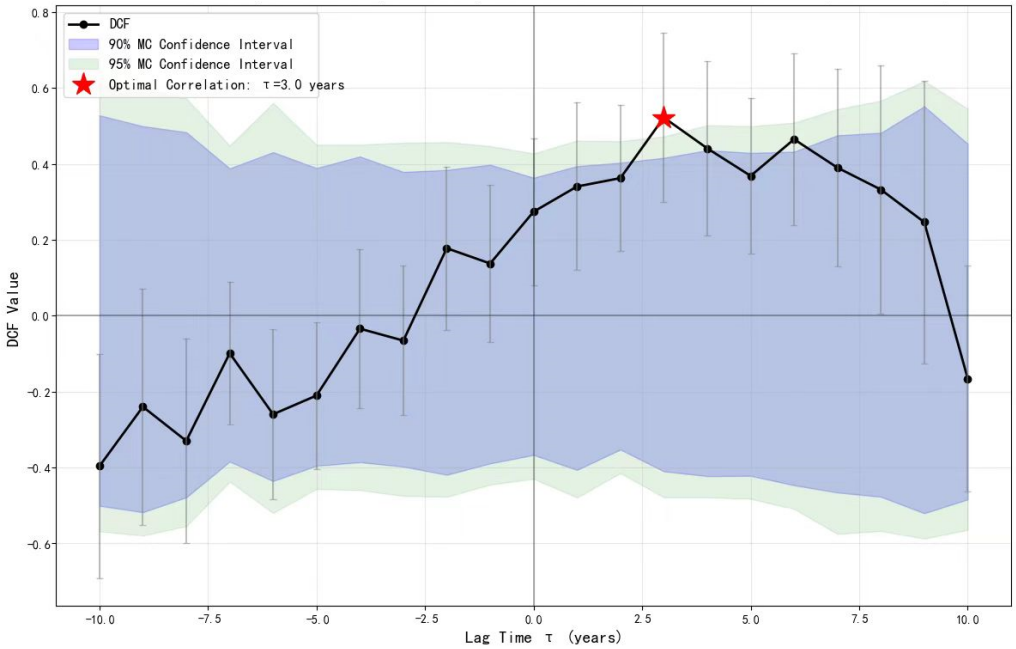


Figure 4. DCF analysis of policy intensity and e-commerce transaction volume (billion yuan).

This finding demonstrates that stronger regulatory intensity exerts a positive correlation effect on the expansion of e-commerce scale. The three-year lag is consistent with the transmission mechanism whereby policies move from formulation and promulgation, through implementation, to firms' internalization of rules, and then manifest as observable outcomes in market transactions. This pattern supports the theoretical plausibility of

the institutional transmission process and provides evidence for the viability of the policy-intensity measurement model employed in this study.

4.3.2. Correlation Between Digital Platform Regulation and Digital Innovation

Figures 5 and 6 illustrate the correlation relationship between the intensity of China's digital-platform regulatory policies and indicators of digital innovation. Figure 6 shows the correlation between policy intensity and the number of patent applications in the digital economy. The DCF curve reaches its optimal lag at $\tau = 3$ years, where the coefficient peaks at DCF = 0.564, indicating that an increase in policy intensity is most strongly associated with patent applications filed approximately three years later.

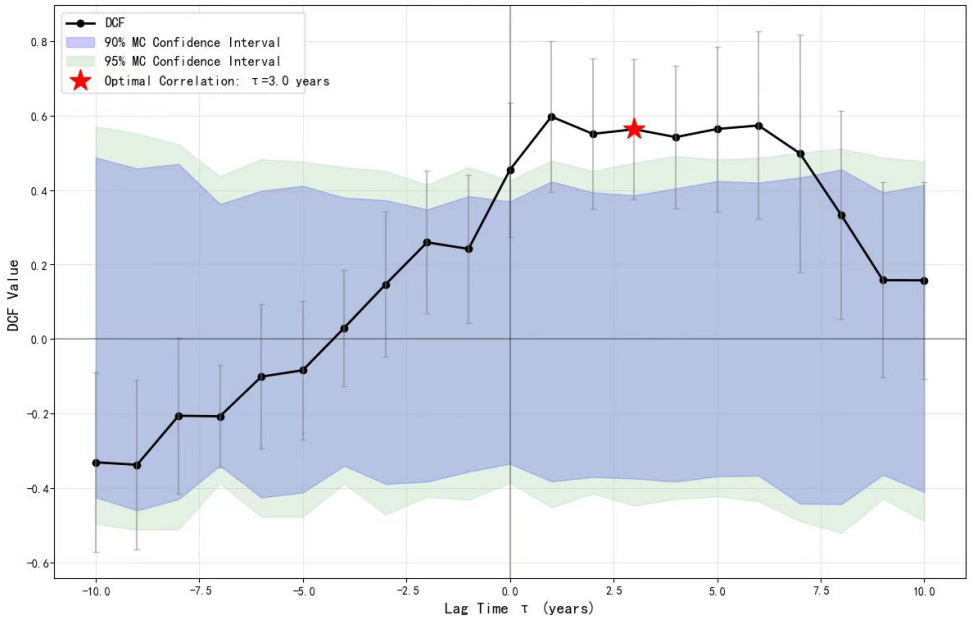


Figure 5. DCF analysis of policy intensity and digital patent application.

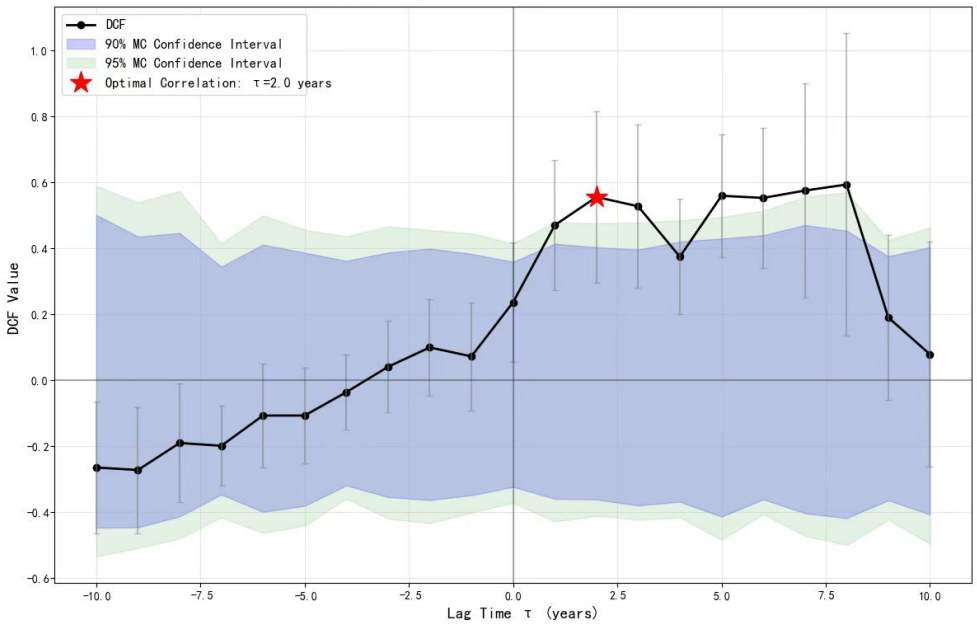


Figure 6. DCF analysis of policy intensity and digital innovation patents granted.

Figure 6 presents the results for granted patents of digital innovations. The DCF curve exhibits a statistically significant positive correlation at $\tau = 2$ years, where the value reaches DCF = 0.555, exceeding both the 90% and 95% Monte Carlo confidence intervals.

Taken together, both figures demonstrate that the correlations are consistently positive and significant when $\tau > 0$, implying that changes in regulatory intensity precede subsequent innovation outputs. The estimated lag—roughly two to three years—is consistent with the time required for policy measures to be implemented, internalized by firms, and reflected in observable innovation outcomes. This pattern highlights the guiding role of digital-platform regulation in shaping the trajectory of technological innovation in China.

Based on the analysis presented in Section 4.3, changes in digital regulatory intensity lead the growth of e-commerce transactions and digital innovation by approximately two to three years, and the two exhibit a stable positive association. This pattern indicates that increases in policy intensity and subsequent improvements in market performance tend to move in the same direction. In other words, when the government strengthens guidance and regulation of digital platforms, both e-commerce activities and digital innovation respond positively in the following years.

The observed lag can be explained from the perspectives of institutional transmission and behavioral adjustment. First, because this study analyzes national-level policy documents, it is important to recognize that national regulations typically undergo a multi-stage implementation process. After issuance, ministries and local governments must translate overarching policy principles into supporting measures, technical standards, and enforcement mechanisms. Platform firms then need to adjust their algorithmic architectures, data-processing procedures, compliance systems, and internal governance rules accordingly. E-commerce platforms and merchants further adapt their operational and marketing strategies to meet new interface requirements and compliance expectations. These layers of institutional transmission and organizational response ultimately shape market outcomes and become visible in macro-level indicators, naturally generating a temporal lag between policy change and observable effects.

Second, this lag structure aligns with Rogers' (1962) diffusion-of-innovation theory [55], which emphasizes that innovation adoption is not an instantaneous reaction to external policy stimuli but a temporal, socially embedded process. Shifts in policy intensity function as salient signals about the institutional environment and policy orientation surrounding digital-platform development, reshaping firms' expectations about the relative returns and risks of pursuing digital innovation. Upon receiving these signals, platform firms convert them into concrete innovation investment decisions and strategic actions, but this requires time for internal assimilation, resource reallocation, and phased implementation. As a result, policy impacts emerge gradually rather than immediately.

The positive association revealed by our analysis further underscores the crucial role of government in building an institutional environment conducive to digital innovation and in providing directional policy guidance. It also suggests that a well-functioning regulatory environment plays an important role in facilitating the diffusion and adoption of innovative technologies in e-commerce contexts [56].

5. Discussion and Conclusions

Our central argument is that the continuous development and application of emerging technologies in e-commerce have spurred the refinement and evolution of digital-platform regulation, thereby driving the dynamic change in policy intensity and shaping the regulatory agenda. We further contend that this dynamic adjustment of policy intensity—through the dual roles of disciplining and enabling technological applications—exerts a significant influence on the sustainable development of e-commerce. To investigate this

relationship, we employ topic-modeling techniques to construct a comprehensive Chinese Digital Platform Policy Corpus and develop a text-based index of policy intensity that quantifies the salience of domain-specific terms in each document. This provides empirical leverage to analyze the link between regulatory intensity and e-commerce development.

5.1. Theoretical Contributions

With the rapid iteration and application of emerging technologies, digital platforms—as the core infrastructure mediating these technologies—have played a pivotal role in the growth of e-commerce, and their regulation has attracted substantial scholarly attention. Our study makes significant contributions to the literature on digital-platform regulation and its economic and innovation consequences.

First, we examine the dynamic evolution of China's digital-platform regulatory policies against the backdrop of emerging technological change. Prior studies have primarily focused on the impact of individual landmark regulations [13,14], yet the development of digital platforms depends on successive waves of technological innovation and adoption. As digital technologies evolve, their economic and social implications also shift, suggesting that research on platform governance should move beyond single-policy analysis toward a dynamic and systemic perspective that accounts for changing technological contexts.

Second, we compile a comprehensive corpus of Chinese platform-regulation policies and employ the EvoLDA-DB model to generate a content-based measure of policy intensity. While previous research has explored regulatory trajectories, it has largely relied on qualitative approaches [12,57]. By contrast, our method captures the temporal continuity of policy texts, mitigating the topic-overlap problems of traditional clustering techniques. The resulting indicator provides a more nuanced representation of the evolving regulatory priorities embedded in policy discourse. This approach can also be applied in future research to quantify policy intensity across different industries and regions.

Third, our findings reveal that policy intensity is positively and significantly associated with both e-commerce transaction value and digital innovation, with market responses lagging regulatory changes by two to three years. This evidence advances theory in two ways. On the one hand, it highlights the composition effects of regulation: the aggregate outcome is shaped by the interaction of enabling intensity and disciplining intensity. On the other hand, the results demonstrate the asymmetric effects of regulatory direction: while disciplining-oriented policies may slow transaction growth in the short run, they simultaneously reallocate innovative effort toward trust-enhancing, safety-related, and compliance-oriented technologies. Together, these insights enrich our understanding of how digital-platform regulation influences both the scale and the quality of e-commerce development.

5.2. Managerial Implications

Our findings carry important implications for policymakers in the field of digital-platform governance.

First, governance approaches should shift from a "single-policy, static-effect" mindset toward a perspective that emphasizes "policy portfolios and dynamic evolution." The policy-intensity index constructed through textual analysis, together with its temporal association with e-commerce activity and digital innovation, suggests that market and innovation outcomes reflect the net effect of the broader regulatory environment rather than the influence of any individual policy in isolation. Accordingly, policy design and legislation should systematically assess the cumulative and time-sequenced effects of multiple regulatory instruments, avoiding abrupt or frequent adjustments in policy intensity and guiding regulatory transitions along a stable and predictable trajectory.

Second, policymakers should fully leverage the directional market-shaping function of regulation. The findings demonstrate a significant and positive association between policy intensity, e-commerce scale, and digital innovation, with effects emerging after a lead time of approximately two to three years. This implies that clear and credible policy direction can meaningfully shape firms' expectations and innovation behaviors. To reduce uncertainty and enhance the anticipated returns of compliant digital innovation, regulators are encouraged to establish medium- and long-term roadmaps, articulate clear compliance pathways, and maintain stable institutional expectations. In designing policy portfolios, enabling tools, such as infrastructure development, standard-setting, and SME support, should be coordinated with disciplining tools, such as competition policy, data and algorithm governance, and consumer protection, to achieve a dynamic balance between safeguarding the system and incentivizing innovation.

Third, a regularized system of policy evaluation and feedback should be established. To support sustainable development in digital-platform innovation and technology-driven e-commerce, the dynamic linkages among policy intensity, market outcomes, and innovation performance should be incorporated into ongoing monitoring frameworks. Drawing on the policy-intensity index and topic-based corpus methods developed in this study, policymakers can periodically quantify the magnitude and structure of various regulatory instruments and jointly analyze them with indicators such as e-commerce transaction value, patent applications, and patent grants. Such a data-driven approach to policy learning and iterative refinement would enable the progressive formation of a governance model that is scenario-oriented, industry-anchored, security-grounded, and development-driven, thereby providing institutional support for the high-quality and sustainable growth of China's digital economy.

5.3. Limitations and Future Research

Our results confirm the existence of a significant correlation relationship between digital-platform regulatory policies and the development of e-commerce, as well as the lagged effects of regulation on both transaction volume and technological innovation. This study seeks to examine how changes in digital-platform policy intensity shape the development of e-commerce markets and the trajectory of digital technological innovation. However, the analysis presented here focuses primarily on correlation rather than causation, and several important questions remain for future inquiry:

First, the study emphasizes the compositional effect of policy intensity on e-commerce development but does not fully disentangle causality. Future research should control for broader macroeconomic conditions, technological progress, and industry competition to more precisely identify under what conditions, through which mechanisms, and to what extent policy intensity influences e-commerce growth, marketing innovation, and digital innovation.

Second, although the present study focuses on the aggregate net effect of digital-platform policies, it offers limited discussion on how such findings can inform more systematic approaches to policy design and decision-making. Within this overall effect, the asymmetric influence of policy direction, the differing impacts of enabling versus disciplining regulatory instruments, has not been fully unpacked, particularly with respect to innovation outcomes. Building on our findings, future work could employ richer empirical data to investigate the causal mechanisms of digital-platform regulation and pay closer attention to the heterogeneous impacts of specific policy tools. In terms of innovation outcomes, the analysis treats digital innovation as a unified category and does not distinguish among its internal forms. Given that different types of innovation (e.g., incremental versus

radical) may respond differently to regulatory interventions, subsequent research could more finely classify innovation outputs to reveal potential differential effects of regulation.

Future research should not only examine the heterogeneity of policy design but also analyze its underlying layers and structural components in order to systematically assess how different digital-platform policies influence the development and efficiency of e-commerce markets. Such an approach would help identify the key policy measures and core regulatory variables that most significantly and effectively shape market outcomes. To achieve this, scholars may employ the Analytic Hierarchy Process (AHP) or its advanced variants integrated with other hybrid models [58] to conduct a systematic and structured analysis of the policy factors that contribute to the success of e-commerce markets.

Finally, the empirical analysis in this study is grounded in the context of China's national-level policy environment. Future research may extend the analytical framework in two directions. One avenue is cross-national or cross-regional comparison, which would enhance our understanding of how regional characteristics and market structures moderate policy effects and help identify divergent developmental trajectories across institutional environments. Another promising direction is multi-level policy analysis, drawing on provincial or municipal policy documents to examine implementation and translation mechanisms across administrative levels. Such work would deepen our understanding of how spatial heterogeneity shapes e-commerce dynamics and provide more targeted guidance for promoting sustainable, innovation-driven platform development.

Author Contributions: Conceptualization, D.H. and Y.C.; methodology, D.H. and Y.C.; software, D.H. and Y.C.; validation, D.H., Y.C. and Z.W.; formal analysis, D.H.; data curation, D.H.; writing—original draft preparation, D.H.; writing—review and editing, Y.C., H.Z. and Z.W.; visualization, D.H. and Y.C.; supervision, H.Z. and Z.W.; project administration, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 71972175 and Beijing Social Science Fund Project, grant number 22GLB028.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

To assess the effectiveness of the EvoLDA-DB model in processing policy-text data, we benchmark its performance against several alternative topic-clustering approaches using different coherence-calculation methods. Specifically, we employ four widely used measures of topic coherence: C_UMass (log-conditional probability based on word co-occurrence), C_V (cosine similarity derived from Word2Vec embeddings), C_NPMI (normalized pointwise mutual information), and C_UCI (PMI-based UCI coherence). These metrics are standard tools for evaluating topic consistency. Using these four measures, we compare the performance of the EvoLDA-DB model with that of LDA, LDA-Canopy, NMF, and PCA on the policy-text corpus employed in this study.

The comparative results are shown in the Figure A1. As illustrated, EvoLDA-DB (in blue) gets the highest topic-coherence score than the other models (in grey), demonstrat-

ing not only its superior ability to extract coherent and interpretable themes from policy documents but also its robustness as a topic-clustering approach for policy-text analysis.

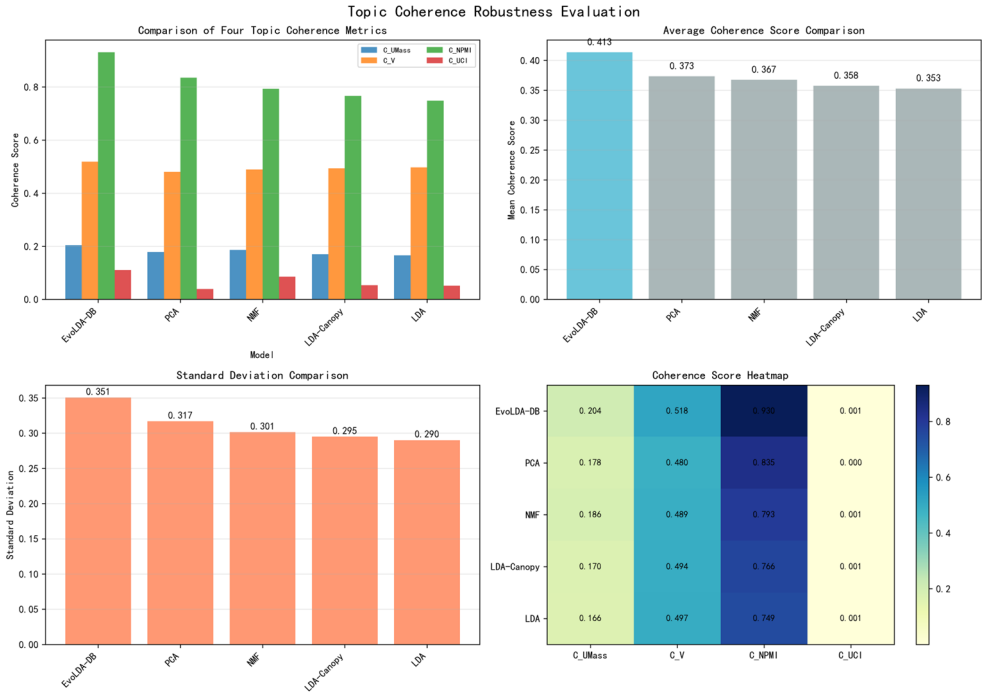


Figure A1. Evaluation of Topic Coherence for Digital Platform Policy Subject Headings.

Appendix B

Example of Policy-Intensity Quantification.

Policy document: Guiding Opinions of the General Office of the State Council on Promoting the Standardized and Healthy Development of the Platform Economy.

Year of issue: 2019.

Document length: 730 valid tokens.

Policy-intensity score: 3073.49.

Computation procedure:

- Step 1: Text preprocessing

Following the preprocessing steps described in the main text, the original policy document is transformed into a set of valid keywords, and the frequency of each keyword is calculated. For example:

regulation (监管): 45 occurrences.

platform (平台): 38 occurrences.

service (服务): 32 occurrences.

market (市场): 28 occurrences.

standardization (规范): 25 occurrences.

… (730 valid terms in total).

- Step 2: Term-level intensity scoring

Based on the EvoLDA-DB topic-clustering framework and the topic-level quantification of the domain-specific corpus, we obtain pre-computed intensity weights for 258 core terms. For example, words in Table A1:

In this step, each preprocessed keyword from the policy document is matched to its corresponding term in the domain-specific corpus.

Table A1. Example of words' pre-computed intensity weights.

Words	Cluster ID	Average Score of Cluster
regulation	2	4.94
platform	2	4.94
service	3	2.88
market	4	7.24
standardization	4	7.24

- Step 3: Policy-intensity computation

Using the policy-intensity formula presented in the main text, the topic-level weights are mapped onto each matched keyword. The intensity contribution of each term is obtained by multiplying its assigned weight by its frequency within the policy document. For example:

regulation: 4.94 × 45 = 222.30.
platform: 4.94 × 38 = 187.72.
service: 2.88 × 32 = 92.16.
market: 7.24 × 28 = 202.72.
standardization: 7.24 × 25 = 181.00.
…

Summing the contributions of all matched keywords yields the final policy-intensity score: 3073.49.

Appendix C

Figure A2 presents a word cloud constructed from the keywords of policy corpurs. The visualization highlights their relative salience: the darker the color and the larger the font size, the more frequently the term appears in the policy texts and the greater its significance within the regulatory discourse.

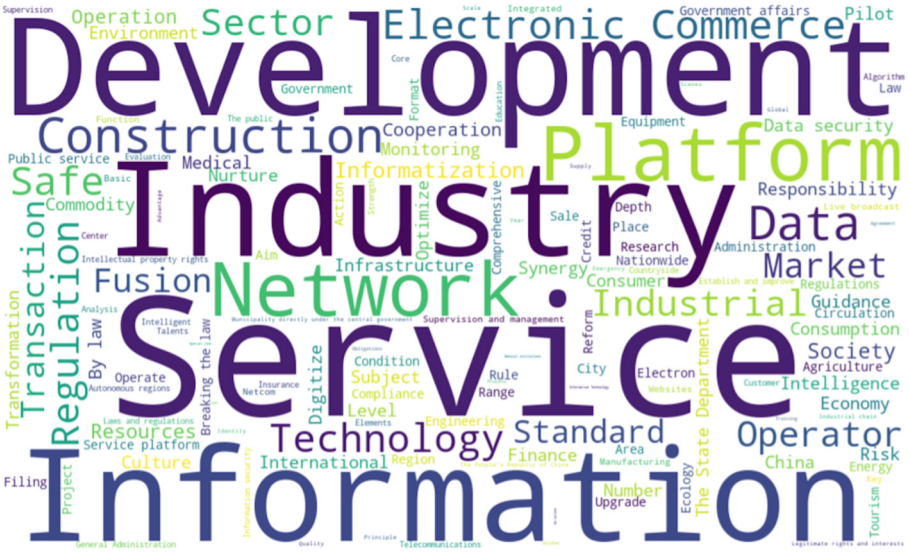


Figure A2. Word cloud of policy corpus.

References

1. Shen, B.; Dong, C.; Ng, C.T. *Technology-Driven Supply Chain Management in Industrial 4.0 Era: Resilience, Responsiveness and Reliability*; World Scientific: Singapore, 2025; ISBN 978-981-98-1170-0.
2. Singh, S.; Sai Vijay, T. Technology Roadmapping for the E-Commerce Sector: A Text-Mining Approach. *J. Retail. Consum. Serv.* **2024**, *81*, 103977. [CrossRef]
3. Shen, B.; Cheng, M.; Dong, C.; Xiao, Y. Battling Counterfeit Masks during the COVID-19 Outbreak: Quality Inspection vs. Blockchain Adoption. *Int. J. Prod. Res.* **2023**, *61*, 3634–3650. [CrossRef]

4. Ameen, N.; Tarhini, A.; Reppel, A.; Anand, A. Customer Experiences in the Age of Artificial Intelligence. *Comput. Hum. Behav.* **2021**, *114*, 106548. [CrossRef]
5. Shen, B.; Xu, X.; Yuan, Q. Demand Learning Through Social Media Exposure in the Luxury Fashion Industry: See Now Buy Now Versus See Now Buy Later. *IEEE Trans. Eng. Manage.* **2023**, *70*, 1295–1311. [CrossRef]
6. Shen, B.; Dong, C.; Tong, X.; Ngai, E.W.T. Emerging Technologies in E-Commerce Operations and Supply Chain Management. *Electron. Commer. Res. Appl.* **2022**, *55*, 101203. [CrossRef]
7. Rotolo, D.; Hicks, D.; Martin, B.R. What Is an Emerging Technology? *Res. Policy* **2015**, *44*, 1827–1843. [CrossRef]
8. Cusumano, M.A.; Gawer, A.; Yoffie, D.B. *The Business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power*, 1st ed.; Harper Business: New York, NY, USA, 2019; ISBN 978-0-06-289632-2.
9. De Reuver, M.; Sørensen, C.; Basole, R.C. The Digital Platform: A Research Agenda. *J. Inf. Technol.* **2018**, *33*, 124–135. [CrossRef]
10. Gawer, A. (Ed.) *Platforms, Markets and Innovation*; Edward Elgar: Cheltenham, UK, 2011; ISBN 978-1-84980-331-1.
11. Pozzebon, M. Future of Information Systems. In *Encyclopedia of Information Systems*; Elsevier: Amsterdam, The Netherlands, 2003; pp. 391–401. ISBN 978-0-12-227240-0.
12. Kim, J.; Ahn, S. The Platform Policy Matrix: Promotion and Regulation. *Policy Internet* **2025**, *17*, e414. [CrossRef]
13. Blind, K.; Niebel, C.; Rammer, C. The Impact of the EU General Data Protection Regulation on Product Innovation. *Ind. Innov.* **2024**, *31*, 311–351. [CrossRef]
14. Chen, W.; Wang, Y.; Wu, D.; Yin, X. Can the Establishment of a Personal Data Protection System Promote Corporate Innovation? *Res. Policy* **2024**, *53*, 105080. [CrossRef]
15. Knill, C.; Schulze, K.; Tosun, J. Regulatory Policy Outputs and Impacts: Exploring a Complex Relationship. *Regul. Gov.* **2012**, *6*, 427–444. [CrossRef]
16. Carley, S.; Miller, C.J. Regulatory Stringency and Policy Drivers: A Reassessment of Renewable Portfolio Standards. *Policy Stud. J.* **2012**, *40*, 730–756. [CrossRef]
17. Holzinger, K.; Knill, C.; Sommerer, T. Is There Convergence of National Environmental Policies? An Analysis of Policy Outputs in 24 OECD Countries. *Environ. Politics* **2011**, *20*, 20–41. [CrossRef]
18. Liefferink, D.; Arts, B.; Kamstra, J.; Ooijevaar, J. Leaders and Laggards in Environmental Policy: A Quantitative Analysis of Domestic Policy Outputs. *J. Eur. Public Policy* **2009**, *16*, 677–700. [CrossRef]
19. Vayansky, I.; Kumar, S.A.P. A Review of Topic Modeling Methods. *Inf. Syst.* **2020**, *94*, 101582. [CrossRef]
20. Hoffman, D.L.; Novak, T.P. Consumer and Object Experience in the Internet of Things: An Assemblage Theory Approach. *J. Consum. Res.* **2018**, *44*, 1178–1204. [CrossRef]
21. Davenport, T.H. *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*; The MIT Press: Cambridge, MA, USA, 2018; ISBN 978-0-262-35063-1.
22. Wedel, M.; Kannan, P.K. Marketing Analytics for Data-Rich Environments. *J. Mark.* **2016**, *80*, 97–121. [CrossRef]
23. Huang, M.-H.; Rust, R.T. Artificial Intelligence in Service. *J. Serv. Res.* **2018**, *21*, 155–172. [CrossRef]
24. Appel, G.; Grewal, L.; Hadi, R.; Stephen, A.T. The Future of Social Media in Marketing. *J. Acad. Mark. Sci.* **2020**, *48*, 79–95. [CrossRef]
25. Davis, B.; Grewal, D.; Hamilton, S. The Future of Marketing Analytics and Public Policy. *J. Public Policy Mark.* **2021**, *40*, 447–452. [CrossRef]
26. Gleiss, A.; Degen, K.; Pousttchi, K. Identifying the Patterns: Towards a Systematic Approach to Digital Platform Regulation. *J. Inf. Technol.* **2023**, *38*, 180–201. [CrossRef]
27. Cusumano, M.A.; Gawer, A.; Yoffie, D.B. Can Self-Regulation Save Digital Platforms? *Ind. Corp. Change* **2021**, *30*, 1259–1285. [CrossRef]
28. Zuboff, S. Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. *J. Inf. Technol.* **2015**, *30*, 75–89. [CrossRef]
29. Gawer, A. Digital Platforms and Ecosystems: Remarks on the Dominant Organizational Forms of the Digital Age. *Innovation* **2022**, *24*, 110–124. [CrossRef]
30. Aridor, G.; Che, Y.-K.; Nelson, W.; Salz, T. The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR. *NBER Working Papers* **2020**. Available online: http://www.nber.org/papers/w26900 (accessed on 1 December 2025). [CrossRef]
31. Rong, K.; Sokol, D.D.; Zhou, D.; Zhu, F. Antitrust Platform Regulation and Entrepreneurship: Evidence from China. *Manag. Sci.* 2025, *ahead of print*. [CrossRef]
32. Zhang, G.; Gao, Y.; Li, J.; Su, B.; Chen, Z.; Lin, W. China's Environmental Policy Intensity for 1978–2019. *Sci. Data* **2022**, *9*, 75. [CrossRef]
33. Hurka, S.; Knill, C.; Steinebach, Y. Rules as Policy Data? Measuring and Linking Policy Substance and Legislative Context. *Regul. Gov.* **2024**, *18*, 724–739. [CrossRef]

34. Schmidt, T.S.; Sewerin, S. Measuring the Temporal Dynamics of Policy Mixes—An Empirical Analysis of Renewable Energy Policy Mixes' Balance and Design Features in Nine Countries. *Res. Policy* **2019**, *48*, 103557. [CrossRef]
35. Debrun, X.; Moulin, L.; Turrini, A.; Ayuso-i-Casals, J.; Kumar, M.S. Tied to the Mast? National Fiscal Rules in the European Union. *Econ. Policy* **2008**, *23*, 297–362. [CrossRef]
36. Combes, J.; Debrun, X.; Minea, A.; Tapsoba, R. Inflation Targeting, Fiscal Rules and the Policy Mix: Cross-effects and Interactions. *Econ. J.* **2018**, *128*, 2755–2784. [CrossRef]
37. Mao, J.; Tang, S.; Xiao, Z.; Zhi, Q. Industrial Policy Intensity, Technological Change, and Productivity Growth: Evidence from China. *Res. Policy* **2021**, *50*, 104287. [CrossRef]
38. Huang, C.; Yang, C.; Su, J. Policy Change Analysis Based on "Policy Target–Policy Instrument" Patterns: A Case Study of China's Nuclear Energy Policy. *Scientometrics* **2018**, *117*, 1081–1114. [CrossRef]
39. Schaffrin, A.; Sewerin, S.; Seubert, S. Toward a Comparative Measure of Climate Policy Output. *Policy Stud. J.* **2015**, *43*, 257–282. [CrossRef]
40. Pan, Y.; Wu, D.; Luo, C.; Dolgui, A. User Activity Measurement in Rating-Based Online-to-Offline (O2O) Service Recommendation. *Inf. Sci.* **2019**, *479*, 180–196. [CrossRef]
41. Gottfried, A.; Hartmann, C.; Yates, D. Mining Open Government Data for Business Intelligence Using Data Visualization: A Two-Industry Case Study. *JTAER* **2021**, *16*, 1042–1065. [CrossRef]
42. Nanda, G.; Douglas, K.A.; Waller, D.R.; Merzdorf, H.E.; Goldwasser, D. Analyzing Large Collections of Open-Ended Feedback From MOOC Learners Using LDA Topic Modeling and Qualitative Analysis. *IEEE Trans. Learn. Technol.* **2021**, *14*, 146–160. [CrossRef]
43. Shah, G.H. An Improved DBSCAN, a Density Based Clustering Algorithm with Parameter Selection for High Dimensional Data Sets. In Proceedings of the 2012 Nirma University International Conference on Engineering (NUiCONE), Ahmedabad, India, 6–8 December; 2012; pp. 1–6.
44. Creţulescu, R.G.; Morariu, D.I.; Breazu, M.; Volovici, D. DBSCAN Algorithm for Document Clustering. *Int. J. Adv. Stat. ITC Econ. Life Sci.* **2019**, *9*, 58–66. [CrossRef]
45. Ram, A.; Jalal, S.; Jalal, A.S.; Kumar, M. A Density Based Algorithm for Discovering Density Varied Clusters in Large Spatial Databases. *Int. J. Comput. Appl.* **2010**, *3*, 1–4. [CrossRef]
46. Sewerin, S.; Kaack, L.H.; Küttel, J.; Sigurdsson, F.; Martikainen, O.; Esshaki, A.; Hafner, F. Towards Understanding Policy Design through Text-as-Data Approaches: The Policy Design Annotations (POLIANNA) Dataset. *Sci. Data* **2023**, *10*, 896. [CrossRef]
47. Edelson, R.A.; Krolik, J.H. The Discrete Correlation Function—A New Method for Analyzing Unevenly Sampled Variability Data. *ApJ* **1988**, *333*, 646. [CrossRef]
48. Hufnagel, B.R.; Bregman, J.N. Optical and Radio Variability in Blazars. *ApJ* **1992**, *386*, 473. [CrossRef]
49. Zhang, G.; MacCarthy, B.L.; Ivanov, D. The Cloud, Platforms, and Digital Twins—Enablers of the Digital Supply Chain. In *The Digital Supply Chain*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 77–91, ISBN 978-0-323-91614-1.
50. Bonina, C.; Koskinen, K.; Eaton, B.; Gawer, A. Digital Platforms for Development: Foundations and Research Agenda. *Inf. Syst. J.* **2021**, *31*, 869–902. [CrossRef]
51. Tallon, P.P.; Queiroz, M.; Coltman, T.; Sharma, R. Information Technology and the Search for Organizational Agility: A Systematic Review with Future Research Possibilities. *J. Strateg. Inf. Syst.* **2019**, *28*, 218–237. [CrossRef]
52. Münter, M.T. Platforms and Multi-Sided Markets. In *International Encyclopedia of Business Management*; Elsevier: Amsterdam, The Netherlands, 2026; pp. 813–821, ISBN 978-0-443-13702-0.
53. CAICT. Insights into the Development of the Platform Economy 2024. Available online: http://www.caict.ac.cn/kxyj/qwfb/ztbg/202407/P020240729569788593896.pdf (accessed on 25 September 2025).
54. Rodríguez-Ardura, I.; Meseguer-Artola, A.; Vilaseca-Requena, J. Factors Influencing the Evolution of Electronic Commerce: An Empirical Analysis in a Developed Market Economy. *J. Theor. Appl. Electron. Commer. Res.* **2008**, *3*, 18–29. [CrossRef]
55. Rogers, E.M. *Diffusion of Innovations*, 5th ed.; Free Press: New York, NY, USA; London, UK; Toronto, ON, Canada; Sydney, Australia, 2003; ISBN 978-0-7432-5823-4.
56. Zhu, K.; Kraemer, K.L.; Xu, S. The Process of Innovation Assimilation by Firms in Different Countries: A Technology Diffusion Perspective on E-Business. *Manag. Sci.* **2006**, *52*, 1557–1576. [CrossRef]
57. McKnight, S.; Kenney, M.; Breznitz, D. Regulating the Platform Giants: Building and Governing China's Online Economy. *Policy Internet* **2023**, *15*, 243–265. [CrossRef]
58. Simjanovic, D.; Vesic, N.; Ignjatovic, J.; Randjelovic, B. A Novel Surface Fuzzy Analytic Hierarchy Process. *Filomat* **2023**, *37*, 3357–3370. [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.