

Making innovation visible: Innovation narrative focus and configurations of collaborative technological output

Yongchang Jiang

Harbin University of Commerce

Chunjing Bai

bc.j1633@163.com

Harbin University of Commerce

Research Article

Keywords: innovation narrative focus, collaborative technological output, corporate disclosure, configurational analysis, dynamic QCA

Posted Date: September 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11016732/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Making innovation visible: Innovation narrative focus and configurations of collaborative technological output

Abstract

Structured indicators capture aspects of firms' innovation conditions, whereas formal innovation narratives indicate which innovation issues receive relatively greater emphasis in public communication. How this textual information combines with technological foundations and environmental conditions in relation to subsequent collaborative technological output remains underexplored. Using Management Discussion and Analysis (MD&A) texts and firm-year data for Chinese A-share-listed firms from 2011 to 2023, we apply a structural topic model (STM) to construct firm-year measures of innovation narrative focus and then use dynamic qualitative comparative analysis (QCA) to identify configurations associated with high collaborative technological output in the following year. No individual condition meets the threshold for stable necessity; instead, multiple narrative – technology – environment configurations are associated with the outcome. Across the identified configurations, transformation and upgrading orientation recurs, whereas collaboration orientation does not emerge as a core condition. These findings show that relative issue emphasis within formal innovation narratives constitutes an organisational information dimension distinct from structured innovation conditions, and that its association with subsequent collaborative technological output is configurationally dependent.

Keywords

innovation narrative focus; collaborative technological output; corporate disclosure; configurational analysis; dynamic QCA

1. Introduction

Cross-organisational collaboration has become an important means through which firms access complementary knowledge, share R&D risks and improve the effectiveness of innovation (Bogers et al., 2018; Barros et al., 2020; Chemmanur et al., 2023). However, potential resource complementarity does not automatically translate into effective collaboration. Poor partner selection can increase coordination costs and disputes over value appropriation, while exposing firms to knowledge leakage, technological imitation and collaboration failure (Arora et al., 2016; Chemmanur et al., 2023). Before a relationship is formed, firms must therefore identify potential partners and assess their technological foundations, resource complementarity and potential collaborative value (Gulati, 1995; Ahuja, 2000). This process is subject to substantial information constraints. Innovation activities are specialised, long-term,

and uncertain, and firms' technological trajectories, resource-allocation logics, and prospective courses of action are not fully captured by publicly available structured indicators (Connelly et al., 2011; Sarpong et al., 2026). Structured indicators such as R&D expenditure and patent counts capture parts of firms' innovation inputs and accumulated technological outputs, but reveal less about how firms interpret technological opportunities and articulate future development. Formal organisational texts, including Management Discussion and Analysis (MD&A) sections, provide complementary information. As recurring corporate disclosures, MD&A sections contain management commentary on operating conditions, strategic priorities, resource commitments, and future directions (Li, 2008; Loughran and McDonald, 2011), providing a recurring and accessible source for observing how firms organise and express innovation priorities. When partner screening relies on limited observable information, understanding how this textual information is associated with subsequent collaborative technological output is theoretically and practically important.

However, external accessibility does not establish that formal innovation narratives accurately represent firms' innovation capabilities, nor does it show that potential partners have accepted these narratives and acted on them. Firms may selectively present and strategically package innovation activities to build legitimacy, communicate with capital markets or manage impressions (Zott and Huy, 2007). Recent research also indicates that firms adjust annual-report language in response to innovation-protection and communication objectives (Gounopoulos et al., 2025). Moreover, generic references to innovation, digitalisation, transformation and collaboration are relatively easy to imitate, so firms with different technological foundations may use similar narrative content. The issues foregrounded in formal texts therefore do not, by themselves, establish that the underlying claims are supported by substantive capabilities. Signalling theory similarly holds that signal credibility depends on the relationship between a signal and the sender's underlying attributes, as well as on the cost of imitation for senders that lack those attributes (Connelly et al., 2011). Accordingly, we treat formal innovation narratives neither as direct reflections of innovation capability nor as signals that external actors have necessarily recognised as credible. We instead conceptualise them as externally observable organisational information structures.

Research on innovation narratives provides an important basis for understanding formal corporate texts. Existing studies have examined the intensity of innovation or R&D disclosure (Li and Zou, 2024; Liang et al., 2024), textual attributes such as readability, tone, complexity and semantic similarity (Li, 2008; Loughran and McDonald, 2011; Qi and Chen, 2025), and narrative distinctiveness relative to industry conventions or category referents (Wu et al., 2025). This work documents heterogeneity in the amount, linguistic form and relative distinctiveness of innovation information, together with its economic consequences. It pays less attention, however, to which issues firms place in relatively prominent positions within their innovation

narratives. Firms with similar levels of disclosure and similar textual attributes may foreground different innovation issues and therefore present different innovation priorities. At the same time, research on collaborative innovation has mainly explained inter-organisational innovation and its outcomes through structured conditions such as resource endowments, organisational capabilities and relational networks (Cohen and Levinthal, 1990; Gulati, 1995; Ahuja, 2000; Xie et al., 2023). Systematic attention to the issue-emphasis structure within formal innovation narratives remains limited. Existing research therefore does not adequately explain which innovation issues firms emphasise relative to others in these narratives, or under what conditions this issue-emphasis structure is associated with subsequent collaborative technological output.

To capture which innovation issues firms relatively foreground in formal texts, we introduce the construct of innovation narrative focus. We define it as the content structure formed by the relative emphasis that a firm places on different innovation issues within its formal innovation narratives. Its focus on what a firm foregrounds distinguishes innovation narrative focus from disclosure intensity, which concerns information volume; textual attributes, which concern linguistic form; and narrative distinctiveness, which concerns deviation from a referent. The attention-based view holds that organisations selectively direct attention towards particular problems, opportunities and courses of action in specific contexts (Ocasio, 1997, 2011; Ocasio et al., 2018). Issue emphasis in formal texts can therefore be treated as an observable expression of innovation-related attention at the level of formal communication, but not as a direct measure of actual managerial attention or innovation capability. The theoretical core of innovation narrative focus is the relative issue-emphasis structure within formal innovation narratives. The specific narrative orientations identified through text analysis are context-bound empirical manifestations of this construct rather than fixed dimensions that necessarily apply across settings.

Innovation narrative focus alone is insufficient to explain subsequent collaborative technological output. The same narrative focus may occupy different configurational positions under different technological and environmental conditions, while different combinations of conditions may be associated with the same high-output state (Fiss, 2011; Misangyi et al., 2017). We therefore do not estimate a stable net effect of any single condition. Instead, we ask: What configurations of innovation narrative focus, technological foundations and environmental conditions are associated with high collaborative technological output in the following year? To address this question, we combine MD&A texts with annual structured data for Chinese A-share-listed firms from 2011 to 2023. We first use a structural topic model (STM) to identify latent topics in the innovation-related corpus and construct firm-year measures of context-specific narrative orientations. We then use the technology-organisation-environment (TOE) framework to organise the narrative, technological and environmental conditions, and apply

dynamic qualitative comparative analysis (QCA) to identify configurations associated with next-year high collaborative technological output and assess their stability across years and firms.

The results show that no individual condition meets the threshold for stable necessity. Instead, multiple combinations of narrative, technological and environmental conditions are associated with high collaborative technological output. Across the identified configurations, transformation and upgrading orientation recurs, whereas collaboration orientation does not emerge as a core condition. This pattern indicates that the formal innovation narratives associated with high collaborative technological output are not primarily centred on claims about collaboration; they more often foreground future technological directions, implementation pathways and upgrading priorities.

The remainder of the article is organised as follows. Section 2 reviews the relevant literature and develops the research framework. Section 3 describes the research design, sample and measures. Section 4 reports the dynamic QCA results and robustness checks. Section 5 discusses the findings, theoretical contributions, practical implications and limitations.

2. Theoretical background and research framework

2.1. Innovation narrative focus: relative issue emphasis within formal innovation narratives

The attention-based view provides a theoretical basis for explaining why firms develop different issue priorities in their innovation-related communication. Its central premise is that firm behaviour depends on how organisations channel and distribute decision-makers' attention towards particular issues and courses of action. What receives attention is, in turn, shaped by the situation and by organisational rules, resources, relationships and communication procedures (Ocasio, 1997). Subsequent work distinguishes among attentional perspective, attentional engagement and attentional selection, and emphasises that organisational attention is distributed, interactive and temporally dynamic rather than a fixed cognitive stock that can be freely allocated across issues (Ocasio, 2011; Joseph et al., 2024; Ocasio, 2025). Differences in innovation-related attention therefore concern not only whether firms attend to innovation, but also the relative priority assigned to different innovation problems, opportunities and courses of action at a given time.

Formal organisational communication provides observable empirical material for examining such priority structures. Communication does more than transmit pre-existing information: communication practices, organisational vocabularies, and spoken and written texts also participate in selecting, organising and sustaining strategic issues (Ocasio et al., 2018). Prior studies have accordingly used the relative salience of strategic issues in annual reports to

identify firms' strategic knowledge structures and exploration – exploitation orientations (Kabanoff and Brown, 2008; Uotila et al., 2009). Formal texts are nevertheless selected, organised and worded communication products whose content may also reflect disclosure requirements, communication objectives and competitive-protection concerns (Gounopoulos et al., 2025). Issue emphasis in formal texts thus offers an empirical entry point into how firms organise innovation-related attention in formal expression, but it cannot be equated with actual internal attention, resource allocation or innovation capability.

Building on this reasoning, we define innovation narrative focus as the content structure formed by the relative emphasis that a firm places on different innovation issues within its formal innovation narratives. The construct has two related properties. First, it is content-specific: it concerns which problems, opportunities and courses of action a firm foregrounds when communicating about innovation. Second, it is relative: the focality of any issue is assessed against the relative prominence of other innovation issues in the same firm's narrative during the same period. Innovation narrative focus therefore concerns neither the mere mention of an innovation issue nor the absolute frequency of related terms. It captures how a firm distributes relative prominence across multiple innovation issues in its formal communication. This definition is consistent with prior research that uses the relative salience or relative prevalence of textual issues to characterise strategic orientations (Kabanoff and Brown, 2008; Uotila et al., 2009).

Innovation narrative focus is therefore distinct from related text-based constructs. Innovation disclosure intensity captures how much innovation- or R&D-related information a firm discloses (Li and Zou, 2024; Liang et al., 2024). Textual attributes such as readability, tone and complexity capture how that information is expressed (Li, 2008; Loughran and McDonald, 2011; Qi and Chen, 2025). Innovation narrative distinctiveness captures the degree to which a firm's innovation text departs from industry conventions, category prototypes or other referents (Wu et al., 2025). By contrast, innovation narrative focus concerns the internal content configuration of a single firm's formal innovation narratives and asks what the firm relatively foregrounds across innovation issues. Firms with similar disclosure volumes and linguistic attributes may thus display different innovation narrative focus because they assign different relative priorities to those issues.

The theoretical content of innovation narrative focus is relatively stable: it concerns the relative issue-emphasis structure formed within formal innovation narratives. The issues that receive emphasis, and the higher-order narrative orientations that they form, may nevertheless vary across institutional environments, industries, observation periods and corpora. An STM can identify latent content structures in a given corpus and estimate systematic relationships between document characteristics and topic prevalence (Roberts et al., 2014). However, statistical topic patterns do not automatically carry managerial meaning. Their semantic

interpretation, aggregation and theoretical labelling require researcher judgement grounded in the corpus and research question (Hannigan et al., 2019). Recent work similarly argues that topic models have no single validation standard independent of the analytical task and corpus, and that researchers should explain and transparently report model choices and topic judgements (Bernhard-Harrer et al., 2026). We therefore do not prespecify a fixed set of cross-context dimensions for innovation narrative focus. Instead, we use STM to identify latent topics in the innovation-related MD&A texts of Chinese listed firms and aggregate them into higher-order narrative orientations on the basis of their semantic content and inter-topic relationships.

2.2. The conditional association between innovation narrative focus and collaborative technological output

Inter-organisational collaborative innovation is a broad concept encompassing knowledge exchange, joint R&D, technology transfer and joint commercialisation (Bogers et al., 2018; Arya et al., 2018). This study examines a narrower outcome: collaborative technological output, defined as patentable technological results jointly produced by a firm and external actors and formally recorded through joint patent applications. A joint patent application indicates that multiple applicants claim rights to the same invention and thus provides an observable record of cross-organisational technological output and shared ownership (Hagedoorn, 2003). Research also shows systematic relationships between knowledge-creating alliances and the quantity and quality of firms' subsequent patent output (Hottenrott and Lopes-Bento, 2016). Joint patents, however, capture only formal and patentable joint technological results. They do not cover informal knowledge exchange, collaboration negotiations, alliance formation, joint commercialisation or other collaborative innovation processes, nor do they reveal when the relationships underlying those patents were formed. Our theoretical interpretation is therefore confined to subsequent collaborative technological output and does not equate joint patenting with the broader process of collaborative innovation.

Inter-organisational technological collaboration involves bilateral matching and task fit. Its potential value depends not only on the resources held by an individual firm, but also on whether the parties' attributes are complementary and whether that complementarity meets the knowledge and capability requirements of a specific task. Research shows that assessments of partner attractiveness vary with collaboration tasks and project contexts (Shah and Swaminathan, 2008), and that relationship formation is a process of mutual selection around heterogeneous resources and capabilities (Mindruta et al., 2016). Resource complementarity is more likely to support relationship formation when it fits task requirements (Furlotti and Soda, 2018). This literature indicates that collaboration matching involves assessments of resource complementarity, technological attributes and task fit, creating theoretical space for organisational information that supplements conventional resource and capability indicators.

In this context, innovation narrative focus may provide complementary information. By assigning different levels of relative salience to innovation issues, formal innovation narratives present the problems, technological directions and courses of action that firms choose to foreground at a particular time. Research on organisational communication shows that texts and communication practices organise and highlight strategic issues rather than merely transmit pre-existing information (Ocasio et al., 2018). Recent research further finds that narrative R&D disclosures contain information on innovation activity not fully captured by conventional R&D indicators and are associated with subsequent patents and patent citations (Andreou et al., 2026). These arguments motivate the possibility that different issue-emphasis structures in formal texts are systematically associated with subsequent collaborative technological output. This proposition concerns the empirical relevance of formal innovation narratives as observable information; it does not assume that external actors have read, accepted or acted on that information.

The relationship between innovation narrative focus and collaborative technological output should not, however, be understood as a stable direct effect independent of other conditions. The relevance of partner attributes depends on their fit with the collaboration task, the other party's resources and the relational context (Shah and Swaminathan, 2008; Furlotti and Soda, 2018). Recent open-innovation research likewise shows that the innovation returns to external collaboration depend not only on partner selection but also on firms' adaptive capacity, partners' ecosystem membership and broader contextual conditions (Heaton and Min, 2025). The relationship between R&D investment and innovation performance also varies with external partner structures and industry environments (Oruganti et al., 2024). Innovation narrative focus is therefore better treated as an organisational information condition to be examined jointly with technological foundations and environmental conditions, rather than as an independently sufficient cause of collaborative technological output. We consequently use the TOE framework to organise the conditions and a configurational perspective to examine how their combinations are associated with high collaborative technological output.

2.3. Research framework

The TOE framework provides an appropriate analytical structure for selecting and classifying the conditions considered in this study. The framework situates organisational adoption and implementation of technological innovation within three interrelated contexts: technology, organisation and environment (Chau and Tam, 1997; Zhu et al., 2006). We do not use TOE as a causal theory of joint patent formation, nor do we seek to exhaust all possible conditions associated with collaborative technological output. Instead, TOE provides a parsimonious and theoretically interpretable classification structure. The technological dimension captures R&D investment and digital-technology-related expression; the

organisational dimension captures the issue-emphasis structure within formal innovation narratives; and the environmental dimension captures policy resources, regional market institutions and industry competition. TOE thus organises the condition set, while the configurational perspective addresses interdependence among the conditions.

At the technological level, R&D intensity and digitalisation level capture two aspects of firms' technological foundations. R&D intensity reflects the extent to which a firm commits resources to knowledge production and technological development relative to its operating scale, and provides an important basis for identifying, assimilating and exploiting external knowledge (Cohen and Levinthal, 1990). Digitalisation level reflects the extent to which a firm integrates digital technologies into its activities and advances related transformation; such technologies can connect knowledge, processes and organisational boundaries (Nambisan et al., 2019; Vial, 2019). We do not treat either technological condition as independently sufficient for high collaborative technological output. Instead, we examine how they combine with innovation narrative focus and environmental conditions in configurations associated with the outcome.

At the organisational level, innovation narrative focus represents the relative issue-emphasis structure within formal innovation narratives. Unlike R&D intensity and the text-based digitalisation indicator, it captures neither the scale of resource inputs nor the amount of disclosure devoted to a single technological domain. It instead characterises how firms distribute relative prominence across multiple innovation issues. We include innovation narrative focus in the organisational dimension to examine the technological and environmental conditions under which different narrative orientations co-occur with subsequent high collaborative technological output. We do not treat it as a direct proxy for innovation capability, narrative credibility or partner cognition.

At the environmental level, we include government support, marketisation level and industry competitive intensity to capture, respectively, firms' policy-resource context, regional market institutions and product-market competition. Government support can alter the innovation resources available to firms and the conditions under which they bear risk through subsidies and other forms of support, although its relevance may vary with firm characteristics and other contextual conditions (Zhao et al., 2024; Liu and Xu, 2025). Marketisation level captures regional development in market-based resource allocation, property-rights protection and contract enforcement, and thus the transaction-cost and governance environment for inter-organisational collaboration (Zhou et al., 2017). Industry competition may intensify firms' incentives for innovation and technological renewal (Giroud and Mueller, 2011). At the same time, greater reliance on external knowledge may heighten concerns about imitation, knowledge leakage and the appropriation of innovation returns (Arora et al., 2016). We therefore make no assumption that these environmental conditions have uniformly enabling

effects. Instead, we examine how they combine with innovation narrative focus and technological conditions in configurations associated with high collaborative technological output.

From a configurational perspective, the conditions above are not treated as independent sources of net effects. Complex organisational outcomes commonly reflect conjunctions of conditions and exhibit equifinality and causal asymmetry (Fiss, 2011; Misangyi et al., 2017). We therefore examine which combinations of innovation narrative focus, technological foundations and environmental conditions are associated with high collaborative technological output in the following year. Because the sample has a multi-year panel structure, dynamic QCA is also used to assess the stability of these configurational relationships across time and cases (García-Castro and Ariño, 2016). Accordingly, our empirical claims concern set-theoretic configurational associations rather than strict causal effects. Figure 1 presents the analytical framework.

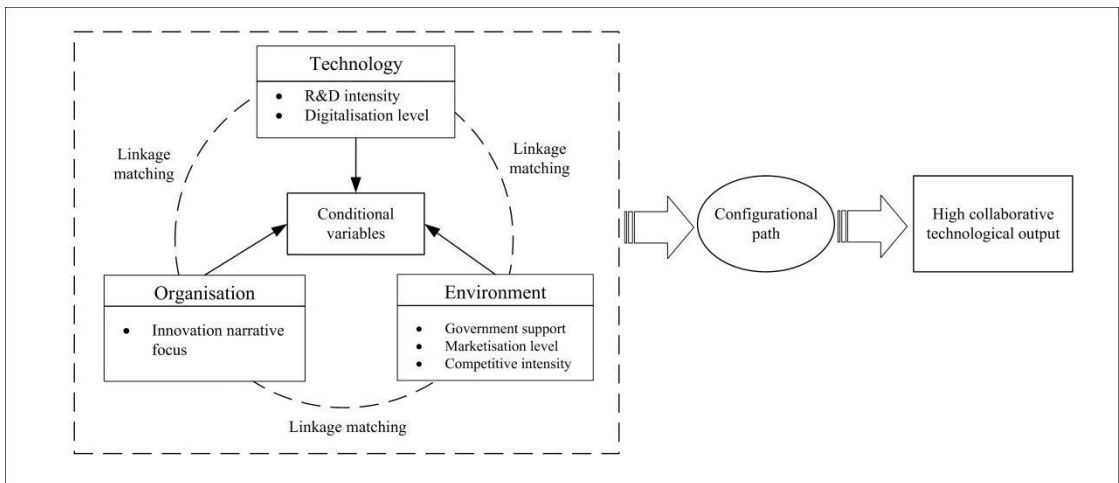


Figure 1. Analytical framework

3. Research design

This study adopts a two-stage research design. In the first stage, innovation-related narrative content is extracted from the MD&A sections of annual reports issued by Chinese A-share-listed firms. An STM is then used to identify the latent topic structure of the innovation-related corpus and construct firm-year measures of innovation narrative focus. In the second stage, these narrative measures are combined with technological and environmental conditions organised by the TOE framework, and dynamic QCA is used to identify configurations associated with high collaborative technological output in the following year. The design converts unstructured text into firm-year organisational information conditions and examines how these narrative measures combine with technological and environmental conditions in relation to the outcome. This section describes the sample and data sources, the construction of the innovation narrative focus measures, and the measurement of the outcome and antecedent

conditions. Section 4 reports calibration, necessity analysis, sufficiency analysis and parameter-based robustness checks.

3.1. Sample and data

This study uses Chinese A-share-listed firms on the Shanghai and Shenzhen stock exchanges from 2011 to 2023 as the research sample. Based on the research purpose and data availability, the initial sample was screened as follows. First, financial firms were excluded because they differ substantially from non-financial firms in asset structure, regulatory constraints, and disclosure rules. Second, firms under special treatment, including ST and *ST firms, were excluded to reduce contamination of the innovation measures by financial distress and going-concern uncertainty. Third, because this study identifies intertemporal configurations at the firm-year level, firms with fewer than five years of valid observations and firms delisted during the sample period were further excluded to reduce the influence of short time series and structural breaks on the analysis. Finally, observations with missing values on key variables were deleted, and all continuous variables were winsorised at the 1st and 99th percentiles to mitigate the influence of extreme values.

To introduce temporal separation between the antecedent conditions and the outcome, the conditions are measured in year t and collaborative technological output in year $t + 1$. Accordingly, the observation period for the conditional variables is 2011 to 2023, and that for the outcome variable is 2012 to 2024. The effective unit of analysis is the configuration of conditions for a given firm-year t and its corresponding collaborative technological output in the following year. After the above procedures, the final sample contains 2,508 firms and 11,170 valid observations, forming an unbalanced panel dataset. The data are mainly obtained from the CSMAR, RESSET, and CNRDS databases, as well as the Cninfo website.

3.2. Construction of innovation narrative focus measures

Building on the definition developed in Section 2, this study operationalises innovation narrative focus as the relative issue-emphasis structure within innovation-related MD&A narratives. Figure 2 summarises the measurement procedure.

3.2.1. Construction of the innovation narrative corpus

This study first used Python to collect annual reports of Chinese A-share-listed firms from 2011 to 2023. This process yielded 31,140 annual reports from 3,431 firms. The MD&A sections were then extracted from these annual reports to form the original text corpus. The MD&A section is a recurring component of the annual report in which management discusses operating conditions, strategic direction, and future development. It is therefore suitable for observing how firms organise and express their innovation priorities.

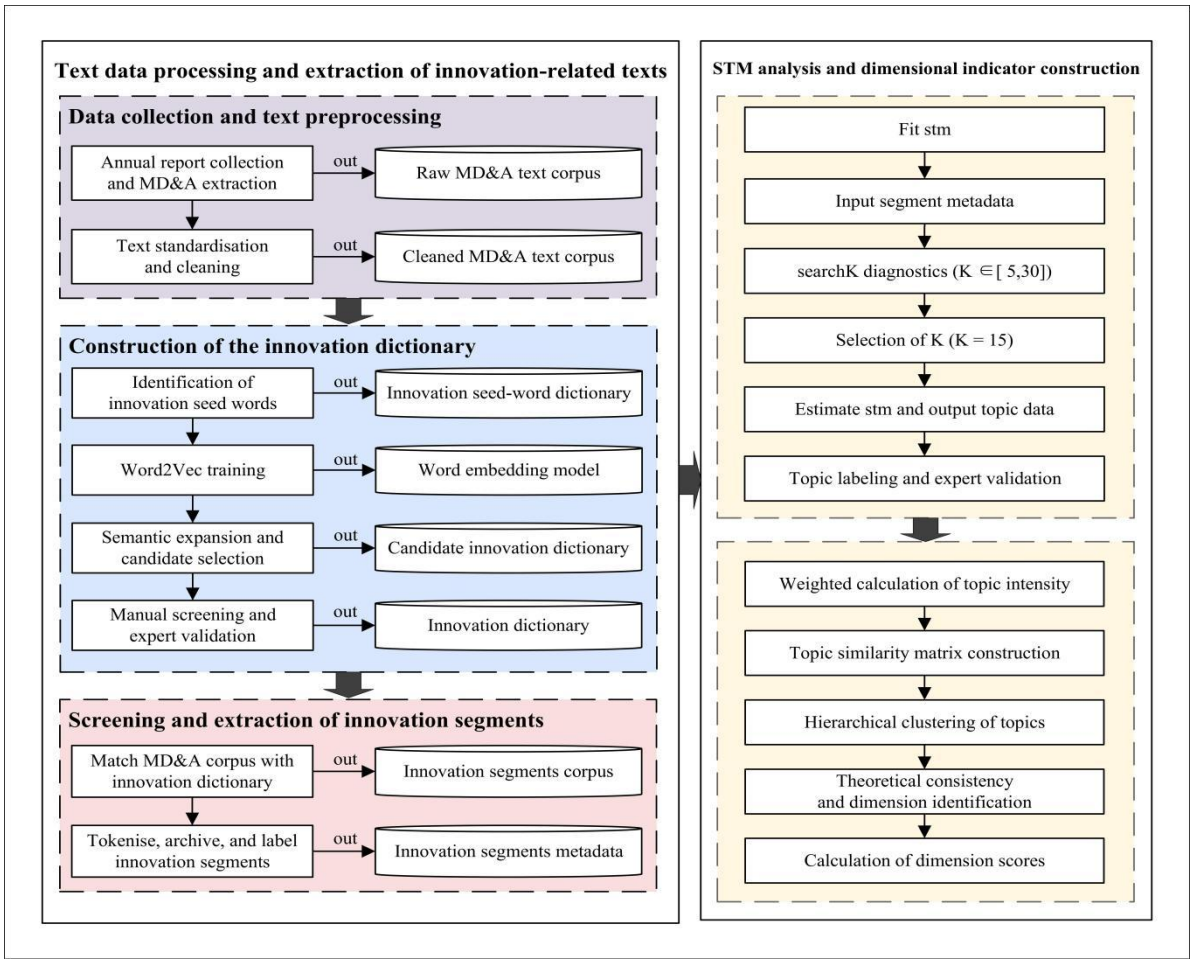


Figure 2. Measurement workflow for innovation narrative focus

In the text preprocessing stage, the extracted MD&A texts were standardised and cleaned by removing non-textual content such as punctuation marks, numbers and formatting characters. Chinese word segmentation was then performed using the jieba package in Python, and stop words were removed. The stop-word list combined the stop-word lists from the Machine Intelligence Laboratory of Sichuan University, Harbin Institute of Technology and Baidu. It was further supplemented with manually screened domain-specific stop words from the financial reporting context, such as ‘annual report’ and ‘during the reporting period’, to reduce the interference of standardised reporting expressions in subsequent topic identification.

To locate innovation-related narratives within the MD&A texts, this study constructed a context-adapted innovation dictionary. The dictionary was used solely to identify and extract innovation-related text; it was not used to calculate the innovation narrative focus measures. Those measures were derived from STM-estimated topic proportions and their subsequent aggregation into higher-order narrative orientations. A static general-purpose dictionary was not adopted because the meaning and usage of innovation-related expressions can vary across industries, policy environments and technological contexts, creating potential semantic mismatch in financial-reporting texts.

Specifically, this study first identified candidate seed words from the large-scale MD&A corpus and relevant innovation policy texts. After expert review, five seed words were retained as semantic anchors. Word2Vec was then used to expand the candidate terms semantically (Rodriguez and Spirling, 2022; Li and Zou, 2024). Candidate terms were ranked by cosine similarity to the seed words, and the 100 highest-ranked terms were included in the initial expanded list. The initial list was then manually screened to improve the validity of the dictionary in the MD&A context. Terms were removed if they were semantically ambiguous, weakly related to innovation activities, overly broad, unsuitable for the financial reporting context, or highly overlapping with other retained terms. This procedure produced a final innovation dictionary containing 69 words and phrases. Appendix Table A1 reports the five seed words and the final retained innovation-dictionary terms.

In the innovation-related narrative extraction stage, this study used text segments rather than full MD&A documents as the basic unit of identification. Full MD&A documents usually contain extensive operating reviews, risk disclosures, market analyses and financial discussions. Modelling the entire MD&A section directly may therefore introduce substantial noise unrelated to innovation. By contrast, text segments can preserve the local semantic context of innovation narratives more precisely. The MD&A texts were segmented using punctuation marks and line breaks as boundaries. Short segments were merged, and long segments were split along semantic boundaries to maintain a segment length of 200 – 800 Chinese characters. Very short segments may contain insufficient semantic information, whereas very long segments may mix multiple issues; this range therefore helps balance semantic completeness and textual comparability.

The innovation dictionary was then used to match text segments. This study retained segments containing at least five dictionary-matched innovation terms. This threshold was used to ensure that the retained segments had sufficient innovation-related semantic density, rather than only scattered mentions of innovation-related words. This procedure yielded 91,671 retained segments, which were indexed by firm and year for topic estimation and aggregation.

3.2.2. Structural topic modelling

After constructing the innovation-related corpus, this study used STM to identify its latent topic structure. Unlike traditional latent Dirichlet allocation, STM allows document-level metadata to be incorporated into the topic-prevalence component of a generative model and is therefore suitable for examining heterogeneous topic distributions across industries and years (Roberts et al., 2014; Roberts et al., 2019). We use STM as a measurement tool rather than as a stand-alone classification exercise: it recovers semantically coherent and interpretable topic patterns that provide the basis for constructing firm-year measures of innovation narrative focus.

We estimated the STM at the segment level, treating each retained innovation-related segment as one document represented by a distribution over latent topics. Because the dynamic

QCA is conducted at the firm-year level, we aggregated segment-topic proportions into firm-year topic shares after model estimation. The STM procedure included tokenisation, stop-word removal, vocabulary restriction and construction of a document-term matrix, and the models were estimated using the stm package in R.

Because innovation narratives may be shaped by annual changes in macro policy, technological cycles and disclosure environments, as well as by industry-specific technological foundations and discourse systems, this study included reporting year and industry information as topic-prevalence covariates in the STM. Firm identifiers were used only to trace and aggregate segment-level topic proportions back to the firm-year level. They were not included as topic-prevalence covariates.

The number of topics, K , was selected by combining statistical diagnostics with substantive interpretability rather than treating any single index as mechanically decisive (Roberts et al., 2019; Tonidandel et al., 2022). Using the searchK procedure in the stm package, this study compared candidate models with K ranging from 5 to 30. The comparison considered held-out likelihood, residuals, semantic coherence and exclusivity, together with the substantive clarity of high-probability and FREX words and representative segments. Figure 3 reports the diagnostic profiles across the candidate values of K .

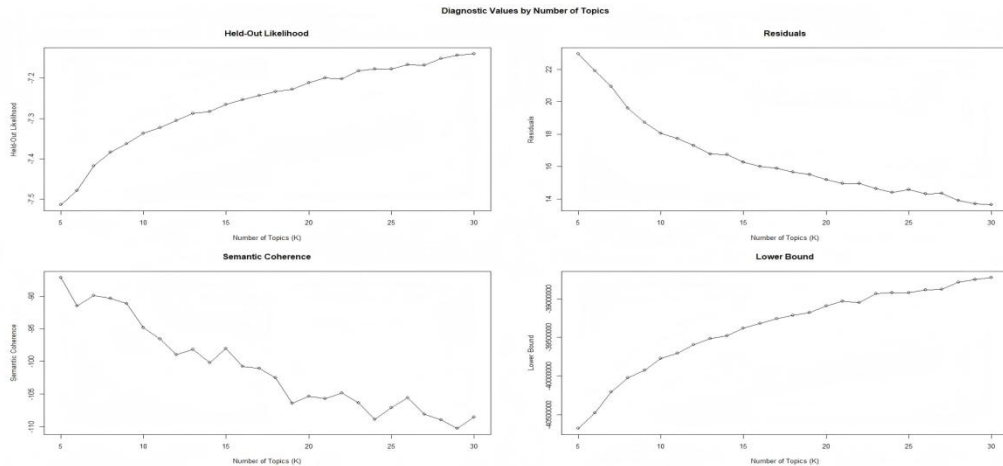


Figure 3. Diagnostic comparison for selecting the number of STM topics

The comparison did not indicate a single topic number that dominated every diagnostic. Models with fewer topics tended to combine substantively distinct innovation issues into broad categories, thereby obscuring differences among resource commitment, technological implementation, strategic upgrading and cross-organisational collaboration. Models with more topics, by contrast, tended to fragment semantically coherent issue domains into narrowly defined or partially overlapping topics, reducing parsimony and making higher-order interpretation less stable. The 15-topic solution retained substantively consequential distinctions without producing excessive fragmentation and offered the most defensible balance among statistical fit, semantic coherence, exclusivity and interpretability. It was therefore

selected as the benchmark representation of the corpus and used in the subsequent measure-construction procedure.

For topic interpretation and labelling, this study followed the principle of combining statistical diagnostics, semantic interpretation and manual validation (Tonidandel et al., 2022; Bernhard-Harrer et al., 2026). Each topic was first interpreted using its high-probability words, FREX words and representative segments, and a preliminary label was assigned. Two experts familiar with firm innovation and text analysis then reviewed the proposed labels independently. Differences in interpretation were resolved through discussion with reference to the representative text, reducing reliance on isolated keywords or a single researcher's judgement. Table 1 reports the labels, corpus proportions and representative high-frequency words for the 15 topics.

Table 1. Topic labels and high-frequency words.

Topic label	Topic proportion	High-frequency words (top 5)
Topic 1: Patent technology	1.56%	utility model, method, invention, patent, device
Topic 2: R&D investment	5.24%	R&D investment, R&D personnel, capitalisation, investment activities, cash inflow
Topic 3: Supply and market risk	9.99%	customers, risk, sales, supplier, new product
Topic 4: Intellectual property	4.37%	patent, intellectual property, utility model, newly added, software
Topic 5: Manufacturing development	6.35%	equipment, electronics, development, design, process
Topic 6: Pharmaceutical R&D	4.71%	pharmaceuticals, biotech, research, clinical, medical
Topic 7: Research platform	6.45%	science and technology, center, engineering, demonstration, laboratory
Topic 8: Industrial upgrading	9.35%	industry, intelligentization, manufacturing, policy, policy planning
Topic 9: Strategic goals	9.93%	management, strategy, innovation, quality, goal
Topic 10: Intelligent services	10.56%	services, platform, intelligent, digitalization, intelligentization
Topic 11: Equipment development	6.23%	system, automobiles, intelligent, development, robots
Topic 12: Talent team	4.81%	talent, team, recruitment, expertise, R&D team
Topic 13: New materials R&D	3.98%	development, new materials, raised funds, performance, high performance
Topic 14: Collaborative foundation	11.20%	innovation, collaboration, technological innovation, brand, foundation
Topic 15: Green process	5.29%	production, process, environmental protection, research, capacity

3.2.3. Construction of firm-year narrative focus measures

STM estimates the proportion of each innovation-related segment assigned to each of the 15 topics. These proportions represent the relative contribution of the topics to the semantic content of a segment. Because the subsequent configurational analysis is conducted at the firm-year level, the segment-level topic proportions were aggregated to obtain firm-year topic shares.

The number and length of innovation-related segments vary across firms and years. An unweighted average would allow numerous short paragraphs to exert disproportionate influence on the annual measure. Segment-level topic proportions were therefore weighted by segment word count. This procedure gives greater weight to segments containing more textual information while preserving the relative topic structure of the firm-year corpus. The measure is defined in Eq. (1):

$$P_{f,t,k} = \frac{\sum_{i=1}^{D_{f,t}} L_i \cdot Q_{i,k}}{\sum_{i=1}^{D_{f,t}} L_i} \quad (1)$$

In Eq. (1), $P_{f,t,k}$ denotes the share of topic k in the innovation-related narratives of firm f in year t ; $D_{f,t}$ denotes the number of innovation-related segment identified for firm f in year t ; L_i

denotes the word count of segment i , and $Q_{i,k}$ denotes the STM-estimated proportion of topic k in segment i . Applying Eq. (1) to all innovation-related segment yields a firm-year topic-share vector for all topics.

The 15 STM topics provide a fine-grained representation of narrative content, but treating all 15 as separate QCA conditions would sharply expand the truth table, aggravate limited diversity and reduce theoretical tractability. The topics were therefore aggregated into a smaller set of higher-order narrative orientations. This step does not replace the topic structure or imply that STM directly estimates the higher-order orientations. It translates the fine-grained topics into theoretically interpretable measures suited to the configurational research question.

To reduce reliance on any single interpretive criterion, topic-to-orientation aggregation drew on three complementary sources of evidence rather than a mechanical classification rule. Three types of evidence were considered. First, STM-based topic-correlation networks and hierarchical clustering were used to identify semantic associations and proximity structures among the 15 topics. Figures A1 and A2 in Appendix A show the STM-based topic-correlation network and hierarchical clustering results, respectively. Second, the high-frequency words, FREX words and representative segments of each topic were examined to assess substantive textual meaning and avoid interpretation based only on isolated keywords. Third, the construct meaning of innovation narrative focus and expert opinions were used to assess the functional role of each topic in firms' innovation narratives and to determine which topics formed similar higher-order narrative orientations.

This procedure yielded four context-specific narrative orientations. Each orientation score is calculated as the sum of the topic shares assigned to that orientation. Innovation resource orientation (I1) comprises Topics 2, 6, 7, 12 and 13 and captures relative emphasis on R&D investment, research platforms, talent and the resource base of innovation. Technology implementation orientation (I2) comprises Topics 1, 4, 5, 11 and 15 and captures relative emphasis on technological development, intellectual property, and the implementation and commercialisation of innovation outcomes. Transformation and upgrading orientation (I3) comprises Topics 8, 9 and 10 and captures relative emphasis on strategic transformation, industrial upgrading and future development direction. Collaboration orientation (I4) comprises Topics 3 and 14 and captures relative emphasis on cross-organisational interaction, external linkages and collaboration-related foundations in innovation narratives.

The four orientations are context-specific empirical manifestations of innovation narrative focus, not fixed cross-context dimensions or independent organisational traits. Nor do they directly measure actual innovation capability, managerial attention or narrative credibility. Because the topic shares form a relative distribution within each firm-year corpus, the orientation scores must be interpreted comparatively: a higher score indicates that an orientation occupies a larger share of the firm's innovation-related narrative, not that the firm discloses a

greater absolute volume of innovation information. Their use in QCA therefore concerns membership in sets of relatively strong narrative emphasis within the analytic sample.

3.3. Variable measurement

Guided by the TOE framework, this study specifies one outcome and nine antecedent conditions. The outcome is collaborative technological output. The technological conditions are R&D intensity and digitalisation level; the organisational conditions are the four context-specific orientations of innovation narrative focus; and the environmental conditions are government support, marketisation level and industry competitive intensity. TOE is used to classify this parsimonious condition set rather than as a causal theory of joint patent formation. Other firm characteristics, including size, age, ownership and financial slack, may also be relevant, but entering a large number of additional set conditions would increase limited diversity and reduce the interpretability of the solution space. The analysis therefore focuses on conditions most directly connected to the research question; omission from the focal condition set should not be interpreted as evidence that other firm attributes are unimportant.

3.3.1. Collaborative technological output

Collaborative technological output is the outcome of the analysis. As discussed in Section 2.2, this term refers to formal and patentable technological outcomes produced across organisational boundaries rather than to the full process of collaborative innovation. Collaborative innovation may also involve informal knowledge exchange, alliance formation, joint R&D projects, co-development, joint commercialisation and ecosystem coordination, none of which is comprehensively observed in the present firm-year data. The outcome measure and all subsequent interpretations are therefore confined to formal collaborative technological output.

Specifically, collaborative technological output is measured by the number of patent applications jointly filed by a firm and external co-applicants. Joint patent applications indicate that multiple actors have participated in formal R&D activities and produced observable technological outcomes. Compared with alliance announcements, cooperation agreements or project-level disclosures, joint patent applications provide a more comparable and verifiable indicator of collaborative technological output at the firm-year level (Liu et al., 2023; Nalmpanti et al., 2024). At the same time, this measure has a clear empirical boundary. It captures formal and patentable technological collaboration, but it does not capture informal collaboration, non-patentable knowledge exchange, alliance negotiation, joint commercialisation, or the quality and duration of collaborative relationships. The empirical interpretation of this study is therefore limited to collaborative technological output.

A patent application is classified as joint when the focal firm and at least one legally distinct co-applicant are listed on the same application. Collaborative technological output is

measured in year $t + 1$ and assigned to years according to patent application dates. Because joint-patent counts are highly right-skewed, the measure is transformed as $\ln(1 + \text{number of joint patent applications})$. The transformation reduces the influence of extreme counts while preserving the ordering of observations. In the subsequent fuzzy-set calibration, high collaborative technological output is defined relative to the distribution of the retained analytic sample rather than by an externally fixed substantive threshold.

3.3.2. Technological conditions

R&D intensity. R&D intensity captures the scale of a firm's resource commitment to knowledge production and technological development relative to its operating activities. It is associated with technological accumulation and the capacity to identify, assimilate and exploit external knowledge (Cohen and Levinthal, 1990; Leung and Sharma, 2021). R&D intensity is measured as R&D expenditure divided by operating revenue, thereby expressing innovation investment relative to the firm's operating scale.

Digitalisation level. Digitalisation level captures the extent to which digital-technology-related activities and initiatives are reflected in the firm's formal reporting. Conceptual research describes digital transformation as organisational change enabled by digital technologies (Vial, 2019; Nambisan et al., 2019), whereas text-based empirical work operationalises firm digitalisation using the prevalence of digital-technology vocabulary in annual reports (Zareie et al., 2024). Following this established text-based approach, this study retains the construct label digitalisation level while interpreting the measure as an indicator constructed from firms' MD&A disclosures, rather than as a comprehensive direct measure of latent digital capability.

Annual reports of Chinese A-share-listed firms from 2011 to 2023 were collected from the Cninfo website. The MD&A texts were extracted and preprocessed by removing non-Chinese characters and stop words and performing Chinese word segmentation. Drawing on digital-economy policy documents and prior research, digital seed words were identified and expanded semantically to construct a context-specific dictionary containing 123 terms. Digitalisation level is calculated as the frequency of digitalisation-related terms divided by the total number of words in the full MD&A text and multiplied by 100. Higher values therefore indicate greater relative prominence of digitalisation-related content in the firm's MD&A.

Although digitalisation level and innovation narrative focus are both derived from annual-report text, they differ in source corpus, measurement procedure and construct meaning. The digitalisation dictionary is developed independently of the innovation dictionary and is applied to the full MD&A to measure the relative prevalence of a predefined technological domain. Innovation narrative focus is constructed only from innovation-related segments and uses STM to estimate the internal distribution of multiple latent innovation topics before their aggregation into higher-order orientations. The former is a domain-specific text-based indicator of digitalisation level; the latter is a relative issue-emphasis structure within formal innovation

narratives. The two variables should therefore not be treated as interchangeable measures of general disclosure intensity.

3.3.3. Organisational condition

The organisational condition is innovation narrative focus. As detailed in Section 3.2, STM first identifies 15 latent topics in the innovation-related MD&A corpus. The topics are then aggregated into four context-specific narrative orientations: innovation resource orientation, technology implementation orientation, transformation and upgrading orientation, and collaboration orientation. For each firm-year, an orientation score equals the sum of the topic shares assigned to that orientation. These measures capture the relative allocation of narrative emphasis across innovation issues and enter the QCA as four organisational information conditions.

3.3.4. Environmental conditions

Government support. Government support captures the volume of policy resources received by a firm through subsidies and grants. Prior studies commonly regard government subsidies or grants received by firms as a direct manifestation of government support. Compared with policy documents or project lists, subsidy amounts provide a more direct observable measure of policy resources received by the firm (Liu and Xu, 2025; Tong et al., 2022). This study therefore measures government support using the amount of government subsidies received by a firm in a given year. Because the amount of government subsidies is skewed, the variable is transformed using the natural logarithm after adding one.

Marketisation level. Marketisation level reflects the degree of market development in the region where the firm is located, including resource allocation, property-rights protection and contract enforcement. In China, provincial marketisation indices are commonly used to capture variation in regional market institutions. We measure marketisation level using the overall marketisation index of the province in which the firm is registered, matched by the firm's registered province and year. A higher value indicates a higher level of marketisation in the firm's region.

Industry competitive intensity. Industry competitive intensity reflects the degree of competitive pressure in the product market in which a firm operates. In empirical research, market concentration is commonly measured by the Herfindahl – Hirschman Index (HHI), with lower concentration indicating stronger competition (Giroud and Mueller, 2011). Accordingly, this study measures industry competitive intensity as $1 - \text{HHI}$, so that a higher value indicates stronger competition. Specifically, the HHI is calculated using firms' operating revenue shares within each industry-year, and industry competitive intensity is then derived as shown in Eq. (2).

$$COMP_{j,t} = 1 - HHI_{j,t} = 1 - \sum_{f=1}^{N_{j,t}} \left(\frac{REV_{f,t}}{\sum_{f=1}^{N_{j,t}} REV_{f,t}} \right)^2 \quad (2)$$

In Eq. (2), $COMP_{j,t}$ denotes the competitive intensity of industry j in year t ; $HHI_{j,t}$ denotes the HHI of industry j in year t ; $N_{j,t}$ denotes the number of firms in industry j in year t , and $REV_{f,t}$ denotes the operating revenue of firm f in year t . A higher value indicates stronger industry competition.

4. Dynamic QCA analysis

4.1. Variable calibration

Dynamic QCA requires raw variables to be calibrated into fuzzy-set membership scores ranging from 0 to 1. Because the focal conditions and outcome do not have generally accepted external thresholds, this study uses direct calibration and combines substantive interpretation with the empirical distribution (Pappas and Woodside, 2021). For each antecedent condition, the 75th, 50th and 25th percentiles define full membership, the crossover point and full non-membership, respectively. These anchors operationalise relative membership in the corresponding sets within the analytic sample; they are not interpreted as universal substantive cut-offs. Collaborative technological output represents a comparatively scarce high-output state and is therefore calibrated more stringently, using the 90th, 50th and 10th percentiles. Scores equal to 0.5 are adjusted to 0.501 to avoid maximum ambiguity in truth-table assignment (Fiss, 2011). Table 2 reports the descriptive statistics and calibration anchors.

Table 2. Calibration and descriptive statistics.

Sets of conditions	Indicator	Descriptive statistics				Fuzzy set calibrations		
		Max.	Min.	Mean	SD	Fully in	Crossover	Fully out
Collaborative technological output	CTO	10.024	0.693	2.522	1.421	4.500	2.303	0.693
R&D intensity	RD	0.283	0.001	0.057	0.047	0.073	0.045	0.028
Digitalisation level	DD	10.749	0.029	1.257	1.335	1.473	0.763	0.446
	I1	0.965	0.001	0.290	0.188	0.403	0.252	0.141
	I2	0.998	0.008	0.245	0.167	0.346	0.217	0.113
Innovation narrative focus	I3	0.949	0.000	0.277	0.193	0.390	0.238	0.124
	I4	0.940	0.000	0.187	0.114	0.244	0.164	0.105
Government support	GV	19.625	0.000	8.078	7.917	15.950	10.463	0.000
Marketisation level	MARKET	13.356	5.216	10.192	1.686	11.379	10.444	9.236
Industry competitive intensity	COMP	0.977	0.362	0.891	0.108	0.950	0.927	0.872

4.2. Necessity analysis

Before conducting the sufficiency analysis, this study first examines whether any single condition constitutes a necessary condition for the outcome. In QCA, necessity is assessed mainly through consistency and coverage. A condition is generally regarded as necessary only when its consistency exceeds 0.90 and its coverage exceeds 0.50. Because this study uses panel data, dynamic QCA further considers inter-group adjustment distance and intra-group consistency distance to evaluate whether the aggregate necessity relationship is stable across years and cases. When the adjustment distance is below 0.20, the aggregate consistency result

can be used directly for necessity assessment. When the adjustment distance exceeds 0.20, the aggregate result may conceal substantial temporal or case-level heterogeneity. In such cases, year-specific results and X – Y scatterplots are examined as supplementary evidence. Table 3 reports the necessity analysis for all conditions.

Table 3. Results of necessary condition analysis.

Condition	High Collaborative Technological Output				Low Collaborative Technological Output			
	Consistency	Coverage	inter-group adjustment distances	intra-group consistency distance	Consistency	Coverage	inter-group adjustment distances	intra-group consistency distance
MARKET	0.595	0.602	0.366	0.587	0.532	0.553	0.450	0.624
~MARKET	0.558	0.537	0.342	0.624	0.617	0.611	0.309	0.551
COMP	0.585	0.561	0.153	0.587	0.601	0.592	0.129	0.624
~COMP	0.575	0.583	0.173	0.624	0.554	0.579	0.125	0.624
RD	0.555	0.561	0.133	0.624	0.572	0.595	0.096	0.624
~RD	0.600	0.576	0.104	0.624	0.579	0.572	0.121	0.624
DD	0.588	0.595	0.201	0.587	0.535	0.558	0.189	0.624
~DD	0.563	0.541	0.181	0.624	0.612	0.604	0.109	0.624
GV	0.535	0.539	0.454	0.587	0.562	0.582	0.482	0.514
~GV	0.585	0.565	0.832	0.440	0.554	0.551	0.904	0.477
I1	0.552	0.548	0.109	0.551	0.586	0.599	0.080	0.551
~I1	0.597	0.584	0.096	0.551	0.558	0.562	0.113	0.551
I2	0.583	0.578	0.133	0.551	0.560	0.572	0.157	0.551
~I2	0.569	0.557	0.141	0.514	0.586	0.591	0.177	0.514
I3	0.600	0.595	0.189	0.477	0.545	0.555	0.177	0.551
~I3	0.551	0.540	0.185	0.551	0.602	0.608	0.109	0.514
I4	0.562	0.557	0.080	0.477	0.581	0.593	0.080	0.477
~I4	0.589	0.577	0.080	0.477	0.566	0.570	0.080	0.514

As shown in Table 3, no condition reaches the conventional threshold for a stable necessary condition for either high or low collaborative technological output. Some conditions, particularly MARKET, ~MARKET, GV and ~GV, display relatively large inter-group adjustment distances, indicating temporal heterogeneity rather than a stable set-theoretic inclusion relationship. Table A2 and Figure A3 in Appendix A report the year-specific necessity results and X–Y scatterplots, respectively. These supplementary analyses show that the set relationships vary substantially across years and do not hold consistently over the full sample period.

Overall, the necessity analysis indicates that high collaborative technological output is not associated with the stable presence or absence of any single technological, organisational or environmental condition. This finding is consistent with the configurational logic of this study: collaborative technological output is unlikely to be characterised by an isolated condition alone. The analysis therefore proceeds to the sufficiency analysis, where the focus shifts from single conditions to combinations of conditions.

4.3. Configurations associated with high collaborative technological output

Dynamic QCA is used to identify combinations of technological, organisational and environmental conditions that are sufficient for membership in the high collaborative technological output set and to assess whether those combinations are stable across time and cases. With nine antecedent conditions, the truth table contains a large potential solution space and cases are unevenly distributed across logically possible rows. The benchmark analysis therefore uses a frequency threshold of 2 to retain empirically recurrent configurations while limiting single-case solutions, a raw consistency threshold of 0.85, and a proportional reduction

in inconsistency (PRI) threshold of 0.70. Because prior research provides no unambiguous directional expectation for all nine conditions and the study seeks to avoid imposing unsupported counterfactual assumptions, no directional assumptions are specified. Table 4 reports the configurations associated with high collaborative technological output.

Table 4. Configurational results for high collaborative technological output.

Configurational conditions	S1	S2	S3	S4	S5	S6	S7	S8
MARKET	●	●	●	●	●			●
COMP	⊗			⊗	●	⊗	●	●
GV	●		●	●	●	●	●	●
RD	⊗	⊗		⊗		⊗	⊗	●
DD	●	●	●	●	●	●	●	⊗
I1		●	●	⊗	⊗	●	⊗	⊗
I2		●	●	●	⊗	●	⊗	●
I3	●	●	●	●	●	●	●	●
I4	⊗	⊗	⊗		⊗	⊗	⊗	⊗
Consistency	0.912	0.940	0.962	0.941	0.956	0.966	0.947	0.971
PRI	0.720	0.769	0.779	0.749	0.809	0.797	0.795	0.778
Raw coverage	0.054	0.055	0.050	0.048	0.038	0.040	0.041	0.028
Unique coverage	0.010	0.016	0.006	0.012	0.003	0.006	0.007	0.006
Inter-group adjustment distances	0.044	0.028	0.036	0.044	0.052	0.044	0.068	0.052
Intra-group consistency distance	0.110	0.073	0.073	0.073	0.073	0.037	0.073	0.037
Total PRI					0.708			
Total consistency					0.902			
Total coverage					0.127			

Note: ● indicates that the core condition is present; ● indicates that the marginal condition is present; ⊗ indicates that the core condition is absent; ⊗ indicates that the marginal condition is absent; blank cells indicate that the condition may be either present or absent.

Table 4 identifies eight configurations associated with high collaborative technological output. Overall solution consistency is 0.902, indicating that the combined solution satisfies the benchmark requirement for set-theoretic sufficiency. Configuration-level consistency ranges from 0.912 to 0.971, and all PRI values exceed 0.70. The reported inter-group adjustment distances and intra-group consistency distances are also within the accepted ranges, providing no indication that the solutions are dominated by pronounced temporal or case-group instability. Overall solution coverage is 0.127. The solution therefore accounts for a limited but empirically identifiable region of the high-output set; it should not be read as an exhaustive explanation of all firm-years with high collaborative technological output. The results support equifinality within that region: more than one conjunction of narrative, technological and environmental conditions is associated with the outcome.

For exposition, the eight configurations are organised into four broader patterns according to the conditions they share and the dominant narrative-technology-environment structure they display. These patterns are interpretive summaries of the solution terms in Table 4, not additional solutions or latent classes. A blank condition in a solution term means that its presence or absence is not specified by that configuration; it must not be interpreted as evidence that the condition is empirically low, absent or unimportant.

(1) Digitalisation and upgrading under market and policy support

This pattern corresponds to Configurations S1 and S4. It is characterised by the joint presence of digitalisation level, transformation and upgrading orientation, marketisation level

and government support. This common structure associates high collaborative technological output with a conjunction in which relatively high text-based digitalisation, strong emphasis on transformation and upgrading, and supportive market and policy conditions co-occur. Substantively, the pattern captures alignment between a firm's reported digitalisation-related activity, a future-oriented upgrading emphasis and an institutional environment characterised by more developed market mechanisms and greater government support. This interpretation concerns the simultaneous presence of calibrated conditions; it does not imply that the external environment causes the narrative emphasis or that narrative content is observed or acted on by potential partners. R&D intensity and collaboration orientation are unspecified in both solution terms, so their presence or absence is not required by this pattern rather than empirically unimportant. Technology implementation orientation is present in S4 but unspecified in S1, indicating within-pattern variation in whether an implementation emphasis accompanies the shared four-condition structure.

(2) Resource, implementation and upgrading alignment

This pattern corresponds to Configurations S2, S3 and S6. It is characterised by the joint presence of innovation resource orientation, technology implementation orientation and transformation and upgrading orientation. Together, these conditions form a comparatively broad narrative-technology profile in which firms place concurrent relative emphasis on innovation resources, technological implementation and future upgrading while also displaying relatively high digitalisation-related textual prevalence. Innovation resource orientation captures relative narrative emphasis on R&D investment, research platforms, talent, and related resource foundations; technology implementation orientation captures relative emphasis on patents, intellectual property, manufacturing development, equipment improvement, and process innovation; and transformation and upgrading orientation captures relative emphasis on strategic renewal and future development direction. Their joint presence indicates alignment among these emphases, not a demonstrated temporal mechanism through which resources are converted into outputs. Environmental conditions differ across S2, S3 and S6, suggesting that this narrative-technology profile can be associated with high collaborative technological output under more than one environmental combination. Collaboration orientation is unspecified in these solution terms and should therefore not be interpreted as absent or unimportant.

(3) Digitalisation and upgrading under competitive conditions

This pattern corresponds to Configurations S5 and S7. The pattern therefore associates high collaborative technological output with a conjunction of relatively high digitalisation-related textual prevalence, strong upgrading emphasis, competition and policy support. Competitive intensity is interpreted here only as part of that conjunction rather than as a standalone determinant. From a theoretical standpoint, competition may create incentives for external knowledge search and technological renewal while also increasing concerns about

imitation and knowledge appropriation; the present QCA, however, does not test which mechanism operates or estimate a uniformly positive net effect of competition. Marketisation level is present in S5 but unspecified in S7, so the comparison does not support characterising the entire pattern as a high-marketisation, low-marketisation or generally constrained context. Instead, the two solutions show that the shared digitalisation-upgrading structure can be associated with high collaborative technological output under more than one environmental specification. Collaboration orientation is also unspecified in both solution terms and therefore does not define this pattern.

(4) R&D-anchored implementation and upgrading

This pattern corresponds to Configuration S8. It differs from the other patterns because digitalisation level is not a prominent condition. Instead, S8 combines R&D intensity, technology implementation orientation, and transformation and upgrading orientation with high membership in the marketisation, industry-competition, and government-support sets. The solution describes a conjunction associated with high collaborative technological output; it does not identify a substitution mechanism or show that external actors used the focal firm's narrative in evaluation.

Across the four patterns, transformation and upgrading orientation is the most stable narrative condition. It appears in all configurations associated with high collaborative technological output. This does not mean that transformation and upgrading orientation is independently necessary or sufficient, since the necessity analysis shows that no single condition reaches the threshold for a stable necessary condition. Rather, its repeated presence suggests that a future-oriented upgrading narrative provides a common interpretive frame across heterogeneous technological and environmental contexts. By emphasising industrial upgrading, strategic renewal, intelligent services and future development direction, firms may make their innovation trajectory more legible to external actors. The recurring role of this condition indicates that high-output configurations are not merely associated with general innovation claims, but with narratives that specify where the firm is heading technologically and strategically.

Collaboration orientation, by contrast, is not specified as either present or absent in any of the eight solution terms. This result does not show that collaboration is unimportant, nor does it demonstrate that collaboration-related communication is ineffective. It indicates only that, under the benchmark topic aggregation and calibration, membership in high or low collaboration-oriented narrative emphasis does not distinguish the identified sufficient configurations. The result concerns the configurational position of a text-based orientation; it provides no direct evidence about collaboration relationships, partner-evaluation criteria or the formation of collaborative ties.

4.4. Robustness checks

To assess the sensitivity of the configurational results to key QCA operational settings, this study re-estimates the sufficiency analysis by increasing the frequency threshold from 2 to 3 and the raw consistency threshold from 0.85 to 0.90. Following QCA robustness guidelines, the stability of the re-estimated solutions is assessed by examining whether they preserve a subset relationship with the main solution and whether the changes in consistency and coverage remain limited (Greckhamer et al., 2018; Oana and Schneider, 2024). The re-estimated solutions preserve this relationship, and the overall distribution of configurations and the combinations of core and absent conditions remain substantively unchanged, with only minor changes in configuration consistency, raw coverage, unique coverage, overall consistency and overall coverage. These results suggest that the main configurational patterns are not driven by the selected frequency or consistency threshold.

5. Conclusions and implications

5.1. Conclusions and discussion

Drawing on MD&A texts and annual data from Chinese A-share-listed firms, this study examines how formal innovation narratives combine with technological foundations and environmental conditions in relation to subsequent collaborative technological output. The findings indicate that no single narrative, technological or environmental condition meets the threshold for stable necessity. Instead, multiple configurations of innovation narrative focus, technological conditions and environmental conditions are associated with high collaborative technological output, and the identified configurations show no marked instability across years or cases.

Transformation and upgrading orientation recurs across the identified high-output configurations, indicating that issues concerning future technological directions, strategic renewal and industrial upgrading repeatedly occupy prominent positions within the relevant narrative configurations. Research on organisational communication emphasises that formal texts participate in the selection and sustained presentation of strategic issues (Ocasio et al., 2018), while research on narrative R&D disclosure shows that textual content can supplement information not fully captured by conventional innovation indicators (Andreou et al., 2026). The present findings further indicate that the empirical relevance of formal innovation narratives lies not only in the amount of information disclosed or the form in which it is expressed, but also in the relative positioning of different innovation issues within the narrative. Transformation and upgrading orientation is not, however, a stable necessary condition in its own right. Its configurational position remains dependent on the technological and environmental conditions with which it is combined.

Collaboration orientation does not emerge as a core condition in the high-output configurations, indicating that the cross-organisational nature of the outcome is not directly mirrored in narrative content. Although high collaborative technological output is inherently cross-organisational, the formal innovation narratives associated with it are not necessarily centred on collaboration, coordination or external linkages. Research on partner selection emphasises that inter-organisational collaboration rests on resource complementarity, technological relatedness and task fit rather than on general expressions of willingness to collaborate (Shah and Swaminathan, 2008; Mindruta et al., 2016; Furlotti and Soda, 2018). The present findings are consistent with this view: the relative emphasis placed on collaboration-related issues does not form a stable core condition, and its empirical position depends on how it is configured with technological and environmental conditions. This finding concerns only the configurational position of collaboration orientation within formal narratives. It does not address the importance of collaborative relationships for joint technological outcomes, nor does it provide direct evidence of partner-evaluation criteria or collaboration-formation processes.

5.2. Theoretical contributions

First, this study shifts the theoretical focus of innovation narrative research from external textual characteristics to the internal structure of issue emphasis. Prior research has mainly examined how much innovation information firms disclose, how such information is expressed linguistically, and whether firms' narratives differ from relevant reference groups. It has paid less attention to why firms with similar disclosure intensity or textual styles may nevertheless present different innovation priorities. This study introduces the construct of innovation narrative focus to capture the content structure formed by the relative emphasis that firms place on different innovation issues within their formal innovation narratives. Formal innovation narratives are therefore not merely vehicles for information disclosure; they also constitute information structures through which firms organise and present innovation priorities. Issue emphasis is thus not simply a statistical output of text analysis. It can also be understood as an observable manifestation of how innovation-related attention is allocated and presented in formal communication.

Second, this study identifies the configurational role of formal innovation narratives as an organisational information condition. By incorporating innovation narrative focus, technological foundations, and environmental conditions into the same analytical framework, it broadens the range of organisational conditions associated with collaborative technological output beyond firms' resources and capabilities to include the relative emphasis they place on innovation-related issues in formal texts. This extension complements rather than substitutes for resource- and capability-based explanations and keeps narrative content conceptually distinct from actual capability, narrative credibility, and partner cognition.

5.3. Practical implications

First, firms should incorporate formal innovation narratives into the governance of innovation information rather than treat them as substitutes for substantive technological foundations. When communicating their innovation activities, firms should present more specific accounts of the technological problems they address, their future development directions, implementation arrangements and upgrading priorities. These accounts should be presented alongside observable information on R&D activities, technology projects, intellectual property and interim technological progress. The purpose of formal innovation communication is not simply to increase the volume of references to innovation or collaboration, but to clarify the relative priority and technological substance of different innovation issues. The recurrence of transformation and upgrading orientation across the identified configurations suggests that firms can articulate their future technological directions and implementation priorities more fully. Because this orientation is not a stable necessary condition in its own right, however, it should not be converted mechanically into a uniform disclosure template for all firms. The non-core position of collaboration orientation likewise indicates that firms should not confine their formal innovation communication to general expressions of willingness to collaborate, but should provide more specific information about technological directions and implementation arrangements.

Second, policy makers and innovation intermediaries should not limit innovation-information support to the compilation of structured indicators such as R&D expenditure, patent counts and project funding. They may also incorporate the technological directions, implementation priorities and upgrading issues foregrounded in firms' formal narratives into the organisation and presentation of innovation information. Disclosure guidance, technology-information platforms and intermediary services can improve the clarity, comparability and verifiability of corporate innovation information by standardising information categories, identifying source materials and presenting narrative content alongside structured indicators. Formal innovation narratives should complement structured indicators rather than be used in isolation to assess firms' technological capabilities, collaborative value or future collaborative technological output. Nor should any single narrative orientation be established as a universal model for corporate innovation communication.

5.4. Limitations and future research directions

This study has several limitations that also suggest directions for future research. First, collaborative technological output is measured using joint patent applications. This measure captures formal and patentable joint technological outcomes but does not encompass broader forms of collaborative innovation, such as informal knowledge exchange, alliance negotiations,

joint commercialisation and ecosystem coordination. Future research could integrate alliance announcements, project-level collaboration records and inter-organisational network data to extend the empirical measurement of collaborative innovation activities and outcomes.

Second, this study relies on MD&A texts to measure innovation narrative focus. Although MD&A texts are standardised, periodic and publicly accessible, their content may still be affected by selective emphasis, strategic packaging and impression management. Future research could compare MD&A texts with patent documents, investor materials, internal innovation documents and collaboration agreements to examine how innovation narrative focus is manifested across different textual settings.

Third, this study does not directly observe how external actors encounter, interpret or use formal innovation narratives. Although MD&A texts are publicly accessible, the present research design cannot establish whether potential partners read the relevant content or whether differences in narrative content affect partner evaluations or collaboration choices. Future research could use interviews, experiments, surveys or process-level data to examine directly how external actors access, interpret and use formal innovation narratives.

Finally, the condition set is intentionally parsimonious, and the empirical sample is restricted to Chinese A-share-listed firms. Future research could incorporate additional firm-level conditions and assess the applicability of the findings across different institutional environments and disclosure regimes.

References

- Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly*, 45(3), 425–455. <https://doi.org/10.2307/2667105>
- Arya, V., Saraf, A., Chichkanov, N., Papa, A., & Romano, M. (2026). AI-enhanced competency transfer hubs: a conceptual framework for university-industry engagement and knowledge sharing. *The Journal of Technology Transfer*, 51(2), 682–712. <https://link.springer.com/article/10.1007/s10961-025-10233-7>
- Andreou, P. C., Lambertides, N., Magidou, M., & Maruska, A. E. (2026). Measuring R&D activity through 10-K narrative disclosures. *Research Policy*, 55(3), 105409. <https://doi.org/10.1016/j.respol.2025.105409>
- Arora, A., Athreye, S., & Huang, C. (2016). The paradox of openness revisited: Collaborative innovation and patenting by UK innovators. *Research Policy*, 45(7), 1352–1361. <https://doi.org/10.1016/j.respol.2016.03.019>
- Barros, M. V., Ferreira, M. B., do Prado, G. F., Piekarski, C. M., & Picinin, C. T. (2020). The interaction between knowledge management and technology transfer: A current literature review between 2013 and 2018. *The Journal of Technology Transfer*, 45(5), 1585–1606. <https://doi.org/10.1007/s10961-020-09782-w>
- Bernhard-Harrer, J., Ashour, R., Eberl, J.-M., Tolochko, P., & Boomgaarden, H. (2026). Beyond standardization: A comprehensive review of topic modeling validation methods for computational social science research. *Political Science Research and Methods*, 14(3), 833–851. <https://doi.org/10.1017/psrm.2025.10008>
- Bogers, M., Chesbrough, H., & Moedas, C. (2018). Open innovation: Research, practices, and policies. *California Management Review*, 60(2), 5–16. <https://doi.org/10.1177/0008125617745086>
- Chau, P. Y. K., & Tam, K. Y. (1997). Factors affecting the adoption of open systems: An exploratory study. *MIS Quarterly*, 21(1), 1–24. <https://doi.org/10.2307/249740>
- Chemmanur, T. J., Shen, Y., & Xie, J. (2023). Innovation beyond firm boundaries: Strategic alliances and corporate innovation. *Journal of Corporate Finance*, 80, 102418. <https://doi.org/10.1016/j.jcorpfin.2023.102418>
- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1), 128–152. <https://doi.org/10.2307/2393553>
- Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, 37(1), 39–67. <https://doi.org/10.1177/0149206310388419>
- Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of Management Journal*, 54(2), 393–420. <https://doi.org/10.5465/AMJ.2011.60263120>
- Furlotti, M., & Soda, G. (2018). Fit for the task: Complementarity, asymmetry, and partner selection in alliances. *Organization Science*, 29(5), 837–854. <https://doi.org/10.1287/orsc.2018.1205>
- García-Castro, R., & Ariño, M. A. (2016). A general approach to panel data set-theoretic research. *Journal of Advanced Management Science and Information Systems*, 2, 63–76. <https://doi.org/10.6000/2371-1647.2016.02.06>
- Giroud, X., & Mueller, H. M. (2011). Corporate governance, product market competition, and equity prices. *The Journal of Finance*, 66(2), 563–600. <https://doi.org/10.1111/j.1540-6261.2010.01642.x>
- Gounopoulos, D., Huang, W., & Yang, M. (2025). Innovation and annual report readability. *Research Policy*, 54(8), 105301. <https://doi.org/10.1016/j.respol.2025.105301>

- Greckhamer, T., Furnari, S., Fiss, P. C., & Aguilera, R. V. (2018). Studying configurations with qualitative comparative analysis: Best practices in strategy and organization research. *Strategic Organization*, 16(4), 482–495. <https://doi.org/10.1177/1476127018786487>
- Gulati, R. (1995). Social structure and alliance formation patterns: A longitudinal analysis. *Administrative Science Quarterly*, 40(4), 619–652. <https://doi.org/10.2307/2393756>
- Hagedoorn, J. (2003). Sharing intellectual property rights—An exploratory study of joint patenting amongst companies. *Industrial and Corporate Change*, 12(5), 1035–1050. <https://doi.org/10.1093/icc/12.5.1035>
- Hannigan, T. R., Haans, R. F. J., Vakili, K., Tchaljian, H., Glaser, V. L., Wang, M. S., Kaplan, S., & Jennings, P. D. (2019). Topic modeling in management research: Rendering new theory from textual data. *Academy of Management Annals*, 13(2), 586–632. <https://doi.org/10.5465/annals.2017.0099>
- Heaton, S., & Min, J. (2025). Open innovation in ecosystems: Exploring how the affiliation of an ecosystem partner impacts the benefits of collaboration in open innovation. *Research Policy*, 54(2), 105168. <https://doi.org/10.1016/j.respol.2024.105168>
- Hottenrott, H., & Lopes-Bento, C. (2016). R&D partnerships and innovation performance: Can there be too much of a good thing? *Journal of Product Innovation Management*, 33(6), 773–794. <https://doi.org/10.1111/jpim.12311>
- Joseph, J., Laureiro-Martinez, D., Nigam, A., Ocasio, W., & Rerup, C. (2024). Research frontiers on the attention-based view of the firm. *Strategic Organization*, 22(1), 6–17. <https://doi.org/10.1177/14761270231223397>
- Kabanoff, B., & Brown, S. (2008). Knowledge structures of prospectors, analyzers, and defenders: Content, structure, stability, and performance. *Strategic Management Journal*, 29(2), 149–171. <https://doi.org/10.1002/smj.644>
- Leung, T. Y., & Sharma, P. (2021). Differences in the impact of R&D intensity and R&D internationalization on firm performance: Mediating role of innovation performance. *Journal of Business Research*, 131, 81–91. <https://doi.org/10.1016/j.jbusres.2021.03.060>
- Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2–3), 221–247. <https://doi.org/10.1016/j.jacceco.2008.02.003>
- Li, X., & Zou, L. (2024). Does mandating narrative disclosure of innovation help unveil the curtain of R&D expenditure? Evidence from regulation change in China. *International Review of Financial Analysis*, 91, 103000. <https://doi.org/10.1016/j.irfa.2023.103000>
- Liang, Q., Yin, Y., Liu, Y., & Zhao, Q. (2024). Narrative R&D disclosure and bank loans: Evidence from China. *Finance Research Letters*, 62, 105259. <https://doi.org/10.1016/j.frl.2024.105259>
- Liu, L., & Xu, X. (2025). The effect of government subsidies on firms' innovation performance: Does subsidy continuity matter? *Applied Economics*, 57(5), 471–487. <https://doi.org/10.1080/00036846.2024.2305160>
- Liu, M., Shan, Y., & Li, Y. (2023). Heterogeneous partners, R&D cooperation and corporate innovation capability: Evidence from Chinese manufacturing firms. *Technology in Society*, 72, 102183. <https://doi.org/10.1016/j.techsoc.2022.102183>
- 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>

- Mindruta, D., Moeen, M., & Agarwal, R. (2016). A two-sided matching approach for partner selection and assessing complementarities in partners' attributes in inter-firm alliances. *Strategic Management Journal*, 37(1), 206–231. <https://doi.org/10.1002/smj.2448>
- Misangyi, V. F., Greekhamer, T., Furnari, S., Fiss, P. C., Crilly, D., & Aguilera, R. V. (2017). Embracing causal complexity: The emergence of a neo-configurational perspective. *Journal of Management*, 43(1), 255–282. <https://doi.org/10.1177/0149206316679252>
- Nalmpanti, A. D., Wong, C. Y., & Oghazi, P. (2024). Collaborating for innovation: The inhibiting role of constraints. *Journal of Innovation & Knowledge*, 9(3), 100504. <https://doi.org/10.1016/j.jik.2024.100504>
- Nambisan, S., Wright, M., & Feldman, M. (2019). The digital transformation of innovation and entrepreneurship: Progress, challenges and key themes. *Research Policy*, 48(8), 103773. <https://doi.org/10.1016/j.respol.2019.03.018>
- Oana, I.-E., & Schneider, C. Q. (2024). A robustness test protocol for applied QCA: Theory and R software application. *Sociological Methods & Research*, 53(1), 57–88. <https://doi.org/10.1177/00491241211036158>
- Ocasio, W. (1997). Towards an attention-based view of the firm. *Strategic Management Journal*, 18(S1), 187–206. [https://doi.org/10.1002/\(SICI\)1097-0266\(199707\)18:1<187::AID-SMJ936>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1097-0266(199707)18:1<187::AID-SMJ936>3.0.CO;2-K)
- Ocasio, W. (2011). Attention to attention. *Organization Science*, 22(5), 1286–1296. <https://doi.org/10.1287/orsc.1100.0602>
- Ocasio, W., Laamanen, T., & Vaara, E. (2018). Communication and attention dynamics: An attention-based view of strategic change. *Strategic Management Journal*, 39(1), 155–167. <https://doi.org/10.1002/smj.2702>
- Ocasio, W. (2025). Attentional control: Institutions, management, organizations, and algorithms. *Journal of Management Inquiry*, 34(1), 3–18. <https://doi.org/10.1177/10564926241278520>
- Oruganti, V., Bodas Freitas, I. M., & Jacob, J. (2024). R&D investment and innovation performance under vertical partner concentration. *Industry and Innovation*, 31(10), 1264–1301. <https://doi.org/10.1080/13662716.2024.2340627>
- Pappas, I. O., & Woodside, A. G. (2021). Fuzzy-set qualitative comparative analysis (fsQCA): Guidelines for research practice in information systems and marketing. *International Journal of Information Management*, 58, 102310. <https://doi.org/10.1016/j.ijinfomgt.2021.102310>
- Qi, D., & Chen, W. (2025). R&D narrative disclosure and semantic similarity: Evidence from China. *Economic Change and Restructuring*, 58, 77. <https://doi.org/10.1007/s10644-025-09915-x>
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4), 1064–1082. <https://doi.org/10.1111/ajps.12103>
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: An R package for structural topic models. *Journal of Statistical Software*, 91(2), 1–40. <https://doi.org/10.18637/jss.v091.i02>
- Rodriguez, P. L., & Spirling, A. (2022). Word embeddings: What works, what doesn't, and how to tell the difference for applied research. *The Journal of Politics*, 84(1), 101–115. <https://doi.org/10.1086/715162>
- Sarpong, D., Meissner, D., Moraes, G. H. S. M. D., & Vismara, S. (2026). The impact of university-industry engagement and the rise of competency transfer partnerships. *The Journal of Technology Transfer*, 51(2), 637–652. <https://doi.org/10.1007/s10961-025-10251-5>

- Shah, R. H., & Swaminathan, V. (2008). Factors influencing partner selection in strategic alliances: The moderating role of alliance context. *Strategic Management Journal*, 29(5), 471–494. <https://doi.org/10.1002/smj.656>
- Tong, X., Wang, Z., & Li, X. (2022). The influence of government subsidy on enterprise innovation: Based on Chinese high-tech enterprises. *Economic Research-Ekonomska Istraživanja*, 35(1), 1481–1499. <https://doi.org/10.1080/1331677X.2021.1972818>
- Tonidandel, S., Summerville, K. M., Gentry, W. A., & Young, S. F. (2022). Using structural topic modeling to gain insight into challenges faced by leaders. *The Leadership Quarterly*, 33(5), 101576. <https://doi.org/10.1016/j.leaqua.2021.101576>
- Uotila, J., Maula, M., Keil, T., & Zahra, S. A. (2009). Exploration, exploitation, and financial performance: Analysis of S&P 500 corporations. *Strategic Management Journal*, 30(2), 221–231. <https://doi.org/10.1002/smj.738>
- Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *Journal of Strategic Information Systems*, 28(2), 118–144. <https://doi.org/10.1016/j.jsis.2019.01.003>
- Wu, X., Chen, X., & Wang, X. (2025). With whom to be distinctive? Examining the impact of innovation narrative distinctiveness from the category prototype vs. category exemplar on firm performance. *Technovation*, 143, 103210. <https://doi.org/10.1016/j.technovation.2025.103210>
- Xie, X., Liu, X., & Chen, J. (2023). A meta-analysis of the relationship between collaborative innovation and innovation performance: The role of formal and informal institutions. *Technovation*, 124, 102740. <https://doi.org/10.1016/j.technovation.2023.102740>
- Zareie, M., Attig, N., El Ghoul, S., & Fooladi, I. (2024). Firm digital transformation and corporate performance: The moderating effect of organizational capital. *Finance Research Letters*, 61, 105032. <https://doi.org/10.1016/j.frl.2024.105032>
- Zhao, Y., Yang, Y., Ma, J., Lu, A., & Xu, X. (2024). Policy-induced cooperative knowledge network, university-industry collaboration and firm innovation: Evidence from the Greater Bay Area. *Technological Forecasting and Social Change*, 200, 123143. <https://doi.org/10.1016/j.techfore.2023.123143>
- Zhou, K. Z., Gao, G. Y., & Zhao, H. (2017). State ownership and firm innovation in China: An integrated view of institutional and efficiency logics. *Administrative Science Quarterly*, 62(2), 375–404. <https://doi.org/10.1177/0001839216674457>
- Zhu, K., Kraemer, K. L., & Xu, S. (2006). The process of innovation assimilation by firms in different countries: A technology diffusion perspective on e-business. *Management Science*, 52(10), 1557–1576. <https://doi.org/10.1287/mnsc.1050.0487>
- Zott, C., & Huy, Q. N. (2007). How entrepreneurs use symbolic management to acquire resources. *Administrative Science Quarterly*, 52(1), 70–105. <https://doi.org/10.2189/asqu.52.1.70>

Appendix TableA1. Innovation lexicon based on Word2Vec.

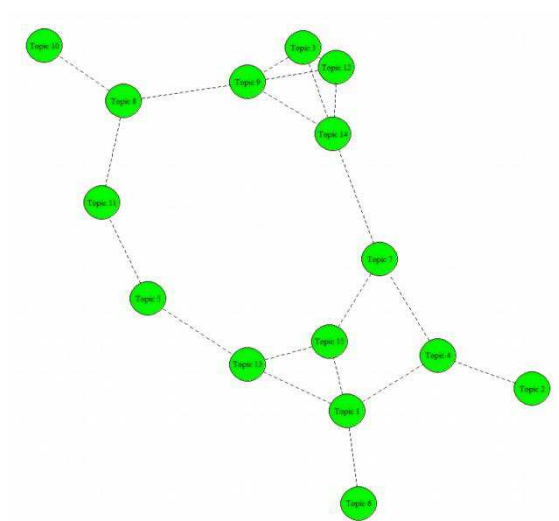
Seed words (5)	Expanded terms (64)
technological innovation; R&D; new product development; technological upgrading; commercialization of research outcomes	R&D investment; process improvement; talent; core technology; application and commercialization; tackling key problems; R&D team; product upgrading; upgrading; technology route; integration; scientific research achievements; joint problem-solving; strategic transformation; R&D personnel; breakthrough; technology development; industrialization; team; key laboratory; key technology; application implementation; innovation-driven; new process; R&D center; technological breakthrough; new product; experimental development; integrated innovation; advanced process; process upgrading; innovation strategy; process optimization; R&D platform; commercialization; intelligent manufacturing; research outcomes; product development; technology reserve; transformation and upgrading; iteration; engineering center; collaborative R&D; major breakthrough; equipment; frontier technology; commercialization; innovation direction; experimental validation; technological problem-solving; bottleneck breakthrough; laboratory; cooperative R&D; product innovation; key research project; green manufacturing; indigenous innovation; intelligent equipment; scientific research platform; technological innovation; green process; joint R&D; technology transfer; R&D system

Source: Authors own work

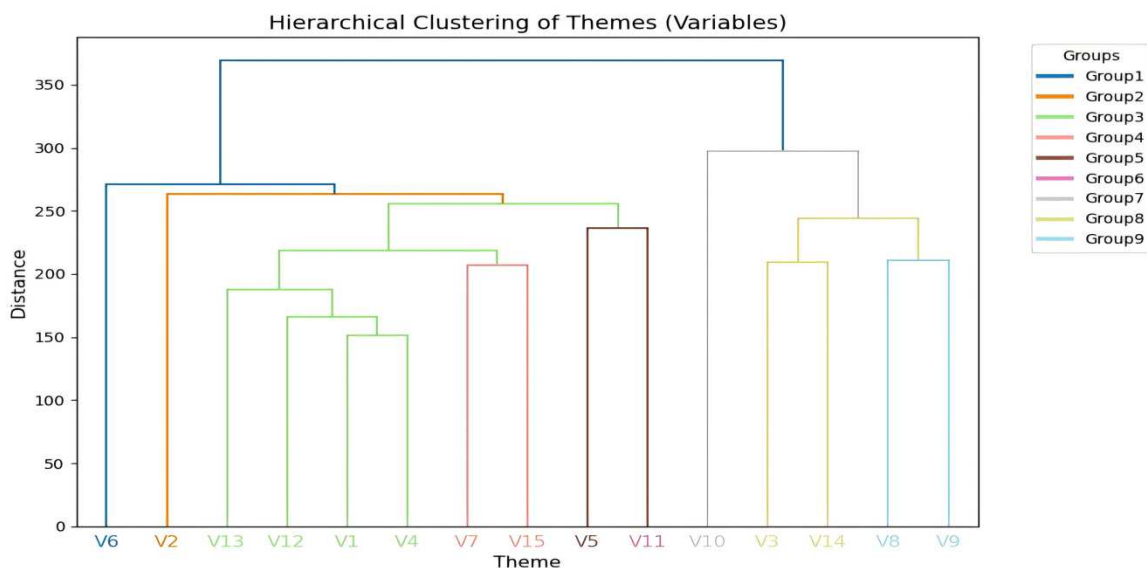
Appendix TableA2. Inter-group results for condition–outcome combinations with adjustment distances greater than 0.20.

Condition-outcome combinations			2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Conditions 1	MARKET/CTO	Consistency	0.274	0.276	0.387	0.447	0.534	0.379	0.418	0.453	0.554	0.613	0.685	0.751	0.827
		Coverage	0.686	0.691	0.646	0.624	0.569	0.677	0.652	0.628	0.614	0.607	0.615	0.633	0.511
Conditions 2	MARKET/~CTO	Consistency	0.208	0.209	0.303	0.375	0.451	0.313	0.353	0.421	0.523	0.593	0.668	0.732	0.794
		Coverage	0.840	0.833	0.782	0.720	0.668	0.694	0.663	0.626	0.534	0.516	0.492	0.464	0.567
Conditions 3	~MARKET/CTO	Consistency	0.936	0.933	0.870	0.799	0.688	0.829	0.784	0.730	0.579	0.511	0.435	0.365	0.300
		Coverage	0.423	0.427	0.447	0.481	0.474	0.493	0.501	0.540	0.569	0.588	0.615	0.644	0.559
Conditions 4	~MARKET/~CTO	Consistency	0.922	0.922	0.863	0.804	0.709	0.854	0.815	0.750	0.622	0.549	0.478	0.421	0.316
		Coverage	0.672	0.669	0.684	0.667	0.679	0.630	0.628	0.596	0.562	0.555	0.554	0.560	0.677
Conditions 5	GV/CTO	Consistency	0.999	0.996	0.999	0.999	0.999	0.999	0.999	0.659	0.601	0.450	0.421	0.390	0.328
		Coverage	0.401	0.404	0.410	0.436	0.433	0.461	0.519	0.572	0.619	0.635	0.695	0.738	0.639
Conditions 6	GV/~CTO	Consistency	0.996	0.992	0.994	0.998	0.998	0.998	0.633	0.558	0.439	0.423	0.356	0.302	0.262
		Coverage	0.645	0.638	0.628	0.600	0.602	0.571	0.601	0.571	0.556	0.560	0.520	0.511	0.635
Conditions 7	~GV/~CTO	Consistency	0.074	0.072	0.066	0.064	0.059	0.058	0.493	0.581	0.699	0.724	0.791	0.845	0.851
		Coverage	0.990	0.970	0.988	0.987	0.985	0.987	0.636	0.610	0.539	0.524	0.515	0.486	0.585
Conditions 8	~GV/CTO	Consistency	0.115	0.105	0.094	0.085	0.082	0.071	0.493	0.549	0.677	0.708	0.730	0.782	0.826
		Coverage	0.948	0.894	0.909	0.964	0.974	0.974	0.527	0.536	0.567	0.583	0.581	0.599	0.492

Source: Authors own work

Appendix FigureA1. STM-based topic-correlation network

Appendix FigureA2. Hierarchical Clustering results



Appendix FigureA3. X-Y scatterplots for the necessity analysis.

