

Text Mining and Machine Learning for New-Media Public-Opinion Spillovers in Manufacturing Supply-Chain Digital Collaboration

Chengyan Li^{1,*}

¹Tianjin Foreign Studies University, Tianjin 300270, China

Abstract

Online media coverage may transmit information across supply-chain relationships, yet converting such signals into evidence requires a clear separation between text measurement, network alignment, and statistical inference. This study examines whether focal-firm media attention is associated with connected partners' annual-report digital disclosure and develops a manufacturing-oriented workflow for directed supply-chain analysis. We construct a firm-partner-year panel for Chinese non-financial A-share firms from 2007 to 2024 by combining aggregated CFND media indicators, annual-report digital-term frequencies, CNRDS supplier-customer links, and CSMAR controls. Fixed-effects models estimate conditional associations on directed edges. The baseline specification shows a positive association between focal-firm coverage and partner digital disclosure ($\beta = 0.048$, $SE = 0.020$), while alternative aggregate coverage measures yield coefficients with the same sign; mechanism and subgroup results remain observational and descriptive. These findings do not establish causality or realized interfirm digital collaboration because annual-report word frequency is an imperfect disclosure proxy. A Chinese-BERT-MLP event-and-tone module is therefore retained only as a future extension pending article-level text, human-reviewed labels, and out-of-sample evaluation.

Keywords: text mining, machine learning, supply-chain networks, digital disclosure, manufacturing-oriented information systems

Received on 29 July 2026, accepted on 08 September 2026, published on 12 October 2026

Copyright © 2026 Chengyan Li, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetsis.14249

*Corresponding author. Email: 3436445636@qq.com

1. Introduction

A manufacturing supply chain is simultaneously a physical network of materials and a digital information network connecting original-equipment manufacturers, component suppliers, distributors, and service partners. Production schedules, order interfaces, quality traceability, inventory status, and delivery information are increasingly coordinated through supplier relationship management, manufacturing execution, enterprise resource planning, and industrial Internet platforms. When a quality defect, delivery disruption, cyber incident, or environmental violation is

reported in online news media, the resulting new-media public-opinion signal may be relevant not only to the affected manufacturer but also to connected manufacturing suppliers and customers. For example, a publicized component defect may prompt an automotive manufacturer to request additional batch genealogy and inspection records from its suppliers, while a reported production interruption may lead downstream manufacturing partners to revise inventory buffers and alternative-sourcing plans. Production-network research shows that firm-specific shocks can travel through input-output linkages and that propagation depends on network position, input specificity, and the substitutability of affected partners [1–3]. Digital platforms and smart,

connected products have therefore become important foundations for interfirm coordination and manufacturing adaptation [4–6]. The corresponding computational challenge is to extract reliable signals from unstructured media information, align them with directed supplier–customer relationships, and examine their association with the connected firm's digital-disclosure index. For a manufacturing firm, the practical value of artificial intelligence lies not only in classifying news but also in linking a verified event to the affected supplier–customer edge and the relevant quality, production, inventory, or delivery process. The proposed framework therefore combines explainable text features, directed network matching, and human-reviewed decision support.

This challenge is nontrivial. Online media reports are noisy, repetitive, and heterogeneous, and the same event may be framed differently or associated with several corporate entities. A useful signal must therefore be resolved to the correct firm, time period, and supply-chain partner before it enters the estimation model. Media information may also propagate through different paths: technical disclosures can diffuse information about industrial Internet interfaces, cloud services, traceability requirements, and digital operating practices, whereas intensified media visibility can increase the salience of connected firms. Moreover, the available annual-report indicator reflects a firm's disclosed digital orientation rather than directly observing shared orders, synchronized production, platform transactions, or joint planning quality. A defensible method must consequently distinguish text measurement, supply-chain edge construction, spillover estimation, and mechanism analysis, while preserving the boundary between a disclosure-based digital-capability proxy and transaction-level collaboration.

Existing studies provide important but incomplete foundations for addressing this problem. Research on changing media environments shows that social media affects knowledge and participation [7], journalistic practice [8], misinformation diffusion [9], and the sentiment composition of recirculated news [10]. Firm-level studies further indicate that media coverage can alter the external information environment [11], analyst attention [12], innovation incentives [13–15], and corporate responses [13–15], while text-mining methods have been used to convert complaints and other unstructured records into structured analytical signals [16]. Nevertheless, much of this literature relies on aggregate coverage counts or static sentiment indicators. It rarely links an extracted media signal to a specific supplier–customer edge, and it generally does not distinguish exposure received by the focal firm from attention subsequently received by its connected partner.

A related stream of research examines digital transformation [17,18], supply-chain resilience, and interfirm performance [19]. Digital-transformation research shows that digital technologies reshape organizational and entrepreneurial processes [4,20]. Supply-chain studies link digitalization to resilience [21,22], visibility [23,24], and cross-organizational communication [25–27]. In manufacturing settings, smart products [5], industrial platforms, artificial intelligence [28–30], cloud computing,

and data analytics [31] support coordinated production, risk response, traceability, and data-driven quality control [6,32,33]. At the industry and firm levels, digital-economy development and digital inputs are associated with manufacturing upgrading and green productivity [34,35]. Vertical partnerships can also transmit digital practices and capabilities across supply-chain boundaries [36,37], and recent research has combined machine learning and text analytics with patent descriptions to measure digital-technology adoption and its relationship with supply-chain structure [38]. However, these studies generally begin with firms' internal digital investment or technological capability and seldom model online media coverage as an external information input associated with outcomes across an observed supply-chain relationship. Consequently, the relationship among media text, directed interfirm connections, and the partner-level digital-disclosure index remains insufficiently represented.

System-oriented studies address complementary parts of this problem. Industrial text analytics documents heterogeneous sources, domain artifacts, annotation shortages, and deployment constraints [39], while collaborative IIoT research emphasizes standardization, acquisition, fusion, and scalable data access [40]. EAI studies illustrate established BERT-based news classification and scalable supply-chain information architectures [41,42]. Existing systems then emphasize different operational tasks: extracting buyer–supplier relations from news to construct supply-chain maps [43], mining streaming social-media data for timely risk identification [44], converting multi-sentence Chinese documents into structured event records [45], deferring uncertain predictions to a downstream expert [46], or defining event-time, windowing, and trigger semantics for scalable data streams [47]. The present workflow does not claim to reproduce those trained or streaming systems. Its incremental contribution is to connect observed aggregate media measures to directed supply-chain edge-years, keep NLP measurement distinct from panel inference, and specify abstention and human review as safeguards; document-level event extraction and continuous stream processing remain future implementations.

Computational methods provide useful components for this workflow. Topic models summarize latent themes [48], contextual encoders [49,50] represent surrounding language [51,52], and elastic-net regularization can control overfitting when text variables are numerous and correlated [53]. Economic and management research [54,55] nevertheless requires prediction, measurement, and causal interpretation to remain distinct [56,57]. Accordingly, classification performance would evaluate a future raw-text measurement layer, whereas the directed panel estimates in Section 4 concern an observational association. The implemented analysis therefore uses transparent aggregate features; Section 3.3 defines only the auditable interface through which a later raw-text module could enter the workflow.

Against this background, the study asks whether focal-firm online media exposure is conditionally associated with the connected partner's digital-disclosure index; whether the reported mechanism regressions are consistent with interfirm

digital-technology diffusion and media-attention spillover channels; and where descriptive subgroup patterns appear in the observed panel. The implemented empirical branch uses the available broad Chinese non-financial A-share supply-chain panel for 2007–2024. The manufacturing focus describes the system architecture and deployment context, not a manufacturing-exclusive sample.

Within these evidence boundaries, the contribution is twofold:

System and method design. We specify an auditable, human-supervised workflow that aligns database-supplied aggregate media measures with directed supplier–customer edge-years and keeps NLP measurement separate from panel estimation. In the conditional raw-text branch, deduplication, firm and plant resolution, confidence-aware abstention, and human review are required before article-level features could enter the panel. Chinese-BERT and the MLP heads are established components; no neural-architecture novelty is claimed.

Empirical evidence. Using the available broad non-financial A-share panel, we document an observational association between focal-firm media coverage and partners' digital-disclosure indices, examine sequential mechanism-consistent relations and descriptive subgroup patterns, and use the estimates only to motivate conditional manufacturing review scenarios. The estimates are not causal, the sample is not manufacturing-only, and disclosure is an imperfect collaboration proxy.

2. Research Foundation and Hypothesis Development

2.1 New-Media Public Opinion and Economic Text Measurement

Business media act as information intermediaries between firms and market participants. By selecting events, repeating claims, and assigning linguistic tone, media outlets can reduce some information-search costs while also concentrating attention on a limited set of firms and issues. Classic evidence links media tone and coverage breadth to investor sentiment [58], returns [59,60], and the processing of corporate fundamentals [11,61]. More recent work [62,63] shows that media visibility can affect analyst coverage [12], innovation incentives [13–15], corporate monitoring [64], reputation [65,66], and firms' environmental responses [12,15]. These findings provide an economic basis for treating media coverage as part of the external information environment rather than as a purely communicative outcome. They do not, however, imply that every article creates an exogenous shock or that a coverage count directly measures sentiment.

Turning narrative information into a panel variable requires an explicit measurement rule. Economic text analysis emphasizes the choice of corpus, unit of observation, dictionary or model, validation target, and aggregation window [67,68]. Generic sentiment dictionaries can

misclassify technical corporate language, while text-derived risk measures require external validation against observable firm behavior [67,69]. Text can also reveal evolving product-market networks that are not captured by static industry codes [70]. In computational terms, keyword counts maximize auditability; latent topic models summarize recurring themes [48]; and contextual encoders represent words according to surrounding language [49]. The appropriate representation is therefore determined by the available corpus and the inferential target, not by model complexity alone.

Coverage volume, linguistic tone, event novelty, and source diversity have different economic meanings and should not be collapsed without justification. Volume can capture the intensity of public information production, but it also reflects firm size, disclosure propensity, and the tendency of outlets to repeat the same event. Tone can distinguish favorable from adverse framing, yet a neutral article about a plant shutdown may contain more operationally relevant information than a positive promotional article. Novelty matters because the first verified report can change beliefs whereas subsequent copies may add little information. Source diversity can improve credibility when independent outlets confirm an event, although syndicated content can create the appearance of corroboration. An ideal event-level pipeline would therefore deduplicate near-identical articles, resolve the mentioned plant and legal entity, identify the event date, classify operational content, and aggregate only after these checks. It would also normalize counts by observable media exposure or document opportunity and compare alternative aggregation windows. These choices affect the economic interpretation of the coefficient: a volume estimate describes association with observed visibility, while an event-study estimate would require a defensible innovation in the information set and a clearly dated response window [60,67,68]. The present design retains the more modest visibility interpretation.

The available data contain firm-year coverage counts and database-provided sentiment categories rather than article text. Accordingly, the empirical analysis treats total, positive, and neutral coverage as observed visibility measures, not exogenous media shocks. Any future document-level features require time-based and firm-disjoint validation before entering the partner-level model.

2.2 Text Measurement and the Proposed Raw-Text Extension

Document-level event extraction is a distinct measurement task that requires article-level annotation and evaluation; validated Chinese systems such as Doc2EDAG illustrate that requirement in financial text [45]. In this study, an established Chinese-BERT encoder and compact MLP heads [52,71] are therefore mentioned only as one possible future implementation. The proposed contribution is the auditable workflow interface—deduplication, entity and plant resolution, directed-edge alignment, abstention, and human

review—not a new neural architecture. Section 3.3 states the minimum technical boundary for that optional branch.

2.3 Production Networks and Directional Spillovers

A production network links firms through repeated input, delivery, quality, and financing relationships. Network models show that idiosyncratic disturbances can affect aggregate outcomes when shocks propagate through economically important connections [1]. Input specificity makes customers more vulnerable to supplier disruptions [2], and evidence from the Great East Japan Earthquake demonstrates that real shocks travel in both upstream and downstream directions [3]. These studies imply that the economic meaning of a firm-level signal depends on who is connected to the firm, the direction of the connection, and the availability of substitutes. A market-wide peer comparison is therefore not equivalent to an observed supplier–customer edge.

The same logic applies to information and digital requirements. A downstream manufacturer may impose electronic order formats, batch genealogy fields, delivery-status interfaces, and quality-reporting standards on its suppliers. An upstream supplier's disruption or process innovation can likewise change a customer's planning and monitoring needs. Vertical partnerships have been shown to transmit digital practices across firm boundaries [36,37], while digital supply-chain research treats information visibility [24,32] and cross-firm integration as central capabilities [25–27]. The empirical representation in this study is therefore edge-conditioned and directional: coverage belongs to focal firm i , the digital-disclosure index belongs to connected partner j , and the edge is observed in year t .

2.4 Manufacturing Digital Disclosure and Collaboration

Digital collaboration is not a single technology purchase. It is an organizational capability that combines compatible information systems, shared process definitions, data quality, and the authority to act on jointly visible information. Digitally enabled process integration can improve supply-chain performance [25], but information-technology resources create value only when they support coordination, information exchange, and response capabilities [26]. Big-data [31–33] and Industry 4.0 research [72–74] similarly stresses that analytics, visibility, and operational flexibility are complementary rather than interchangeable [24,75]. In a manufacturing setting, the relevant outcomes include synchronized schedules, earlier quality containment, lower search costs for alternative supply, and faster tracing of affected materials [76].

The available annual-report measure aggregates terms in five categories—digital applications, blockchain, cloud computing, artificial intelligence, and big data. The index directly measures annual-report digital disclosure and is

interpreted as a disclosure-based proxy for digital capability; it does not measure realized interfirm collaboration. It therefore does not directly observe shared purchase orders, production-plan synchronization, platform transactions, or interorganizational data quality. Throughout the paper, the term digital-disclosure index refers to the measured construct, whereas digital capability denotes the latent capability that the index may imperfectly proxy. This distinction preserves the economic boundary between a firm's disclosed digital resources and the realized productivity of a joint manufacturing process.

The five-category dictionary follows the annual-report text-measurement approach of Wu et al. [77], covering artificial intelligence, big data, cloud computing, blockchain, and digital-technology applications. Word frequency remains sensitive to reporting style, annual-report length, strategic emphasis, and symbolic disclosure. It does not directly reveal technology expenditure, implementation quality, system interoperability, shared transaction data, or realized supplier–customer coordination. The observed variable is therefore consistently termed a digital-disclosure index.

2.5 Mechanisms and Hypotheses

Online media coverage can be associated with a partner's digital-disclosure index through both information and governance incentives. Coverage of a focal firm's product defect, production interruption, digital initiative, or environmental event lowers the cost of discovering a risk that may otherwise remain local. Connected partners may respond by improving traceability, reporting, and interface compatibility; focal firms may also disclose new digital requirements to reassure customers, regulators, and investors. Because the available design is observational and contemporaneous, the predicted relationship is stated as an association rather than a causal impact. H1. Greater online media coverage of a focal firm is positively associated with the digital-disclosure index of its connected supply-chain partner.

The association is not mechanically positive. Adverse coverage may tighten financing constraints, divert managerial attention, or induce symbolic disclosure without changing production routines. Small suppliers may lack the absorptive capacity, bargaining power, or complementary equipment needed to respond to a focal firm's digital standards. Conversely, firms with public-policy obligations or larger digitalization budgets may respond more quickly to external scrutiny. These competing forces motivate the subsequent heterogeneity analysis by supply-chain position, ownership, and region rather than an assumption that every partner experiences the same average relationship.

The first proposed channel is interfirm digital technology diffusion. Repeated transactions make digital interfaces and operating routines partly relationship-specific. When focal firms strengthen digital processes under greater scrutiny or information demand, suppliers and customers have incentives to adopt compatible practices. Evidence on vertical partnerships documents digital spillovers across supply-chain

boundaries [36,37]. More broadly, FDI and spatial digital-economy research documents technology and innovation spillovers [78,79], while evidence from manufacturing SMEs emphasizes formal innovation strategy and structure [80]. H2. Focal-firm online media coverage is positively associated with the focal-firm digital-disclosure index, which in turn is positively associated with the partner digital-disclosure index. The sequential regressions evaluate whether the reported estimates are consistent with this channel; they do not by themselves establish a causal indirect effect.

The second proposed mechanism is a media attention spillover channel. Coverage of a focal firm can direct journalists, investors, regulators, and customers to connected suppliers or buyers. Evidence from other information environments shows that media effects can spill across linked issues [81] and that information can spill across markets [82]. Financial-media research further shows that coverage changes information processing and attention allocation [11,58,60]. A partner facing greater visibility may have

stronger incentives to disclose digital controls or invest in systems that make quality and delivery responses auditable. H3. Focal-firm coverage is positively associated with partner media coverage, which is positively associated with the partner digital-disclosure index. The two proposed channels may coexist and are tested with common variable definitions.

3. Research Design and Estimation Method

3.1 Overall System Architecture

Figure 1 presents the text-network architecture that links aggregated media exposure, annual-report disclosure, directed supply-chain links, panel estimation, and a proposed raw-text measurement extension.

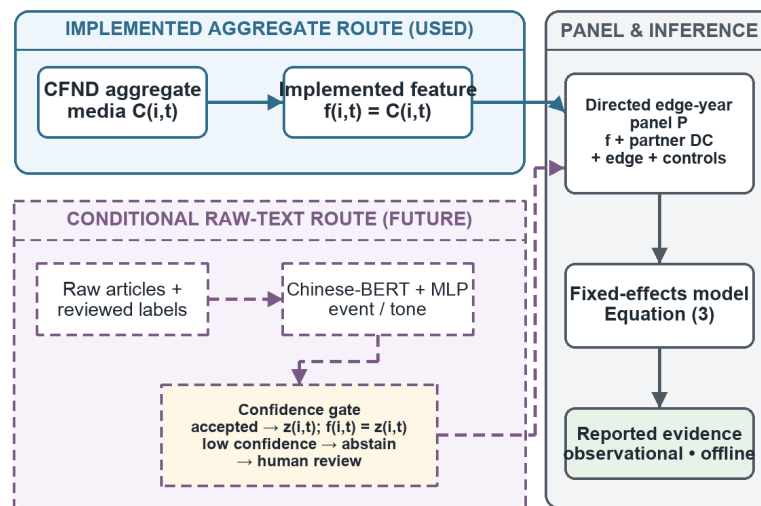


Figure 1. Dual-route text-network architecture for manufacturing-oriented supply-chain public-opinion analysis. The solid aggregate route ($f = C$) is implemented; the dashed raw-text route (accepted z and $f = z$) is conditional on training and validation. Both connect to the directed edge-year panel, but the reported evidence comes only from the aggregate route.

The framework separates data sources, measurement, directed panel construction, estimation, and evidence use. CFND aggregates, annual-report digital terms, CNRDS supplier-customer relationships, and CSMAR controls are aligned by entity and fiscal year and assembled into a firm-partner-year panel. The implemented branch feeds observed aggregate features to the fixed-effects model. A proposed raw-text branch would deduplicate reports, resolve firms and plants, classify events and tone, and apply confidence-based abstention before any output enters the panel. Operationally, the workflow freezes data versions and entity rules, aligns directed records, estimates the low-dimensional panel, and reserves raw-text training for a later labeled-data implementation. This design keeps text-model metrics

separate from panel coefficients and routes any operational signal to human review.

3.2 Problem Definition and Notation

Let i denote a focal listed firm, j its disclosed supplier or customer, and t the fiscal year. The unit of observation is an observed directed edge-year (i, j, t) , not an isolated news article. The dependent variable is the partner's contemporaneous digital-disclosure index in year t , consistent with the reported baseline model. This same-period specification does not eliminate reverse causality or omitted-variable concerns. Table 1 defines the notation used below.

Table 1. Core notation for the proposed method

Symbol	Definition	Role
i, j, t	Focal firm, connected partner, and year	Edge-year index
$a_{ij,t}$	Indicator for a disclosed supplier/customer edge	Supply-chain link indicator
$NMA_{i,t}$	Focal-firm online media coverage feature	Main explanatory variable
$f_{i,t}$	Branch-specific firm-year text feature: aggregate $C_{i,t}$ in the implemented route or raw-text $z_{i,t}$ in the conditional route	Input to the directed edge representation
$z_{i,t}$	Accepted firm-year event-and-tone feature vector from a trained and validated conditional raw-text route	Future optional feature; not observed in this study
$DC_{j,t}$	Partner digital-disclosure index	Dependent variable / capability proxy
$X_{i,t}, X_{j,t}$	Focal-firm and partner control vectors	Covariates
μ_i, τ_t	Focal-firm and year fixed effects	Fixed effects
$NMA2_{j,t}$	Partner online media coverage feature	Partner coverage variable in the media attention spillover channel

3.3 Implemented Text Measure and Optional Raw-Text Branch

The reported analysis uses CFND focal-firm coverage counts and database-provided positive and neutral categories.

Annual-report counts across the five digital categories are aggregated and log-transformed in Equation (1). These observed aggregate measures, rather than neural-model outputs, enter the regressions in Section 4.

$$DC_{j,t} = \ln(1 + \sum_{k=1}^5 c_{j,t,k}), k \in \{\text{applications, blockchain, cloud, AI, big data}\} \quad (1)$$

Here c denotes matched dictionary frequency. If raw articles and human-reviewed labels become available, an optional measurement module could deduplicate reports, resolve firm and plant aliases, encode articles with an established Chinese-BERT model, and use compact MLP heads to estimate event and tone probabilities. After confidence screening and human review, accepted article-level outputs would be aggregated by firm and fiscal year into $z_{i,t}$, the conditional raw-text event-and-tone feature vector. Low-confidence assignments would be withheld from $z_{i,t}$ and

routed to human review. This is an interface specification only: no Chinese-BERT–MLP model was trained, no raw-text prototype or model-validation experiment was conducted, no classification result is reported, and no BERT-derived feature enters the present panel analysis.

Algorithm 1 retains explicit if/else/end if and for/end for boundaries so that the optional raw-text route is visibly separated from the implemented aggregate-data route before directed edge-year assembly.

Algorithm 1: Conditional text-feature construction and directed-edge panel assembly

Input: aggregate dmedia features C ; optional raw corpus D and labeled set Y ; alias dictionary A ; directed edges E ; annual reports R ; controls X .

Output: branch-specific firm-year features $\{f_{i,t}\}$; edge-year signals $\{s_{i \rightarrow j,t}\}$; analysis panel P .

- 1 Freeze database versions, A , E , R , fiscal-year rules, thresholds, and splits.
- 2 if raw corpus D and human-reviewed labels Y are available then
- 3 | Deduplicate D and keep all reports concerning the same event in one split.
- 4 | Train Chinese-BERT–MLP parameters on the training partition only.

```

5      |      for each article  $d_m \in D$  do
6      | |      Align  $d_m$  to fiscal year  $t$  and resolve firm, plant, and supplier aliases.
7      | |      if entity-match confidence  $< \delta$  then
8      | | |      Mark  $d_m$  as abstained.
9      | |      else
10     | | |      Compute event/tone probabilities and retain model metadata.
11     | |      end if
12     |      end for
13     |      Aggregate accepted article-level event/tone outputs as  $z_{i,t}$ ; set  $f_{i,t} = z_{i,t}$ .
14     |      else
15     | |      Set  $f_{i,t} = C_{i,t}$  from the available database-supplied aggregate features.
16     |      end if
17     |      for each directed edge  $(i,j,t) \in E$  do
18     | |      Construct the directed edge-year signal; merge partner disclosure and both firms' controls into  $P$ .
19     |      end for
20     |      Estimate Equation (3) on  $P$ .
21     |      Keep text-model metrics separate from panel coefficients.
22     |      return  $\{f_{i,t}\}$ ,  $\{s_{i \rightarrow j,t}\}$ , and  $P$ .

```

3.4 Edge-Conditioned Spillover Representation

The network module retains only text signals attached to an observed supply-chain edge and preserves edge direction. Let $f_{i,t}$ denote the branch-specific firm-year feature: $f_{i,t} = C_{i,t}$ for the implemented database-supplied aggregate route, whereas $f_{i,t} = z_{i,t}$ only if a future Chinese-BERT-MLP route is trained and validated; $z_{i,t}$ denotes its accepted firm-year event-and-tone feature vector after confidence screening and human review. The representation passed to the estimator is

$$s_{i \rightarrow j,t} = a_{ij,t} \cdot h(f_{i,t}), \quad (2)$$

where h is the identity map for the implemented aggregate coverage feature or a fitted projection available only after a raw-text model has been trained and validated. Edge type, transaction rank, or disclosed purchase/sales share can be appended when available. The present data package contains relationship links but does not provide a complete set of continuously observed transaction weights, so the implemented analysis uses an unweighted edge indicator. This simplification treats disclosed links as equally informative even though production-network research indicates that input specificity and partner substitutability affect shock propagation [1–3]. Transaction shares and product-text similarity [70] would therefore be useful weighting variables in a richer implementation.

3.5 Baseline Panel Specification

The core low-dimensional model estimates the contemporaneous partner digital-disclosure index as a function of focal-firm coverage, focal and partner controls,

and fixed effects:

$$DC_{j,t} = \beta NMA_{i,t} + \gamma^T X_{i,t} + \eta^T X_{j,t} + \mu_i + \tau_t + \varepsilon_{ijt}. \quad (3)$$

The coefficient β is estimated from within-focal-firm changes in reported coverage after accounting for common year shocks and the listed controls. It is therefore a conditional association between exposure at i and the proxy observed at j , not a causal effect. Equation (3) follows the reported model description by assigning firm fixed effects to the focal firm. The available regression summaries do not reveal whether the implemented firm effect is indexed by i , j , or the pair (i, j) , or how standard errors were clustered; inference should therefore be interpreted cautiously. Repeated focal firms, repeated partners, and partners appearing on several edges can create within-network dependence; clustering at the focal-firm level, partner level, or two-way across both dimensions may be appropriate depending on the actual panel structure. These unresolved choices are treated as an identification limitation rather than inferred from the coefficient table.

The fixed-effect choices are not interchangeable. μ_i absorbs time-invariant focal-firm heterogeneity; μ_j absorbs time-invariant partner heterogeneity; $\mu_i + \mu_j$ controls stable characteristics of both endpoints but not pair-specific matching; and μ_{ij} absorbs all persistent attributes of the focal-partner match and identifies β from within-edge changes. The reported specification retains μ_i and year effects because the alternatives cannot be re-estimated from the available summaries. Because $NMA_{i,t}$ varies at the focal-firm-year level and is repeated across i 's edges, focal-firm clustering is the primary recommended inference rule. When repeated partners create dependence through j and cluster counts are adequate, two-way clustering by i and j should be reported as a robustness check.

3.6 Mechanism Specifications

The interfirm digital technology diffusion channel is assessed by first relating focal-firm media exposure to the focal-firm digital-disclosure index and then relating the focal-firm index to the partner digital-disclosure index. The media attention spillover channel is assessed by relating focal-firm coverage to partner coverage and partner coverage to the partner digital-disclosure index. With controls and fixed effects suppressed for readability, the two paths are

$$\begin{aligned} DC_{i,t} &= a_1 NMA_{i,t} + u_{i,t}, DC_{j,t} = b_1 DC_{i,t} + v_{ijt}; \\ NMA_{2,j,t} &= a_2 NMA_{i,t} + e_{ijt}, DC_{j,t} = b_2 NMA_{2,j,t} + \xi_{ijt}. \end{aligned} \quad (4)$$

Table 2 compactly summarizes the workflow described in Section 3.1 and its evidence boundary.

Table 2. Compact processing flow and evidence boundaries

Stage	Input	Operation	Output
Alignment	Versions; names; dates	Freeze rules; match entities and fiscal years	Firm-year keys
Observed panel	Coverage; reports; edges	Aggregate observed text; construct directed edge-year records	NMA; NMA2; DC; $a_{ij,t}$
Estimation	Panel; controls	Fixed effects; mechanisms; descriptive subgroups	β ; SE; N; R^2
Future raw text	Articles; labels	Deduplicate; classify events/tones; abstain	Event/tones; z
Deployment	Estimates; events	Human SRM/MES/ERP review	Auditable alerts

4. Empirical Results and Evidence Boundaries

4.1 Application Setting and Evidence Boundary

The framework is applied to the documented Chinese A-share supply-chain panel for 2007–2024. The sampling procedure

excludes financial firms, firms under special-treatment status, records missing core variables, and non-listed partners; focal firms are matched to disclosed major suppliers or customers. CFND supplies aggregate firm-year coverage and tone categories, annual reports supply digital-term counts, CNRDS supplies directed links, and CSMAR supplies controls. Article text, reviewed labels, row-level estimation records, and predefined holdouts are not available for the present empirical evidence assessment. Numeric checking is therefore limited to the reported tables, and the proposed raw-text branch is not treated as a trained experiment.

The application is manufacturing-oriented because the architecture and deployment scenarios concern production, quality, delivery, traceability, and industrial data interfaces. The available empirical sample, however, is a broad non-financial A-share panel rather than a verified manufacturing-only subsample. Accordingly, the reported coefficients are interpreted as manufacturing-oriented evidence from a broad non-financial sample, not as estimates from a manufacturing-only sample. A sector-specific validation remains an important direction for future research.

The exclusions improve comparability but also shape interpretation. Restricting links to listed focal firms and listed major partners omits many small private suppliers that may have weaker information systems and different exposure to media attention. Using the top five disclosed suppliers or customers emphasizes economically salient relationships, yet disclosure thresholds and partner turnover can make the observed network incomplete. The resulting panel is suitable for examining whether media visibility and digital orientation move together across disclosed supply-chain links; it is not a census of the production network. Table 3 therefore records both the empirical role of each source and the evidence that cannot be independently reconstructed from the aggregate materials.

Table 3. Data sources, empirical roles, and evidence boundaries

Source	Available feature	Empirical role	Evidence boundary
CFND	Online media coverage count; positive and neutral categories	Media exposure and attention features	Aggregates available; AI event/tone model cannot be trained from current package
Listed-firm annual reports	Five-category digital-term frequency	Focal and partner digital-disclosure indices	Disclosure-based capability proxies; no transaction-level collaboration log
CNRDS	Major supplier/customer relationship	Directed edge-year construction	Continuous edge weights unavailable
CSMAR	Firm age, largest-shareholder ownership share, ROA	Focal and partner controls	Complete-case count differs by specification
Reported regression summaries	Coefficients, standard errors, N, R ²	Summary of empirical evidence	No independent row-level re-estimation possible

4.2 Variables and Sample Accounting

Table 4 defines the notation used in the reported empirical results: the suffix 2 identifies the connected upstream or downstream partner. Thus DT is the focal firm's digital-disclosure index, DT2 is the partner's corresponding index, NMA is focal-firm coverage, and NMA2 is partner coverage. Section 3 uses DC as a conceptual label for DT2 to emphasize that it is the dependent digital-disclosure index and a disclosure-based proxy for capability. The change in notation does not alter any reported value. Control variables are constructed for both sides of the relationship so that the estimated NMA coefficient is not simply a comparison among firms differing in age, ownership concentration, or contemporaneous profitability.

Variable	Operational definition	Interpretation in the framework
NMA_NEU	Focal-firm neutral-news count	Alternative exposure feature
NMA2	Partner online media coverage count from CFND	Variable in the media attention spillover channel
AGE / AGE2	ln(1 + years since listing) for focal firm / partner	Maturity controls
TOP / TOP2	Largest-shareholder ratio of focal firm / partner	Ownership-concentration controls
ROA / ROA2	Return on assets of focal firm / partner	Performance controls

Table 4. Variable definitions

Variable	Operational definition	Interpretation in the framework
DT	Log-transformed focal-firm annual-report count across five digital categories following [77]	Focal digital-disclosure index / diffusion-channel variable
DT2	Log-transformed partner annual-report count across the same five categories following [77]	Partner digital-disclosure index (DC); capability proxy
NMA	Focal-firm online media coverage count from CFND	Primary media exposure feature
NMA_POS	Focal-firm positive-news count	Alternative exposure feature

Table 5 reports 4,117 observations for each listed variable. The full two-way fixed-effects model in Table 6 contains 3,700 complete observations after matching the partner outcome and controls. The industry-fixed-effects robustness model contains 4,112 observations, and subgroup totals vary because the reported classifications and complete-case filters do not retain identical records. These sample differences may affect comparability and are treated as a limitation. DT and DT2 are digital-disclosure indices, whereas NMA and NMA2 are media-coverage counts; these constructs should not be interchanged when reading the descriptive distribution. Table 5 reports NMA in count units, but the coefficient summaries do not document any additional estimation scaling. The magnitude of β is therefore not interpreted as a percentage or as a causal response to one additional article.

Table 5. Descriptive statistics

Variable	N	Mean	SD	Min.	Max.
DT	4,117	1.106	1.356	0.000	6.082
DT2	4,117	1.206	1.359	0.000	5.966
NMA	4,117	168.9	186.6	0.000	2,095
NMA2	4,117	1,315	3,218	0.000	49,431
AGE	4,117	1.949	0.999	0.000	3.497
AGE2	4,117	2.381	0.778	0.000	3.497
TOP	4,117	0.355	0.150	0.0225	0.900
TOP2	4,117	0.408	0.186	0.0184	0.900
ROA	4,117	0.0342	0.0855	-1.859	0.545
ROA2	4,117	0.0437	0.0593	-0.488	0.695

As shown in Table 5, the descriptive patterns have two implications for estimation. First, the wide dispersion and high maxima of NMA and especially NMA2 indicate that media visibility is highly concentrated among a subset of listed firms. A small number of prominent partners can therefore influence an untransformed coverage measure, making alternative sentiment-specific counts useful as specification checks. Second, the partner digital index has a slightly higher mean than the focal index but a similar standard deviation; this does not demonstrate diffusion because the variables refer to matched firms and the comparison is unconditional. The regression design instead uses within-firm changes and firm and year fixed effects. Finally, the decline from 4,117 descriptive observations to 3,700 complete cases raises a potential selection issue: missing partner controls or outcomes may be related to firm size, disclosure, or network visibility. Row-level diagnostics should compare retained and omitted records when the underlying data become available.

Complete-case estimation assumes that retention is not systematically related to unobserved determinants of partner disclosure after conditioning on the included variables. A row-level reassessment should compare retained and omitted records by media exposure, digital disclosure, firm size, industry, year, ownership, region, and edge type; it should model the retention indicator and, where appropriate, compare inverse-probability-weighted estimates. The same row-level reassessment should re-estimate the baseline after 1st/99th-percentile winsorization of $NMA_{i,t}$ and $NMA2_{j,t}$ and separately examine $\ln(1 + NMA_{i,t})$ and $\ln(1 + NMA2_{j,t})$, using the same sample and clustering rule. These diagnostics cannot be reconstructed from the available summaries, so no claim of selection neutrality or outlier robustness is made.

4.3 Baseline Results and Specification Stability

Table 6 adds controls in blocks while retaining firm and year fixed effects. Across its four specifications, the focal coefficient ranges only from 0.047 to 0.049 and its reported standard error remains 0.020. R^2 rises from 0.688 to 0.689. The fully controlled model therefore shows stable sign and magnitude across the nested specifications: $\beta = 0.048$, $SE = 0.020$, $p < 0.05$, $N = 3,700$, and $R^2 = 0.689$. Because the dependent variable is a log-transformed digital-disclosure index and the scaling of NMA in the available estimation summaries is not independently recoverable, the coefficient is not converted into an unsupported percentage effect. All four reported specifications include firm and year fixed effects; one, two, and three stars denote $p < 0.10$, $p < 0.05$, and $p < 0.01$, respectively. The significance symbols and standard errors follow the reported results. Because the clustering convention is not stated, the inferential precision should be interpreted cautiously, as discussed in Section 3.5.

Table 6. Nested fixed-effects models and coefficient stability

Specification	Controls added	β on NMA	SE	N	R^2
M1	Fixed effects only	0.049**	0.020	3,700	0.688
M2	+ AGE and AGE2	0.049**	0.020	3,700	0.688
M3	+ TOP and TOP2	0.047**	0.020	3,700	0.689
M4	+ ROA and ROA2	0.048**	0.020	3,700	0.689

The in-sample R^2 values in Table 6 evaluate only the reported aggregate-data regressions; they are not performance measures for the optional raw-text module.

Economically, the stability of β across specifications M1–M4 in Table 6 after age, ownership concentration, and profitability enter the equation suggests that the positive relationship is not driven solely by these observed differences between focal firms and partners. The small increase in R^2 is unsurprising because firm and year effects already absorb substantial persistent and common variation. At the same time, coefficient stability is not a substitute for identification: firms expecting a digital upgrade may attract more coverage, and an unobserved supply-chain event may simultaneously increase media attention and partner disclosure. The baseline result is therefore best read as evidence that media visibility contains partner-relevant information after the listed controls are applied. A stronger design would use lagged exposure,

event timing, an external source of media variation, or transaction-level outcomes.

4.4 Robustness and Descriptive Heterogeneity Analysis

The first three rows of Table 7, also plotted with confidence intervals in Figure 2, report checks that alter the fixed-effects structure or media-coverage measure. Replacing firm fixed effects with industry fixed effects yields $\beta = 0.027$ (SE = 0.014, $p < 0.10$; $N = 4,112$; $R^2 = 0.427$). Replacing total coverage with positive-news coverage yields $\beta = 0.052$ (SE = 0.020, $p < 0.01$), while replacing it with neutral-news coverage yields $\beta = 0.038$ (SE = 0.019, $p < 0.05$); both use $N = 3,700$ and $R^2 = 0.689$. The sign is stable, but the weaker industry-fixed-effects estimate and changed sample require cautious comparison.

The subgroup rows in Table 7 and Figure 2 assess economic heterogeneity rather than out-of-sample predictive generalization. The association is positive and statistically distinguishable from zero for downstream partners, state-owned partners, and firms in central and western China, but not for upstream, privately owned, or eastern groups. A non-significant subgroup estimate is not proof of a zero effect, and coefficient differences should ultimately be tested in a pooled interaction model. Figure 2 displays approximate 95% intervals calculated as coefficient $\pm 1.96 \times$ standard error, and

Table 7 reports the available sample counts. One, two, and three stars denote $p < 0.10$, $p < 0.05$, and $p < 0.01$. Group totals differ because the reported classifications and complete-case filters do not retain identical records.

For a subgroup indicator G_{ij} , a formal comparison requires a pooled model containing $NMA_{i,t}$, G_{ij} , and $NMA_{i,t} \times G_{ij}$, followed by a test of $H_0: \theta = 0$ for the interaction coefficient θ . The same logic applies to supply-chain position, ownership, and region. Row-level records were unavailable for this revision, so Table 7 is interpreted as descriptive subgroup coefficients rather than evidence of statistically established between-group differences.

The robustness estimates in Table 7, together with their intervals in Figure 2, show positive coefficients for both positive and neutral coverage; this pattern is more consistent with a visibility or information channel than with an interpretation driven exclusively by adverse sentiment. Positive coverage can publicize digital practices or interface requirements, whereas neutral coverage can still make a production event and its connected firms easier to discover. This interpretation remains provisional because the positive and neutral counts may be correlated with total coverage and firm size. A raw-text analysis should jointly model tone, volume, novelty, and event type rather than treating each aggregate count as an isolated shock.

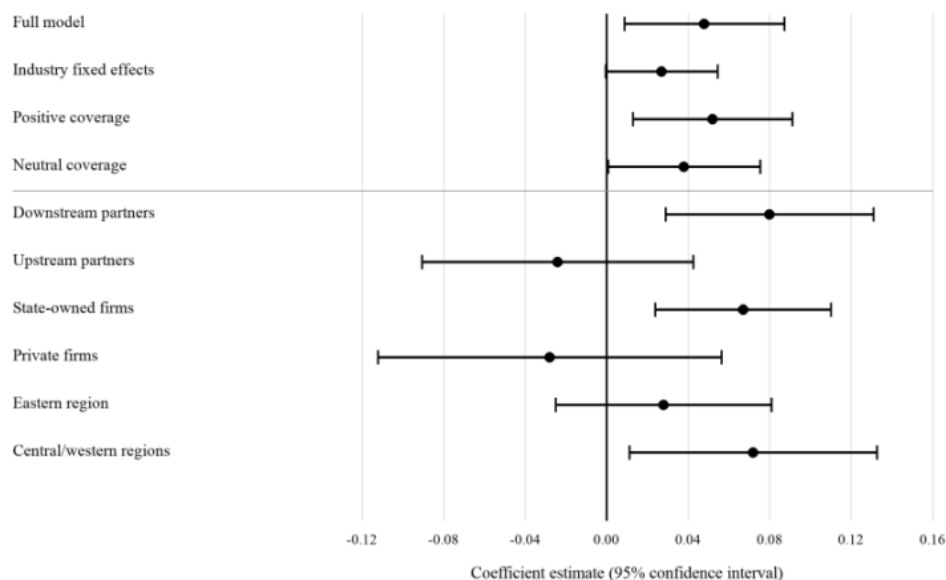


Figure 2. Forest plot of coefficient estimates across robustness and heterogeneity specifications. Error bars are approximate 95% confidence intervals calculated as the estimate $\pm 1.96 \times$ standard error. Subgroup estimates are descriptive comparisons; no formal tests of between-group coefficient differences are reported

Table 7. Robustness and descriptive subgroup estimates

Evidence block	Specification/group	Coefficient	SE	N	R ²
Robustness	Industry FE	0.027*	0.014	4,112	0.427
Robustness	Positive-news count	0.052***	0.020	3,700	0.689
Robustness	Neutral-news count	0.038**	0.019	3,700	0.689
Position	Downstream partner	0.080***	0.026	2,184	0.741
Position	Upstream partner	-0.024	0.034	1,305	0.704
Ownership	State-owned partner	0.067***	0.022	2,062	0.737
Ownership	Privately owned partner	-0.028	0.043	1,401	0.747
Region	Eastern	0.028	0.027	2,307	0.721
Region	Central and western	0.072**	0.031	1,393	0.635

The position rows of Table 7 and their intervals in Figure 2 show a positive downstream estimate and an imprecise upstream estimate. This pattern is economically consistent with customer-facing firms having stronger incentives to protect delivery reliability, product quality, and reputation. When a focal supplier receives coverage, downstream manufacturers may demand richer inspection records, electronic batch genealogy, and real-time delivery visibility. Upstream suppliers, especially small firms, may face the same information demand but lack bargaining power or complementary assets needed for rapid digital investment. Production-network evidence likewise indicates that propagation depends on relationship specificity and substitution possibilities [2,3]. Because the two coefficients were estimated on different subsamples, their contrast is suggestive rather than a formal test that the downstream coefficient exceeds the upstream coefficient.

The ownership and region rows of Table 7, also plotted in Figure 2, show positive estimates for state-owned partners and for central and western firms. These estimates may reflect greater access to digitalization resources, policy-compliance obligations, or more formal reporting systems, while the central and western estimate may reflect greater scope for digital catch-up. These mechanisms cannot be separated with the available variables. Likewise, statistical significance in one group and non-significance in another does not establish a significant between-group difference. Establishing

between-group differences would require pooled interaction terms estimated on a common sample and tests of the relevant linear combinations. The heterogeneity results are therefore used to describe where the average association appears concentrated, not to claim universal structural laws across ownership and regions.

4.5 Mechanism Tests

Table 8 reports two sequential links for each proposed mechanism. The interfirm digital technology diffusion channel is consistent with two positive links: focal coverage is positively associated with the focal-firm digital-disclosure index (0.038, SE = 0.015, $p < 0.01$), and the focal-firm index is positively associated with the partner digital-disclosure index (0.206, SE = 0.025, $p < 0.01$). The media attention spillover channel is likewise consistent with focal coverage being positively associated with partner coverage (0.053, SE = 0.025, $p < 0.05$) and partner coverage being positively associated with the partner digital-disclosure index (0.035, SE = 0.015, $p < 0.05$). The specification including partner coverage also retains a direct NMA coefficient of 0.046 (SE = 0.020, $p < 0.05$). All mechanism regressions include the reported controls and firm and year fixed effects. These are sequential associations consistent with the proposed channels, not a formally identified causal mediation effect.

Table 8. Mechanism-test estimates

Path	Outcome	Predictor	Coefficient	SE	N	R ²
Interfirm digital technology diffusion channel (a_1)	DT	NMA	0.038***	0.015	3,700	0.838
Interfirm digital technology diffusion channel (b_1)	DT2	DT	0.206***	0.025	3,700	0.695
Media attention spillover channel (a_2)	NMA2	NMA	0.053**	0.025	3,700	0.538
Media attention spillover channel (b_2)	DT2	NMA2	0.035**	0.015	3,700	0.689

The first two rows of Table 8 describe the interfirm digital technology diffusion channel. This channel is economically plausible because digital supply-chain benefits arise from compatible routines rather than isolated hardware. A focal firm's reporting, traceability, or interface investment can increase the return to similar investment by its partner, especially when transactions are repeated and process specifications are relationship-specific. Evidence on digitally enabled integration and analytics likewise emphasizes complementarities among information exchange, visibility, flexibility, and organizational capability [24,26,31]. The positive coefficient linking DT to DT2 is consistent with correlated digital orientation across connected firms, although common investment opportunities and assortative matching could produce a similar pattern.

The last two rows of Table 8 describe the media attention spillover channel, which treats the supply-chain relation as an information-search path. Once a focal firm becomes salient, journalists, analysts, regulators, and customers can more easily identify its disclosed suppliers or buyers. Prior financial-media research shows that business coverage changes the information available to market participants and reallocates attention [11,59,60]. Greater partner coverage may then raise the value of auditable digital responses, such as traceability records or delivery dashboards. Yet NMA2 may also capture partner size, disclosure propensity, or a common event that generates coverage of both firms. Establishing mediation would require temporal ordering, a direct test of the indirect effect, and stronger controls for shared news events.

For transparency, Appendix A retains a parameter-level illustration based on independent normal draws calibrated to the reported coefficient and standard error. It is not bootstrap resampling, does not create additional observations, and is

not used as a main empirical result or as evidence of robustness.

4.6 Manufacturing Deployment Scenarios and Current Boundaries

The implemented Stage 1 is a retrospective, offline analysis of licensed aggregate data; it cannot issue real-time event warnings. Its estimates can inform research and later system design, but they are not outputs from an operating industrial service. If raw articles, reviewed labels, and stable data interfaces become available, a later-stage monitoring prototype could deduplicate reports, resolve firms and plants, classify events and tone, abstain on low-confidence cases, and route auditable alerts to human reviewers. Table 9 presents these scenarios as conditional design examples rather than current capabilities.

The three rows of Table 9 distinguish AI-routed quality, delivery, and environmental scenarios. The classifier determines review routing rather than an automatic operational action. In the quality-incident row, new-media coverage is useful because it can provide an early cue before the focal manufacturer has reconciled every lot and supplier record. The cue should trigger entity and event verification, after which batch genealogy, inspection images, nonconformance reports, and corrective-action status can be requested through existing systems. For a delivery interruption, the same logic supports a review of lead-time assumptions, safety stock, and alternative-source readiness. An environmental event requires a different workflow centered on plant identity, affected materials, permits, and legal review. In each case the media variable prioritizes attention; it does not replace operational evidence.

Table 9. Manufacturing-oriented application scenarios and safeguards

Scenario	Proposed AI-assisted signal	Digital collaboration response	Operational safeguard
Quality incident	AI-classified defect reports linked to a component supplier	Request batch genealogy, inspection images, and corrective-action status through SRM/MES	Abstain at low confidence; verify entity and lot; require quality-engineer approval
Delivery disruption	AI-classified delay or shutdown linked to a high-dependency supplier	Refresh lead-time assumptions, inventory buffers, and alternative-source readiness in ERP	Abstain at low confidence; combine signal with purchase orders and logistics data
Environmental/ESG event	AI-classified regulatory or community report involving a connected plant	Collect compliance evidence and trace affected materials across the partner network	Abstain at low confidence; no automatic suspension; legal/compliance review required

Practical implementation faces substantial data, organizational, and governance barriers. Replication depends on continued access to CFND, CNRDS, and CSMAR, their licensing terms, stable interfaces, coverage consistency, and

reliable cross-database entity and plant matching. Unlike prior systems that extract supply-chain relations from news or ingest social-media streams [43,44], Stage 1 has neither document-level input nor continuous ingestion. A raw-text

branch would require lawful article access, a maintained annotation scheme, representative human-reviewed labels, document-level event evaluation [45], and governed deferral of uncertain cases to qualified reviewers [46]. Real-time service would also require event-time and trigger semantics [47], false-alert and missed-event auditing, secure SRM/MES/ERP integration, privacy controls, drift monitoring, version management, computing capacity, staffing, and recurring maintenance. Media undercoverage of small suppliers and publicity bias would remain material risks. For these reasons, the aggregate branch is currently feasible only for offline analysis, whereas real-time industrial monitoring remains a conditional long-term objective and Chinese-BERT-MLP remains an untrained design without a prototype or simulation-based validation.

The operational value of an alert also depends on sector-specific production economics. In automotive assembly, a low-cost component can be economically critical when it has a long qualification cycle, a single approved source, or safety implications; media coverage linked to that supplier should therefore be combined with bill-of-materials position, line-side inventory, lot genealogy, and the time required to approve an alternative. In electronics manufacturing, shorter product cycles and multi-tier subcontracting make plant identity, tooling location, and component substitutability especially important. An interruption reported for one facility should not automatically be assigned to every legal entity in the corporate group. Equipment manufacturers face a different pattern because engineered-to-order products often depend on customized parts and milestone delivery; even a modest delay can propagate to installation, acceptance, and customer-payment schedules. These examples imply that the model output should enter a risk-prioritization rule together with operational severity, relationship exposure, substitutability, and entity-matching confidence. The media feature supplies an external information cue, whereas purchase orders, quality records, and production constraints determine the expected economic consequence. This separation limits popularity bias and makes the intervention auditable for procurement, operations, quality, and compliance teams.

Any future prototype evaluation should distinguish statistical fit from decision value. The first layer concerns signal quality: entity-matching precision, event-classification precision and recall, false-alert rates, and the lead time between the first verified report and an internal operational record. The second layer concerns supply-chain decisions: the share of alerts that trigger a justified review, time saved in locating affected lots, changes in contingency lead time, and reviewer workload. The third layer concerns economic outcomes, including avoided downtime, expedited-freight costs, inventory carrying costs, requalification expenses, and the losses created by unnecessary interventions. Evaluation should use a time-based holdout so that articles and events from later periods are not represented in model development, and all documents concerning the same event should remain in the same split to prevent information leakage. A simple benchmark based on coverage volume and supplier criticality should be compared with any elastic-net or Chinese-BERT-

MLP extension. Thresholds can then be selected according to the relative costs of missed disruptions and false reviews rather than by maximizing a classification metric alone. Such a protocol would convert the proposed architecture into a genuine out-of-sample experiment while preserving the economic distinction between predictive usefulness and causal identification.

If a prototype is eventually deployed, AI governance must continue throughout operation. Manufacturing vocabulary, corporate aliases, supplier relationships, and media practices change over time, so the entity dictionary and event classifier should be versioned and monitored for drift. Confirmed and rejected alerts can provide feedback labels, but reviewer decisions should be sampled for consistency before they are reused for training. Periodic audits should examine whether error rates differ by firm size, ownership, region, or supply-chain tier, because systematically under-covered smaller suppliers may carry high operational risk. Access controls should separate public media evidence from confidential order and quality data, and every recommended action should retain the source article, relationship edge, feature vector, model version, and human decision. These controls support accountability without treating media visibility as a substitute for supplier due diligence.

5. Conclusion

This study documents a positive observational association between focal-firm media coverage and connected partners' annual-report digital-disclosure indices in Chinese non-financial A-share supply-chain links. Its system contribution is a human-supervised workflow that maps external text signals to directed firm-partner-year relations while separating measurement from panel inference. The implemented aggregate-data stage is retrospective and dependent on licensed databases; it is not a real-time warning system. Mechanism paths remain sequential associations, subgroup coefficients remain descriptive, and annual-report word frequency measures disclosure rather than realized collaboration. Any raw-text Chinese-BERT-MLP service requires article access, reviewed labels, prototype development, out-of-sample validation, operational integration, and continued human oversight. The present evidence therefore supports a staged research architecture and a bounded empirical association, not a claim of completed industrial deployment.

Appendix A. Parameter-Level Monte Carlo Illustration

This appendix reports a limited parameter demonstration retained for transparency. It draws 2,000 independent values from a normal distribution calibrated only to the reported full-model coefficient and standard error:

$$\beta^{(b)} \sim N(0.048, 0.020^2), b = 1, \dots, 2000. \quad (A1)$$

Figure A1 displays the resulting normal-approximation distribution, and Table A1 reports its reproducible numerical summary. With seed 20260721, the simulated mean is 0.0477, the standard deviation is 0.0200, the central 95% interval is [0.0088, 0.0875], and 99.15% of draws are

positive. These quantities follow mechanically from independent draws with $\mu = 0.048$ and $\sigma = 0.020$; they are not additional firm observations.

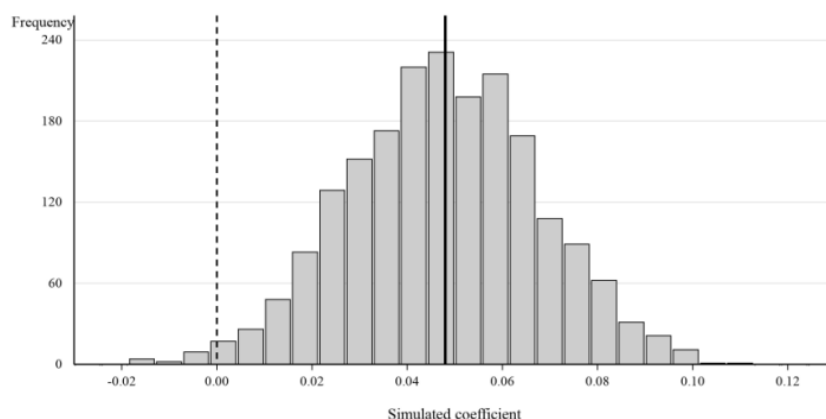


Figure A1. Parameter-level Monte Carlo approximation calibrated to the reported full-model coefficient and standard error. The plotted values are independent normal draws, not bootstrap resamples or additional empirical observations.

Table A1. Reproducible parameter-simulation summary

Seed	B	Mean	SD	Med.	q.025	q.975	Pr($\beta > 0$)
20260721	2,000	0.0477	0.0200	0.0481	0.0088	0.0875	99.15%

The illustration answers only how much sign uncertainty is implied by the reported coefficient and standard error under the assumed normal approximation. It does not evaluate temporal generalization, alternative samples, network dependence, or text-measurement quality and therefore is not treated as independent validation.

An empirical uncertainty analysis would require row-level resampling that preserves edge-year dependence. A focal-firm block bootstrap would resample each focal firm with all associated partners and years before re-estimating Equation (3); multiway or wild-cluster methods may also be appropriate when both endpoints induce dependence. The row-level panel needed for that exercise is unavailable, so no bootstrap interval is reported.

Declarations

Data availability

The empirical variables draw on licensed CFND, CNRDS, and CSMAR databases. Access to the underlying records is governed by the respective database licenses; aggregate results are reported in the manuscript.

Author contributions

C.L. conceived the study, developed the methodology, conducted the formal analysis, curated the data, and wrote and revised the manuscript.

References

- [1] Acemoglu, D., Carvalho, V. M., Ozdaglar, A., & Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80(5), 1977–2016. <https://doi.org/10.3982/ECTA9623>
- [2] Barrot, J.-N., & Sauvagnat, J. (2016). Input specificity and the propagation of idiosyncratic shocks in production networks. *The Quarterly Journal of Economics*, 131(3), 1543–1592. <https://doi.org/10.1093/qje/qjw018>
- [3] Carvalho, V. M., Nirei, M., Saito, Y. U., & Tahbaz-Salehi, A. (2021). Supply chain disruptions: Evidence from the Great East Japan Earthquake. *The Quarterly Journal of Economics*, 136(2), 1255–1321. <https://doi.org/10.1093/qje/qjaa044>
- [4] Plekhanov, D., Franke, H., & Netland, T. H. (2023). Digital transformation: A review and research agenda. *European Management Journal*, 41(6), 821–844. <https://doi.org/10.1016/j.emj.2022.09.007>

- [5] Porter, M. E., & Heppelmann, J. E. (2014). How smart, connected products are transforming competition. *Harvard Business Review*, 92(11), 64–88.
- [6] Duraivelu, K. (2022). Digital transformation in manufacturing industry—A comprehensive insight. *Materials Today: Proceedings*, 68(Part 6), 1825–1829. <https://doi.org/10.1016/j.matpr.2022.07.409>
- [7] Lee, S., & Valenzuela, S. (2024). A self-righteous, not a virtuous, circle: Proposing a new framework for studying media effects on knowledge and political participation in a social media environment. *Social Media + Society*, 10(2), 20563051241257953. <https://doi.org/10.1177/20563051241257953>
- [8] Mäder, A., & Rinsdorf, L. (2023). Constructive journalism as an adaptation to a changing media environment. *Journalism Studies*, 24(3), 329–346. <https://doi.org/10.1080/1461670X.2022.2159859>
- [9] Zilic Fiser, S., & Caks, P. (2023). Strategies for the minimisation of misinformation spread through the local media environment. *Journalism Practice*, 17(10), 2241–2262. <https://doi.org/10.1080/17512786.2023.2183235>
- [10] Kraft, P. W., Krupnikov, Y., Milita, K., Ryan, J. B., & Soroka, S. (2020). Social media and the changing information environment: Sentiment differences in read versus recirculated news content. *Public Opinion Quarterly*, 84(S1), 195–215. <https://doi.org/10.1093/poq/nfaa015>
- [11] Bushee, B. J., Core, J. E., Guay, W., & Hamm, S. J. W. (2010). The role of the business press as an information intermediary. *Journal of Accounting Research*, 48(1), 1–19. <https://doi.org/10.1111/j.1475-679X.2009.00357.x>
- [12] He, G., & Li, A. Z. (2024). Does media coverage of firms' environment, social, and governance (ESG) incidents affect analyst coverage and forecasts? A risk perspective. *International Review of Financial Analysis*, 94, 103289. <https://doi.org/10.1016/j.irfa.2024.103289>
- [13] Gao, J., Wu, H., You, J., & Smith, M. (2025). Green media coverage and corporate green innovation. *Journal of Business Finance & Accounting*, 52(1), 48–90. <https://doi.org/10.1111/jbfa.12802>
- [14] Sun, J., Hou, S., Deng, Y., & Li, H. (2024). New media environment, green technological innovation and corporate productivity: Evidence from listed companies in China. *Energy Economics*, 131, 107395. <https://doi.org/10.1016/j.eneco.2024.107395>
- [15] Li, Z., Huang, Z., & Su, Y. (2023). New media environment, environmental regulation and corporate green technology innovation: Evidence from China. *Energy Economics*, 119, 106545. <https://doi.org/10.1016/j.eneco.2023.106545>
- [16] Manservigi, F., Banzi, M., Tonelli, T., Veronesi, P., Ricci, S., Distanti, D., Faralli, S., & Bortone, G. (2023). Environmental complaint insights through text mining based on the driver, pressure, state, impact, and response (DPSIR) framework: Evidence from an Italian environmental agency. *Regional Sustainability*, 4(3), 261–281. [https://doi.org/10.1016/j.regus.2023.08.002](https://doi.org/10.1016/j.regsus.2023.08.002)
- [17] Zhu, Y., & Zhang, Z. (2024). Supply chain digitalization and corporate ESG performance: Evidence from supply chain innovation and application pilot policy. *Finance Research Letters*, 67, 105818. <https://doi.org/10.1016/j.frl.2024.105818>
- [18] Angelopoulos, S., Bendoly, E., Fransoo, J. C., Hoberg, K., Ou, C., & Tenhiälä, A. (2023). Digital transformation in operations management: Fundamental change through agency reversal. *Journal of Operations Management*, 69(6), 876–889. <https://doi.org/10.1002/joom.1271>
- [19] He, J., Fan, M., & Fan, Y. (2024). Digital transformation and supply chain efficiency improvement: An empirical study from A-share listed companies in China. *PLOS ONE*, 19(4), e0302133. <https://doi.org/10.1371/journal.pone.0302133>
- [20] Nambisan, S. (2017). Digital entrepreneurship: Toward a digital technology perspective of entrepreneurship. *Entrepreneurship Theory and Practice*, 41(6), 1029–1055. <https://doi.org/10.1111/etap.12254>
- [21] Zhao, N., Hong, J., & Lau, K. H. (2023). Impact of supply chain digitalization on supply chain resilience and performance: A multi-mediation model. *International Journal of Production Economics*, 259, 108817. <https://doi.org/10.1016/j.ijpe.2023.108817>
- [22] Birkel, H., Hohenstein, N.-O., & Hähner, S. (2023). How have digital technologies facilitated supply chain resilience in the COVID-19 pandemic? An exploratory case study. *Computers & Industrial Engineering*, 183, 109538. <https://doi.org/10.1016/j.cie.2023.109538>
- [23] Zhang, S., Yu, Q., Wan, S., Cao, H., & Huang, Y. (2024). Digital supply chain: Literature review of seven related technologies. *Manufacturing Review*, 11, 8. <https://doi.org/10.1051/mfreview/2024006>
- [24] Srinivasan, R., & Swink, M. (2018). An investigation of visibility and flexibility as complements to supply chain analytics: An organizational information processing theory perspective. *Production and Operations Management*, 27(10), 1849–1867. <https://doi.org/10.1111/poms.12746>
- [25] Rai, A., Patnayakuni, R., & Seth, N. (2006). Firm performance impacts of digitally enabled supply chain integration capabilities. *MIS Quarterly*, 30(2), 225–246. <https://doi.org/10.2307/25148729>
- [26] Wu, F., Yenyurt, S., Kim, D., & Cavusgil, S. T. (2006). The impact of information technology on supply chain capabilities and firm performance: A resource-based view. *Industrial Marketing Management*, 35(4), 493–504. <https://doi.org/10.1016/j.indmarman.2005.05.003>
- [27] Büyüközkan, G., & Göçer, F. (2018). Digital supply chain: Literature review and a proposed framework for future research. *Computers in Industry*, 97, 157–177. <https://doi.org/10.1016/j.compind.2018.02.010>
- [28] Baryannis, G., Validi, S., Dani, S., & Antoniou, G. (2019). Supply chain risk management and artificial intelligence: State of the art and future research directions. *International Journal of Production Research*, 57(7), 2179–2202. <https://doi.org/10.1080/00207543.2018.1530476>
- [29] Wuest, T., Weimer, D., Irgens, C., & Thoben, K.-D. (2016). Machine learning in manufacturing: Advantages, challenges, and applications. *Production & Manufacturing Research*, 4(1), 23–45. <https://doi.org/10.1080/21693277.2016.1192517>
- [30] Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*, 48, 144–156. <https://doi.org/10.1016/j.jmsy.2018.01.003>
- [31] Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70, 356–365. <https://doi.org/10.1016/j.jbusres.2016.08.009>
- [32] Kache, F., & Seuring, S. (2017). Challenges and opportunities of digital information at the intersection of big data analytics and supply chain management. *International Journal of Operations & Production Management*, 37(1), 10–36. <https://doi.org/10.1108/IJOPM-02-2015-0078>
- [33] Choi, T.-M., Wallace, S. W., & Wang, Y. (2018). Big data analytics in operations management. *Production and Operations Management*, 27(10), 1868–1883. <https://doi.org/10.1111/poms.12838>

- [34] Chen, T., & Zhou, S. (2024). The impact of digital economy on the upgrading of manufacturing structure. *PLOS ONE*, 19(7), e0307184. <https://doi.org/10.1371/journal.pone.0307184>
- [35] Zhang, W., Xu, N., Li, C., Cui, X., Zhang, H., & Chen, W. (2023). Impact of digital input on enterprise green productivity: Micro evidence from the Chinese manufacturing industry. *Journal of Cleaner Production*, 414, 137272. <https://doi.org/10.1016/j.jclepro.2023.137272>
- [36] Geng, Y., Xiang, X., Zhang, G., & Li, X. (2024). Digital transformation along the supply chain: Spillover effects from vertical partnerships. *Journal of Business Research*, 183, 114842. <https://doi.org/10.1016/j.jbusres.2024.114842>
- [37] Li, B., Xu, C., Wang, Y., Zhao, Y., Zhou, Q., & Xing, X. (2024). Digital transformation, supply chain collaboration, and enterprise growth: Theoretical logic and Chinese practice. *European Research on Management and Business Economics*, 30(2), 100249. <https://doi.org/10.1016/j.iedeen.2024.100249>
- [38] Zhou, H., Wang, R., Zhang, X., & Chang, M. (2024). The impact of digital technology adoption on corporate supply chain concentration: Evidence from patent analysis. *Finance Research Letters*, 64, 105413. <https://doi.org/10.1016/j.frl.2024.105413>
- [39] Ittoo, A., Nguyen, L. M., & van den Bosch, A. (2016). Text analytics in industry: Challenges, desiderata and trends. *Computers in Industry*, 78, 96–107. <https://doi.org/10.1016/j.compind.2015.12.001>
- [40] Sun, D., Hu, J., Wu, H., Wu, J., Yang, J., Sheng, Q. Z., & Dustdar, S. (2024). A comprehensive survey on collaborative data-access enablers in the IIoT. *ACM Computing Surveys*, 56(2), Article 50, 1–37. <https://doi.org/10.1145/3612918>
- [41] Shah, P., Patel, H., & Swaminarayan, P. (2024). Multitask sentiment analysis and topic classification using BERT. *EAI Endorsed Transactions on Scalable Information Systems*, 12(1). <https://doi.org/10.4108/eetsis.5287>
- [42] Guo, L., & Sun, C. (2026). An ICT-integrated data-driven framework for optimizing the tobacco supply chain using blockchain and cloud computing. *EAI Endorsed Transactions on Scalable Information Systems*, 12(11). <https://doi.org/10.4108/eetsis.11132>
- [43] Wichmann, P., Brintrup, A., Baker, S., Woodall, P., & McFarlane, D. (2020). Extracting supply chain maps from news articles using deep neural networks. *International Journal of Production Research*, 58(17), 5320–5336. <https://doi.org/10.1080/00207543.2020.1720925>
- [44] Deiva Ganesh, A., & Kalpana, P. (2022). Supply chain risk identification: A real-time data-mining approach. *Industrial Management & Data Systems*, 122(5), 1333–1354. <https://doi.org/10.1108/IMDS-11-2021-0719>
- [45] Zheng, S., Cao, W., Xu, W., & Bian, J. (2019). Doc2EDAG: An end-to-end document-level framework for Chinese financial event extraction. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (pp. 337–346). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1032>
- [46] Mozannar, H., & Sontag, D. (2020). Consistent estimators for learning to defer to an expert. In *Proceedings of the 37th International Conference on Machine Learning* (Vol. 119, pp. 7076–7087). *Proceedings of Machine Learning Research*. <https://proceedings.mlr.press/v119/mozannar20b.html>
- [47] Akidau, T., Bradshaw, R., Chambers, C., Chernyak, S., Fernández-Moctezuma, R. J., Lax, R., McVeety, S., Mills, D., Perry, F., Schmidt, E., & Whittle, S. (2015). The dataflow model: A practical approach to balancing correctness, latency, and cost in massive-scale, unbounded, out-of-order data processing. *Proceedings of the VLDB Endowment*, 8(12), 1792–1803. <https://doi.org/10.14778/2824032.2824076>
- [48] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- [49] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). 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: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1423>
- [50] Goldberg, Y. (2016). A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research*, 57, 345–420. <https://doi.org/10.1613/jair.4992>
- [51] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* 30 (pp. 5998–6008).
- [52] Cui, Y., Che, W., Liu, T., Qin, B., Yang, Z., Wang, S., & Hu, G. (2021). Pre-training with whole word masking for Chinese BERT. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29, 3504–3514. <https://doi.org/10.1109/TASLP.2021.3124365>
- [53] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
- [54] Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87–106. <https://doi.org/10.1257/jep.31.2.87>
- [55] Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11, 685–725. <https://doi.org/10.1146/annurev-economics-080217-053433>
- [56] Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3–28. <https://doi.org/10.1257/jep.28.2.3>
- [57] Agrawal, A., Gans, J., & Goldfarb, A. (2019). Economic policy for artificial intelligence. *Innovation Policy and the Economy*, 19(1), 139–159. <https://doi.org/10.1086/699935>
- [58] Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139–1168. <https://doi.org/10.1111/j.1540-6261.2007.01232.x>
- [59] Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, 64(5), 2023–2052. <https://doi.org/10.1111/j.1540-6261.2009.01493.x>
- [60] Engelberg, J. E., & Parsons, C. A. (2011). The causal impact of media in financial markets. *The Journal of Finance*, 66(1), 67–97. <https://doi.org/10.1111/j.1540-6261.2010.01626.x>
- [61] Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3), 1437–1467. <https://doi.org/10.1111/j.1540-6261.2008.01362.x>
- [62] Ashwin, J. (2024). Financial news media and volatility: Is there more to newspapers than news? *Journal of Financial Markets*, 69, 100896. <https://doi.org/10.1016/j.finmar.2024.100896>
- [63] Xia, J. (2024). Stealing the show: The negative effects of media coverage on peers' stock liquidity. *Finance Research Letters*, 59, 104691. <https://doi.org/10.1016/j.frl.2023.104691>
- [64] Dyck, A., Morse, A., & Zingales, L. (2010). Who blows the whistle on corporate fraud? *The Journal of Finance*, 65(6), 2213–2253. <https://doi.org/10.1111/j.1540-6261.2010.01614.x>
- [65] Syed, R. (2019). Enterprise reputation threats on social media: A case of data breach framing. *The Journal of Strategic*

- Information Systems, 28(3), 257–274. <https://doi.org/10.1016/j.jsis.2018.12.001>
- [66] Meng, T., Yahya, M. H. D. H., Ashhari, Z. M., & Yu, D. (2023). ESG performance, investor attention, and company reputation: Threshold model analysis based on panel data from listed companies in China. *Heliyon*, 9(10), e20974. <https://doi.org/10.1016/j.heliyon.2023.e20974>
- [67] Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35–65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>
- [68] Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3), 535–574. <https://doi.org/10.1257/jel.20181020>
- [69] Hassan, T. A., Hollander, S., van Lent, L., & Tahoun, A. (2019). Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics*, 134(4), 2135–2202. <https://doi.org/10.1093/qje/qjz021>
- [70] Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5), 1423–1465. <https://doi.org/10.1086/688176>
- [71] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536. <https://doi.org/10.1038/323533a0>
- [72] Lee, J., Bagheri, B., & Kao, H.-A. (2015). A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters*, 3, 18–23. <https://doi.org/10.1016/j.mfglet.2014.12.001>
- [73] Kusiak, A. (2018). Smart manufacturing. *International Journal of Production Research*, 56(1–2), 508–517. <https://doi.org/10.1080/00207543.2017.1351644>
- [74] Frank, A. G., Dalenogare, L. S., & Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. *International Journal of Production Economics*, 210, 15–26. <https://doi.org/10.1016/j.ijpe.2019.01.004>
- [75] Ivanov, D., Dolgui, A., & Sokolov, B. (2019). The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics. *International Journal of Production Research*, 57(3), 829–846. <https://doi.org/10.1080/00207543.2018.1488086>
- [76] Tao, F., Qi, Q., Liu, A., & Kusiak, A. (2018). Data-driven smart manufacturing. *Journal of Manufacturing Systems*, 48, 157–169. <https://doi.org/10.1016/j.jmsy.2018.01.006>
- [77] Wu, F., Hu, H., Lin, H., & Ren, X. (2021). Enterprise digital transformation and capital market performance: Empirical evidence from stock liquidity. *Management World*, 37(7), 130–144. <https://doi.org/10.19744/j.cnki.11-1235/f.2021.0097>
- [78] Liu, Z. (2008). Foreign direct investment and technology spillovers: Theory and evidence. *Journal of Development Economics*, 85(1–2), 176–193. <https://doi.org/10.1016/j.jdeveco.2006.07.001>
- [79] Li, G., Li, X., & Huo, L. (2023). Digital economy, spatial spillover and industrial green innovation efficiency: Empirical evidence from China. *Heliyon*, 9(1), e12875. <https://doi.org/10.1016/j.heliyon.2023.e12875>
- [80] Terziovski, M. (2010). Innovation practice and its performance implications in small and medium enterprises (SMEs) in the manufacturing sector: A resource-based view. *Strategic Management Journal*, 31(8), 892–902. <https://doi.org/10.1002/smj.841>
- [81] Bleakley, A., Maloney, E. K., Young, D. G., Hennessy, M., Crowley, J. P., Silk, K. J., & Langbaum, J. B. (2026). Spillover effects of COVID-19 news coverage on willingness to participate in medical research in a diverse sample of US older adults. *Communication Research*, 53(2), 179–203. <https://doi.org/10.1177/00936502231210416>
- [82] Zhang, H., Hong, H., Guo, Y., & Yang, C. (2022). Information spillover effects from media coverage to the crude oil, gold, and Bitcoin markets during the COVID-19 pandemic: Evidence from the time and frequency domains. *International Review of Economics & Finance*, 78, 267–285. <https://doi.org/10.1016/j.iref.2021.12.005>