DOI: https://doi.org/10.62985/j.huit_ojs.vol26.no3E.481

DEVELOPING AND VALIDATING A TEXT MINING DICTIONARY FOR MEASURING DYNAMIC CAPABILITIES IN THE INSURANCE INDUSTRY: A STUDY IN THE VIETNAMESE MARKET

Vuong Duc Hoang Quan[1], Nguyen Duc Chung[2,*]

[1]*Ho Chi Minh City Open University*

[2]*Hong Bang International University*

*Email: *ndc0066@yahoo.com*

Received: 30 March 2026; Revised: 20 April 2026; Accepted: 10 May 2026

ABSTRACT

The study presents the process of constructing and validating a specialized Text Mining Dictionary designed to measure Dynamic Capabilities (DC) in Vietnam's insurance industry. Teece's (2007) microfoundations theoretical framework, encompassing three dimensions - Sensing, Seizing, and Transforming - is applied to guide the construction and classification of keywords. The four-step validation process comprises: construction of a preliminary dictionary (seed dictionary) derived from InsurTech terminology; expert validation through the Fleiss' kappa coefficient; statistical validation of convergent validity against hard financial proxy variables; and controlling for the cheap talk phenomenon by cross-referencing text mining scores with Technology CAPEX data. The dictionary is illustrated through the annual reports of publicly listed insurance companies. The results yield a high-reliability measurement instrument that addresses the inherent limitations of the time lag approach in measuring DC. The insurance industry-specific bilingual Vietnamese–English keyword dictionary is designed for release as an open-source resource to facilitate reuse by the academic community.

Keywords: Dynamic Capabilities, text mining, InsurTech dictionary, insurance industry, methodological validation.

1. INTRODUCTION

1.1. Research Context

Vietnam's insurance industry is entering a phase of structural transformation. As of 2023, the insurance penetration rate stood at approximately 2.31% of GDP, nearly 1 percentage point below the emerging market average and substantially below the 10% threshold observed in developed economies. The commercial insurance market is estimated at approximately USD 4.94 billion in 2025, with a projected compound annual growth rate of 6.90% over the 2026–2034 period. Against this backdrop, a wave of technologies - encompassing artificial intelligence (AI), blockchain, the Internet of Things (IoT), and InsurTech platforms - is reshaping the industry's value chain, spanning product design, distribution, risk pricing, and claims management. The accelerating pace of technological change is placing unprecedented adaptive pressure on industry incumbents.

This adaptive capacity is directly linked to the concept of Dynamic Capabilities (DC). Teece [1] established a theoretical framework comprising three dimensions - Sensing (the perception of opportunities and threats), Seizing (the capture of opportunities), and

Transforming (organizational reconfiguration) - providing the foundation upon which firms deploy and protect intangible assets in order to sustain long-term business performance. Nevertheless, measuring DC empirically remains a substantial challenge, owing to its inherently intangible, complex, and organizationally idiosyncratic nature. The lack of reliable measurement instruments constitutes a binding bottleneck that impedes both academic research and corporate strategic planning.

1.2. Research Gap

The majority of existing empirical studies measure DC through the time lag method, in which the temporal gap between input and output variables is used to indirectly infer the presence of DC. Schilke and Helfat [2] argue that empirical designs incorporating time lag represent a necessary but insufficient condition, due to concerns of reverse causality and context-dependence across industry, technology, and research setting. Two core limitations arise from this: measurement is indirect because outcomes depend on causal assumptions, and the approach fails to distinguish between firms' verbal declarations and their actual actions. The second limitation is precisely the cheap talk phenomenon, which Bingler et al. [3] have demonstrated to be prevalent in voluntary disclosures.

Although text mining has been widely applied in the analysis of financial reports, to date no text mining dictionary has been constructed specifically for the insurance industry in Vietnam, particularly in the context of measuring DC in relation to digital transformation. This methodological gap must be addressed before the academic community can conduct reliable causal tests of the impact of DC on firm performance.

1.3. Research Questions

From the identified methodological gap, the study advances three central research questions:

RQ1: How can a bilingual Vietnamese-English Text Mining Dictionary be constructed and validated to measure the three dimensions of DC (Sensing, Seizing, Transforming) in the Vietnamese insurance industry?

RQ2: Do DC scores measured by text mining exhibit a statistically significant correlation with hard financial proxies reflecting enterprises' actual levels of technology investment?

RQ3: Does the cheap talk control mechanism based on Technology CAPEX data effectively distinguish between enterprises that 'walk the talk' and those that 'talk more than they act' in InsurTech disclosures?

These three research questions correspond respectively to three research objectives and three core validation steps of the process presented in the Methodology section.

1.4. Research Objectives

The study sets out three specific objectives:

(1) To construct a bilingual Vietnamese-English text mining dictionary for measuring the three dimensions of DC (Sensing, Seizing, Transforming) in the insurance industry.

(2) To validate the dictionary through a four-step process: preliminary construction, expert validation, convergent validity assessment, and cheap-talk control.

(3) To illustrate the DC scoring procedure based on annual reports of publicly listed insurance companies in Vietnam.

1.5. Research Contributions

The paper makes contributions in both theoretical and practical dimensions. In theoretical terms, the study proposes a methodological solution that addresses the limitation of time lag in measuring DC, while extending the application of the Teece framework [1] to an emerging market context. Practically, the dictionary is intended to be released as an open-source resource for reuse by the academic community, thereby laying the groundwork for causal studies on the effects of DC on the financial performance of insurance firms.

2. THEORETICAL OVERVIEW

2.1. Dynamic Capabilities and Microfoundations

The concept of Dynamic Capabilities originates from the work of Teece, Pisano, and Shuen [4] and was subsequently developed by Teece [1] into a comprehensive analytical framework in the paper "Explicating Dynamic Capabilities". According to Teece [1], DC is decomposed into three microfoundational components:

Sensing is the capability to scan the environment, innovate, learn, and interpret signals pertaining to technology, markets, and regulation. Firms with strong Sensing capabilities typically exhibit decentralized organizational structures, allowing them to identify changes in the environment earlier than their competitors.

Seizing is the capability to mobilize resources and make strategic decisions in order to exploit identified opportunities. This component encompasses business model design, the delineation of organizational boundaries, and investment commitment.

Transforming is the capability to reconfigure the organization, including the redeployment of tangible and intangible assets, so as to sustain competitive advantage in dynamic environments. This is the component that demands the most profound and sustained structural change among the three dimensions.

It is important to emphasize that DC is not synonymous with 'digital capability'. Digital capability reflects the possession and operation of technology. DC is higher-order organizational processes for Sensing, Seizing, and Reconfiguring resources to adapt to a changing environment [1, 2]. This distinction carries significant practical implications: a firm may possess numerous advanced technologies yet still lack DC if it does not have the organizational processes to reconfigure them in a flexible and purposeful manner.

Enterprises possessing strong Dynamic Capabilities not only adapt but are also capable of shaping business ecosystems through innovation and creativity. In the insurance industry, long regarded as slow to innovate, recent research confirms that firms that systematically integrate Dynamic Capabilities tend to adapt more effectively in product development, business model design, and digitalization strategies [5].

2.2. Measurement Challenges of Dynamic Capabilities

Although the DC theoretical framework has achieved broad consensus, the operationalization of Dynamic Capabilities in empirical settings continues to encounter significant barriers. Schilke and Helfat [2] identify three core challenges. The first is construct validity: each specific dimension must be clearly defined and individually focused, rather than assessed through vague aggregate measures. The second is temporal design: the optimal time lag between the measurement of Dynamic Capabilities and subsequent outcomes is contingent upon the industry and technology in question, with no universally applicable benchmark. The third is measurement methodology: a multi-method approach is recommended to enhance the robustness of results, encompassing surveys, archival proxies, experiments, and qualitative research.

The time lag method is systematically susceptible to the risk of reverse causality: firms exhibiting high performance tend to invest more heavily in DC, rather than DC necessarily engendering high performance. This concern underscores the need to develop direct measurement approaches. Text mining has emerged as a promising solution, capable of assessing DC contemporaneously without relying on assumptions regarding temporal lag.

2.3. Text Mining in Financial Research and DC Measurement

Text mining is a branch of natural language processing (NLP) that enables the extraction of meaningful information from unstructured data such as annual reports, legal filings, or meeting minutes. Loughran and McDonald [6] laid a critical foundation by demonstrating that general-purpose psychological dictionaries are inappropriate for financial text analysis: nearly three-quarters of the words classified as 'negative' do not in fact carry negative meaning in a financial context - for instance, “tax”, “liability”, or “foreign”. This finding opens the path toward an ecosystem of reusable domain-specific dictionaries.

Text mining methods have also been applied to detect cheap talk in corporate disclosures. Bingler et al. [3] developed a cheap talk index based on ClimateBERT to assess the quality of climate disclosures made by companies in the MSCI World index. The authors concluded that voluntary commitments are typically accompanied by significantly higher levels of cheap talk. This finding constitutes a direct cautionary signal for DC research: a firm may employ an extensive set of InsurTech keywords in its reports without actually implementing commensurate investment.

In the field of computational management, several recent studies have extended the application of NLP to the measurement of intangible strategic constructs. Goldfarb and Tucker [7] demonstrate that textual data drawn from regulatory filings and annual reports contain strategic signals with predictive value regarding firms' innovation behavior. Marquis and Toffel [8] as well as Lyon and Maxwell [9] have established a robust theoretical foundation for distinguishing substantive commitment from symbolic commitment in corporate texts. These foundations resonate directly with the objective of constructing a measurement tool for DC through disclosure language. Studies on computational management and greenwashing also suggest an important avenue for extension: employing NLP to identify not only the presence but also the quality of strategic language in voluntary disclosure reports [8, 9].

Expanding the application of NLP in measuring the quality of voluntary disclosure. Recent research indicates that textual attributes such as length, degree of boilerplate (cheap-talk language), complexity, and readability may reflect disclosure quality and the strategic intentions of managers. Lang and Stice-Lawrence [10] analyzed more than 15,000 non-U.S. annual reports and demonstrated that textual attributes correlate with regulatory context, market liquidity, institutional ownership, and analyst following. This study employs the mandatory adoption of IFRS as an exogenous shock and finds that the volume of disclosures increases, the degree of boilerplate decreases, and comparability improves; firms with greater textual improvements also exhibit stronger economic benefits [10]. This finding implies that textual measures may serve as proxies for strategic disclosure choices that influence corporate resources and Dynamic Capabilities.

Huang et al. [11] applied computational linguistics to 10,021 sustainability reports and found that report length and completeness significantly affect ESG ratings, while rating agencies respond differently to tone, degree of boilerplate, and redundancy. This result indicates that NLP indicators of boilerplate and completeness can be mapped onto external assessments of disclosure quality [11]. Similarly, Melloni et al. [12] found that firms with poorer social performance produce more obscure and less complete integrated reports, implying that indicators of conciseness and linguistic balance may signal lower disclosure quality and a propensity for impression management [12].

Boilerplate language and risk disclosure. An important line of research focuses on the phenomenon of boilerplate language in risk disclosures and the narrative sections of financial reports. Cazier et al. [13] examined judicial and regulatory assessments of risk factor language and found that lengthier and more boilerplate-laden risk disclosures are less likely to be deemed inadequate in judicial reviews and less likely to attract SEC comment letters. The study further noted that peer firms adopt language previously assessed as adequate, and that judicial reviews prompt peer firms to expand their disclosures irrespective of the adequacy determination reached [13]. This suggests that standardized language may reduce enforcement attention but does not necessarily improve the informativeness of disclosures.

In contrast, Zhu et al. [14] employed text mining on 14,089 Form 10-K filings from financial enterprises surrounding the subprime crisis and documented that firms used more specific and more negative risk language, while significantly reducing stickiness and boilerplate during the crisis period. Failed organizations disclosed a greater number of crisis-related risks in a negative tone compared to surviving organizations, supporting the informational value of risk disclosures during periods of stress [14]. For DC research, this implies that analyzing shifts in disclosure language over time may reveal firms' substantive responses to environmental shocks direct manifestation of the Sensing and Transforming dimensions within the Teece framework [1].

Lin et al. [15] reports a global increase in the length and boilerplate language of environmental and social (E&S) disclosure sections, and employs word embeddings to characterize generic disclosures, noting the potential trade-off between disclosure volume and informativeness in the context of greenwashing detection [15]. This finding complements the cautionary observations of Bingler et al. [3] regarding cheap talk in climate disclosures, while suggesting that embedding techniques are capable of distinguishing substantive commitments from repetitive cheap-talk language.

The Obfuscation Hypothesis and Readability. Another strand of research examines the obfuscation hypothesis, which posits that managers render adverse disclosures more difficult to read and that readability is strategically managed surrounding litigation, governance, or operating performance. Miller [16] demonstrates that more complex filings - those that are longer and less readable - are associated with lower aggregate trading volume, driven by the diminished activity of retail investors. The negative relationship between report complexity and abnormal trading volume reflects both cross-sectional differences and temporal variation in disclosure attributes, highlighting the distributional effects of investors on readability [16].

Laksmana et al. [17] examine the readability of the Compensation Discussion and Analysis (CD&A) section and test whether readability is associated with managerial incentives and obfuscation in compensation disclosures. The authors identify associations consistent with management employing readability as a channel in compensation narratives, linking textual complexity to disclosure practices surrounding performance benchmarks and compensation incentives [17]. Nadeem [18] extends this line of research by documenting that greater gender diversity on corporate boards improves the readability of narratives in 10-K reports, and that improved readability is associated with superior firm performance. This implies that governance may restrict language obfuscation and enhance the quality of voluntary disclosure [18].

Impression management and retrospective sensemaking. Merkl-Davies et al. [19] provide a social-psychological perspective on impression management and retrospective sensemaking in corporate narratives. Using psychological content analysis of chairpersons' statements, the study finds that firms generally do not create impressions inconsistent with a reading of the full annual report; instead, negative organizational outcomes motivate managers to engage in retrospective sensemaking (framing) rather than direct impression management [19]. Wisniewski and Yekini [20] contribute a market-oriented perspective by applying computational linguistics to UK annual report narratives and finding that linguistic indicators predict subsequent stock returns, suggesting that strategic narrative signals captured by text mining are relevant to market valuation [20].

Evidence from an emerging market. Tran et al. [21] provide evidence from Vietnam demonstrating that the tone of annual reports, classified using manual coding and a naïve Bayes classifier, predicts future corporate operating performance and exhibits greater predictive power in contexts characterized by high information asymmetry [21]. The study also supports machine learning text classifiers as cost-effective alternatives for measuring narrative tone in research on disclosure quality. For the Vietnamese context, this finding is particularly significant in that it confirms the feasibility of NLP methods in less mature information environments, where information asymmetry is pronounced, and voluntary disclosure assumes a more critical role.

Synthesis and implications for DC measurement. Overall, the studies reviewed above demonstrate that textual attributes, including length, boilerplate, readability, tone, specificity, and completeness, can be reliably measured using NLP and reflect both disclosure quality and the strategic intent of managers. These methods have been successfully applied to detect cheap talk, greenwashing, obfuscation, and impression management across a range of disclosure contexts, from climate reports to risk disclosures and sustainability reports. For the current study on measuring DC in the insurance industry, these foundations provide a robust methodological basis alongside an important caveat: the construction of a text mining dictionary must be supplemented by rigorous validation steps to distinguish substantive disclosure from cheap-talk language - a requirement that directly gives rise to the cheap talk control design employing Technology CAPEX data, as presented in Section 3.4.

2.4. The Process of Constructing a Domain-Specific Dictionary

Deng et al. [22] systematized the text mining dictionary construction process under the Design Science methodology, comprising the following steps: objective identification, corpus creation, preliminary dictionary construction, validation, and refinement. This process underscores the concept of confidence level - namely, the degree of balance between the validity and generalizability of each dictionary entry. To date, no study has applied a comparable process to construct a dictionary for measuring DC specifically within the insurance industry, particularly in emerging markets such as Vietnam. This methodology gap constitutes the point of departure for the present paper.

3. RESEARCH METHODOLOGY

This study proposes the adoption of a four-step dictionary validation process, building upon the Design Science framework of Deng et al. [22] and the financial linguistics standards of Loughran and McDonald [6]. Each step is designed to address a specific methodological validity threat. Figure 1 illustrates the overall four-step validation process.

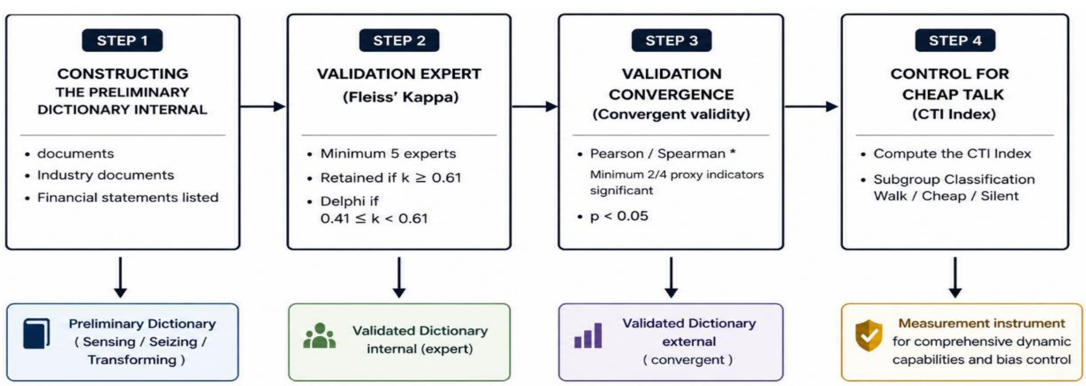

Figure 1. Four-Step InsurTech Dictionary Validation Process
(*Source: Compiled by the authors based on Deng et al. [22] and Loughran & McDonald [6]*)

3.1. Step 1: Constructing the Preliminary Dictionary (Seed Dictionary)

3.1.1. Input Data Sources

The preliminary dictionary was constructed from three primary sources. The first source consists of academic literature: keywords related to InsurTech and DC were collected from Scopus, Web of Science, and Google Scholar using the search string ("InsurTech" OR "insurance technology" OR "digital transformation" AND "insurance"). The second source consists of industry documents, including reports from the Vietnamese Ministry of Finance, the Insurance Association of Vietnam (IAV), and international organizations such as Swiss Re and EIOPA. The third source consists of annual reports of insurance companies listed on HOSE and HNX.

3.1.2. Keyword Screening Process

From the initial keyword pool, which may comprise several hundred entries, the screening process is conducted according to three sequential exclusion criteria. The first criterion is industry specificity: generic technology terms that are not particular to insurance - such as “software” or “information system” in their general sense - are excluded. The second criterion is DC discriminability: keywords that cannot be unambiguously assigned to at least one of the three dimensions - Sensing, Seizing, or Transforming - within the Teece framework [1] are excluded. The third criterion is frequency of occurrence: keywords appearing below the minimum threshold in the pilot annual report corpus are excluded, so as to ensure practical measurability.

After three rounds of screening, the list was reduced to approximately 60–80 candidate keywords and phrases. In this study, the 45 keywords with the highest inter-rater agreement coefficient ($\kappa \geq 0.61$) were retained in the core dictionary published in Appendix. The complete list of 60–80 entries will be made available in the supplementary materials accompanying the official published version, in keeping with the study's commitment to transparency and reproducibility.

3.1.3. Keyword Classification According to the Sensing-Seizing-Transforming Framework

Each keyword in the seed dictionary is classified into one of the three DC dimensions according to Teece [1]. Table 1 summarizes the classification principles and representative keyword examples.

Table 1. Keyword classification according to the three dimensions of dynamic capabilities

DC Dimension	Meaning	Example Vietnamese Keywords	Example English Keywords
Sensing	Environmental scanning, identification of technological opportunities	nghiên cứu thị trường, xu hướng InsurTech, phân tích dữ liệu lớn	market scanning, InsurTech trend, big data analytics
Seizing	Investment, deployment, and opportunity capture	triển khai nền tảng số, đầu tư công nghệ, tự động hóa quy trình	digital platform deployment, technology investment, RPA implementation
Transforming	Organizational restructuring and business model transformation	chuyển đổi số toàn diện, tái cấu trúc nghiệp vụ, hệ sinh thái bảo hiểm số	comprehensive digital transformation, business model innovation, digital insurance ecosystem

Classification notes: Certain keywords — such as “big data analytics” — may reflect both the Sensing dimension (when used to scan and interpret market signals) and the Seizing dimension (when used to deploy analytical applications within operations). In the current bag-of-words approach, each *keyword is assigned to the dimension most predominantly reflecting its function as defined within the Teece framework [1]. The full keyword-to-dimension classification is provided in Appendix (see Appendix for details). This limitation is acknowledged and discussed in detail in Section 5. Advanced NLP techniques such as PhoBERT are proposed as a remedial direction for future research.*

3.2. Step 2: Expert Validation via Fleiss' kappa Coefficient

3.2.1. Validation Procedure

Following the finalization of the seed dictionary, a panel of at least five experts was invited to independently evaluate each keyword. The panel comprises university lecturers specializing in strategic management or finance, senior experts in the insurance industry, and InsurTech specialists. Each expert performs two tasks: classifying keywords into one of the three dimensions - Sensing, Seizing, or Transforming; and rating the degree of relevance on a three-point scale (1 = not relevant; 2 = partially relevant; 3 = fully relevant).

3.2.2. Fleiss' kappa Coefficient

The level of inter-rater agreement among experts is measured using the Fleiss' kappa coefficient [23], a statistical measure of consensus among multiple raters in the categorical classification of items, with correction for chance agreement. This coefficient is calculated using the formula:

$$\kappa = (P^{-} - P^{-}_{e})/(1 - P^{-}_{e})$$

where P^{-} is the observed proportion of agreement and P^{-}_{e} is the expected proportion of agreement due to chance. Kappa values above 0.61 are considered to reflect substantial agreement; values above 0.81 are considered to reflect almost perfect agreement. Only keywords achieving $\kappa \geq 0.61$ are retained in the official dictionary. Contentious terms with $0.41 \leq \kappa < 0.61$ are subjected to an additional round of discussion following the Delphi method.

3.3. Step 3: Statistical Validation of Convergent Validity

3.3.1. Principles

Convergent validity examines whether DC scores measured via text mining exhibit a positive and statistically significant correlation with “hard” financial proxy variables. If the dictionary correctly measures the DC construct, the text mining scores must co-vary with proxy variables that indirectly reflect a firm's digital transformation capabilities.

3.3.2. Financial Proxy Variables

Table 2 lists the four proxy variables employed in this validation step.

Table 2. Financial proxy variables in convergent validity

Proxy variable	Meaning	Data source
IT expenditure / Total revenue	Level of technology investment	Financial statement notes
Technology CAPEX / Total assets	Digital capital investment intensity	Cash flow statement
Number of IT personnel / Total personnel	Technology workforce capability	Annual report
Online contract ratio / Total contracts	Degree of digitalization of distribution channels	Business operations report

3.3.3. Validation Methodology

The correlation between DC text mining scores and proxy variables is examined using Pearson and Spearman correlation coefficients. The dictionary is considered to have achieved convergent validity when the correlation coefficient is positive and statistically significant at $p < 0.05$ with at least two of the four proxy variables.

3.4. Step 4: Controlling for the Cheap Talk Phenomenon

3.4.1. The Cheap Talk Problem in Annual Reports

A significant risk when applying text mining to annual reports is the cheap talk phenomenon: firms employ a high frequency of InsurTech keywords without committing to commensurate investment. Bingler et al. [3] demonstrate that voluntary climate disclosures are frequently accompanied by numerous non-specific commitments, and that cheap talk is positively correlated with actual increases in emissions as well as heightened reputational risk. This phenomenon is entirely likely to occur in InsurTech disclosures and, if left uncontrolled, will significantly distort DC measurement outcomes.

3.4.2. Control Mechanism and Interpretation of the CTI

The study proposes a cheap-talk control mechanism comprising three steps. The first step is to construct the InsurTech Cheap Talk Index (CTI), calculated as the ratio of the DC text mining score to the technology CAPEX-to-total-assets ratio:

$$CTI_i = DC_{TextMining,i} \div (CAPEX_i^T / TotalAssets_i)$$

A high CTI value indicates that the firm engages in extensive discourse while committing minimal investment.

Theoretical and empirical basis of the CTI denominator: Technology CAPEX was selected as the denominator because it constitutes a hard financial indicator reflecting the most substantive investment commitment observable from publicly available financial statements. Brynjolfsson et al. [24] confirm the relationship between capital investment in technology and the J-Curve effect on productivity, implying that technology CAPEX represents a valuable signal of an organization's genuine degree of commitment to digital transformation. Nevertheless, a limitation that must be explicitly acknowledged is that in contemporary digital transformation models predicated on SaaS, cloud computing, or outsourced services, actual technology expenditure may be classified as operating expenses (OPEX) rather than CAPEX. This may result in an elevated CTI that does not reflect cheap talk, but instead signals digital maturity under an asset-light model. To address this limitation, the study proposes two supplementary measures: (a) using total IT expenditure (encompassing both OPEX and CAPEX) as an alternative denominator in the sensitivity analysis; and (b) classifying firms by operating model (asset-heavy vs. asset-light) prior to applying the CTI threshold.

Identifying and Estimating Technology CAPEX: In Vietnam, technology-related capital expenditure items are rarely reported as a discrete line item in financial statements. The study proposes a three-step estimation method: (i) extracting from the notes to the financial statements all investments in “software”, “information systems”, “digital platform”, and equivalent terminology; (ii) supplementing with figures from the “capital expenditure on intangible fixed assets” item in the cash flow statement; (iii) cross-referencing with disclosures in the annual report concerning major digital transformation projects. When data are insufficiently detailed, the alternative proxy variable employed is total IT expenditure / total revenue (from Table 2).

The next step is to categorize firms into three groups: Walk the Talk (high text mining score, high CAPEX), Cheap Talk (high text mining score, low CAPEX), and Silent Doer (low text mining score, high CAPEX). The final step is to test the differences among the three groups in terms of financial performance indicators using ANOVA or Kruskal-Wallis. This mechanism separates genuine signal from noise, substantially enhancing the reliability of the overall measurement instrument.

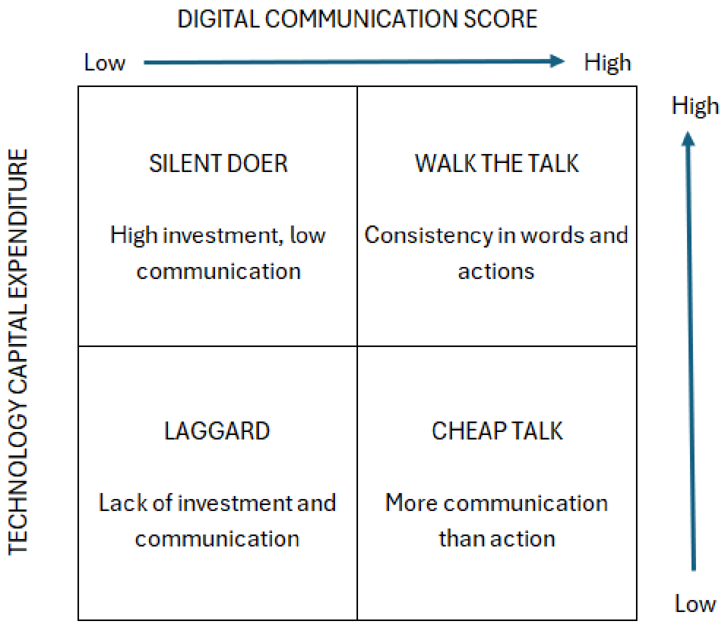

Figure 2. Strategic Classification Matrix: Walk the Talk / Cheap Talk / Silent Doer

Source: Proposed by the authors, developed from Bingler et al. [3]

Matrix reading guide: The vertical axis represents the DC score measured by text mining (low/high); the horizontal axis represents the ratio of technology CAPEX to total assets (low/high). The upper-right quadrant (Walk the Talk) represents the ideal group characterized by high DC disclosure and high investment. The upper-left quadrant (Cheap Talk) represents the group exhibiting high DC disclosure but low investment. The lower-right quadrant (Silent Doer) represents the group that discloses least yet invests substantively.

4. DATA AND ILLUSTRATION CASE

4.1. Scope of Illustrative Data and Sample Representativeness

To illustrate the operationalization of the dictionary, the study employs annual reports covering the 2024–2025 period of listed insurance enterprises in Vietnam, encompassing both life and non-life insurance. The intentional restriction of illustrative data to a short time span reflects a clear and deliberate rationale: the core objective of this paper is to present the methodology, not to report comprehensive empirical findings. The complete panel data for the 2020–2026 period will be the subject of a subsequent study.

An important limitation concerning sample representativeness must be candidly acknowledged. The exclusive focus on insurance companies listed on HOSE and HNX may introduce selection bias in two directions. First, several major life insurers in Vietnam with significant InsurTech investments (such as Prudential, Manulife, and AIA) are not listed on domestic stock exchanges, resulting in the omission of important DC signals. Second, listed companies tend to disclose more strategic information, which may systematically inflate DC text mining scores relative to actual industry conditions. To mitigate this bias, future research is recommended to expand the sample to include unlisted enterprises through the collection of voluntary annual reports and data from the Insurance Supervisory Authority (ISA).

4.2. Scoring Procedure

The DC scoring procedure for each enterprise comprises four sequential steps. Step one is text collection and preprocessing: extracting content from annual reports, removing numerical tables, appendices, and irrelevant material, then standardizing the text (tokenization, stop-word removal, and Vietnamese lemmatization using VnCoreNLP or Underthesea). Step two is keyword Term Frequency counting: for each report, tallying the occurrences of each keyword in the validated dictionary, categorized along the three dimensions of Sensing, Seizing, and Transforming. Step three is computing the normalized score:

$$DC_{dimension,i} = \left[\sum_{k} \quad TF(k,i)\right] / N(i) \times 1000$$

where TF (k, i) denotes the frequency of keyword k in the corporate report of firm i, k belonging to the keyword set of that dimension, and N(i) is the total word count of the report. This normalization controls for variation in report length across firms and across years. Step four involves the aggregation of DC scores: the composite score may be computed as either the arithmetic mean or the weighted average of the three dimensions, contingent on the validation outcomes from Step 3.

4.3. Proposed InsurTech Dictionary

Table 3 presents an excerpt of the proposed dictionary, comprising keywords and phrases distributed with relative balance across the three dimensions. The complete dictionary of 45 core keywords is published in Appendix accompanying this paper.

Table 3. Excerpt of the Proposed InsurTech Dictionary (Sensing-Seizing-Transforming)

No.	Vietnamese	English	DC Dimension
1	trí tuệ nhân tạo	artificial intelligence	Sensing
2	phân tích dữ liệu lớn	big data analytics	Sensing
3	công nghệ bảo hiểm	InsurTech	Sensing
4	chuỗi khối	blockchain	Sensing
5	Internet vạn vật	Internet of Things (IoT)	Sensing
6	điện toán đám mây	cloud computing	Sensing
7	học máy	machine learning	Sensing
8	nền tảng bảo hiểm số	digital insurance platform	Seizing
9	tự động hóa quy trình	robotic process automation	Seizing
10	bancassurance trực tuyến	online bancassurance	Seizing
11	chatbot / trợ lý ảo	chatbot / virtual assistant	Seizing
12	bồi thường tự động	automated claims processing	Seizing
13	đầu tư công nghệ	technology investment	Seizing
14	triển khai nền tảng số	digital platform deployment	Seizing

No.	Vietnamese	English	DC Dimension
15	chuyển đổi số toàn diện	comprehensive digital transformation	Transforming
16	tái cấu trúc mô hình kinh doanh	business model restructuring	Transforming
17	hệ sinh thái bảo hiểm số	digital insurance ecosystem	Transforming
18	đổi mới sáng tạo tổ chức	organizational innovation	Transforming
19	hiện đại hóa hệ thống lõi	core system modernization	Transforming
20	văn hóa số	digital culture	Transforming

Note: Appendix provides a complete list of 45 keywords, including annotations on the dominant dimension, inter-rater agreement coefficient (κ), and contextual usage examples drawn from annual reports.

The keywords were selected on the basis of their specificity to the insurance industry, thereby clearly differentiating the present dictionary from general-purpose text mining dictionaries such as Loughran-McDonald [6]. The dictionary acknowledges the prevalent bilingual phenomenon in Vietnamese disclosure reports, wherein numerous InsurTech terms such as "blockchain", "chatbot", and "IoT" are retained in their original English form within Vietnamese-language texts. Both linguistic variants are incorporated into the dictionary to ensure that no relevant signals are overlooked.

5. RESULTS AND DISCUSSION

5.1. Advantages of Text Mining Relative to Time-Lag Approaches

The text mining method offers three substantive advantages over time-lag approaches in measuring DC. The first is directness: text mining directly measures DC signals through firms' disclosed language, rather than indirectly inferring them from financial outcomes after an indeterminate lag period. The second is granularity: the three dimensions of Sensing, Seizing, and Transforming can be clearly disaggregated, rather than collapsed into a single, ambiguous composite DC index. The third is scalability: once validated, the dictionary can be applied automatically across multiple firms and multiple years, substantially reducing data collection costs relative to traditional survey-based methods.

It must be frankly acknowledged that text mining also carries inherent limitations, the most consequential of which is the risk of cheap talk, as analyzed above. It is precisely for this reason that Step 4 in the validation process was specifically designed to control for this risk, thereby transforming a potential limitation into a valuable additional analytical dimension.

5.2. Limitations of the Bag-of-Words Method and Avenues for NLP Extension

A core limitation of the bag-of-words method is its failure to capture sentence context. This generates two important practical implications. The first is the problem of dual classification: certain keywords such as "big data analytics" may serve both Sensing and Seizing depending on the context of deployment. The second is the problem of negation and conditionality: sentences such as "the company has not yet deployed AI" or "the plan to apply blockchain is under consideration" are still counted as positive DC signals, whereas they in fact reflect the absence or uncertainty of the capability. Both of these limitations must be explicitly acknowledged when interpreting the results.

The most viable remedial direction is the integration of context-sensitive language models. PhoBERT [25], pre-trained on a large-scale Vietnamese corpus, is capable of discriminating word senses based on sentence context. Applying PhoBERT to the keyword classification step will significantly reduce the misclassification rate attributable to dual-context polysemy and will additionally handle negation structures. This represents a high-priority research direction within the dictionary's development roadmap.

5.3. Linguistic Accuracy of Financial Language in the Vietnamese Context

The key lesson from Loughran and McDonald [6] is the necessity of constructing a domain-specific dictionary rather than mechanically applying a general-purpose dictionary. In the context of the Vietnamese insurance industry, this requirement becomes all the more pressing for three compounding reasons. First, the language of Vietnamese annual reports exhibits grammatical and semantic structures that differ fundamentally from English, rendering English-language dictionaries inapplicable without substantial adaptation. Second, many InsurTech terms are used bilingually or retained verbatim in their original English form within Vietnamese texts. Third, Vietnam's distinctive institutional context—characterized by rapidly increasing regulatory compliance pressures and a high pace of policy reform—generates unique linguistic nuances that must be made explicit in the dictionary.

5.4. The J-Curve Effect and Implications for Measurement

In the course of digital transformation, the J-Curve effect represents a widely observed phenomenon: performance undergoes a temporary decline before achieving sustained improvement. Brynjolfsson et al. [24] demonstrate that investment in General Purpose Technologies generates a J-Curve effect on productivity, as complementary intangible investments require time to manifest their full effects. The implication for DC measurement is unambiguous: if only a short time lag is employed, researchers may erroneously conclude that DC exerts no positive impact, when in reality the firm remains in the downward phase of the J-curve. The text mining approach, combined with cheap talk controls, enables DC to be measured at the current point in time without relying on assumptions regarding the optimal lag length, thereby mitigating this source of measurement bias.

5.5. Applications in the Context of Emerging Markets

The Vietnamese insurance market, with a penetration rate of only 2.3–2.8% of GDP - below the ASEAN average (3.35%) and the Asian average (5.37%) exemplifies an institutional void in which substantial growth opportunities coexist with complex regulatory challenges. In this context, InsurTech enterprises such as INSO, OPES, and Papaya are actively leveraging OCR, AI, and mobile platforms to streamline both the insurance purchasing process and the claims settlement process. This transformation was affirmed at the Vietnam Insurance Summit 2025, themed "Strategy for Insurance Industry Development in the Age of the AI Revolution", which convened more than 400 delegates. Musaigwa and Netswera [5] confirm from an international perspective that RPA, AI, blockchain, and online platforms have revolutionized the core processes of the insurance industry. This context underscores the urgency of developing a specialized DC measurement instrument tailored to the insurance industry in Vietnam.

5.6. Scope of Managerial and Policy Implications

As a methodological study, this paper provides a measurement framework for DC rather than causal conclusions regarding the impact of DC on firm performance. The managerial and policy implications drawn at this stage are therefore directional and exploratory in nature, rather than constituting robust empirical evidence. Specifically, the potential of a "Technology Transparency Index" based on CTI requires validation through multi-year panel data and a sufficiently large firm sample before any policy application recommendations can be advanced. Only future research employing comprehensive empirical data will be able to confirm or refute the hypotheses concerning the relationship between DC and the financial performance of insurance firms.

6. CONCLUSION

6.1. Summary

This paper presents the process of constructing and validating a specialized text mining dictionary designed to measure Dynamic Capabilities in Vietnam's insurance industry. The four-step process - encompassing seed dictionary construction, expert validation via Fleiss' kappa, convergent statistical validation against financial proxies, and cheap talk controls using CAPEX data - constitutes a systematic and replicable methodology. A bilingual Vietnamese-English Text Mining Dictionary comprising 45 core keywords, classified across the three dimensions of Sensing, Seizing, and Transforming as defined by Teece [1], is proposed as a novel measurement instrument that overcomes the fundamental limitations of conventional time lag approaches. The complete list drawn from the initial candidate pool (60–80 entries) will be published alongside the official version of the manuscript.

6.2. Limitations and Future Research Directions

The study acknowledges four principal limitations. First, the illustrative data are confined to a 1–2 year window and a limited sample of listed insurance enterprises, thereby excluding large non-listed firms. Moreover, the dictionary requires periodic updates to keep pace with the evolving landscape of technological terminology. Additionally, context effects have not been fully addressed by the current bag-of-words approach: the same keyword may carry different meanings depending on the sentence in which it appears. Finally, the Technology CAPEX denominator in the CTI index may not fully capture the actual technology expenditure of firms operating under an asset-light model.

Four directions for future research are proposed. The first direction is to apply the dictionary to multi-year panel data (2020–2026) to examine the impact of DC on financial performance, including a sample of unlisted firms. The second direction is to integrate PhoBERT [25] to enhance contextual differentiation and improve the handling of negation structures. The third direction is to develop a Technology CAPEX estimation method tailored to the specificities of Vietnamese financial reporting, encompassing OPEX expenditures on technology. A fourth direction is to expand into other financial sectors such as banking and securities in Vietnam, as well as into comparable emerging markets.

REFERENCES

[1] D. J. Teece, "Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance," *Strateg. Manage. J.*, vol. 28, no. 13, pp. 1319–1350, Dec. 2007, doi: https://doi.org/10.1002/smj.640

[2] O. Schilke and C. E. Helfat, "Unlocking dynamic capabilities: Pathways for empirical research," *J. Manage. Sci. Rep.*, vol. 3, no. 2, 2025, doi: https://doi.org/10.1177/27550311251318724

[3] J. A. Bingler, M. Kraus, M. Leippold, and N. Webersinke, "How cheap talk in climate disclosures relates to climate initiatives, corporate emissions, and reputation risk," *J. Banking Finance*, vol. 164, p. 107191, Jul. 2024, doi: https://doi.org/10.1016/j.jbankfin.2024.107191

[4] D. J. Teece, G. Pisano, and A. Shuen, "Dynamic capabilities and strategic management," *Strateg. Manage. J.*, vol. 18, no. 7, pp. 509–533, Aug. 1997, doi: https://doi.org/10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z

[5] M. Musaigwa and F. Netswera, "Transforming insurance: The impact of digital technologies on operations and processes," *Int. J. Appl. Res. Bus. Manage.*, vol. 6, no. 1, pp. 1–18, 2025, doi: https://doi.org/10.51137/wrp.ijarbm.2025.mmti.45818

[6] T. Loughran and B. McDonald, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *J. Finance*, vol. 66, no. 1, pp. 35–65, Feb. 2011, doi: https://doi.org/10.1111/j.1540-6261.2010.01625.x

[7] A. Goldfarb and C. Tucker, "Digital economics," *J. Econ. Lit.*, vol. 57, no. 1, pp. 3–43, Mar. 2019, doi: https://doi.org/10.1257/jel.20171452

[8] C. Marquis, M. W. Toffel, and Y. Zhou, "Scrutiny, norms, and selective disclosure: A global study of greenwashing," *Organ. Sci.*, vol. 27, no. 2, pp. 483–504, Apr. 2016, doi: https://doi.org/10.1287/orsc.2015.1039

[9] T. P. Lyon and J. W. Maxwell, "Greenwash: Corporate environmental disclosure under threat of audit," *J. Econ. Manage. Strategy*, vol. 20, no. 1, pp. 3–41, Mar. 2011, doi: https://doi.org/10.1111/j.1530-9134.2010.00282.x

[10] M. Lang and L. Stice-Lawrence, "Textual analysis and international financial reporting: Large sample evidence," *J. Account. Econ.*, vol. 60, no. 2–3, pp. 110–135, Nov.–Dec. 2015, doi: https://doi.org/10.1016/j.jacceco.2015.09.002

[11] J. Huang, D. D. Wang, and Y. Wang, "Textual attributes of corporate sustainability reports and ESG ratings," *Sustainability*, vol. 16, no. 21, p. 9270, Oct. 2024, doi: https://doi.org/10.3390/su16219270

[12] G. Melloni, A. Caglio, and P. Perego, "Saying more with less? Disclosure conciseness, completeness and balance in Integrated Reports," *J. Account. Public Policy*, vol. 36, no. 3, pp. 220–238, May–Jun. 2017, doi: https://doi.org/10.1016/j.jaccpubpol.2017.03.001

[13] R. A. Cazier, J. L. McMullin, and J. S. Treu, "Are lengthy and boilerplate risk factor disclosures inadequate? An examination of judicial and regulatory assessments of risk factor language," *Account. Rev.*, vol. 96, no. 4, pp. 131–155, Jul. 2021, doi: https://doi.org/10.2308/TAR-2018-0657

[14] X. Zhu, J. Li, and Y. Wang, "Are risk disclosures in financial reports informative? A text mining-based perspective," *Humanit. Soc. Sci. Commun.*, vol. 11, art. no. 1653, Dec. 2024, doi: https://doi.org/10.1057/s41599-024-04169-w

[15] Y. Lin, R. Shen, J. Q. Wang, and Y. Yu, "Global evolution of environmental and social disclosure in annual reports," *J. Account. Res.*, vol. 62, no. 5, pp. 1941–1988, 2024, doi: https://doi.org/10.1111/1475-679X.12575

[16] B. P. Miller, "The effects of reporting complexity on small and large investor trading," *Account. Rev.*, vol. 85, no. 6, pp. 2107–2143, Nov. 2010, doi: https://doi.org/10.2308/accr.00000001

[17] I. Laksmana, W. Tietz, and Y.-W. Yang, "Compensation discussion and analysis (CD&A): Readability and management obfuscation," *J. Account. Public Policy*, vol. 31, no. 2, pp. 185–203, Mar.–Apr. 2012, doi: https://doi.org/10.1016/j.jaccpubpol.2011.08.003

[18] M. Nadeem, "Board gender diversity and managerial obfuscation: Evidence from the readability of narrative disclosure in 10-K reports," *J. Bus. Ethics*, vol. 179, no. 1, pp. 153–177, Jun. 2022, doi: https://doi.org/10.1007/s10551-021-04830-3

[19] D. M. Merkl-Davies, N. M. Brennan, and S. J. McLeay, "Impression management and retrospective sense-making in corporate narratives: A social psychology perspective," *Account. Audit. Account. J.*, vol. 24, no. 3, pp. 315–344, 2011, doi: https://doi.org/10.1108/09513571111124036

[20] T. P. Wisniewski and L. S. Yekini, "Stock market returns and the content of annual report narratives," *Account. Forum*, vol. 39, no. 4, pp. 281–294, Dec. 2015, doi: https://doi.org/10.1016/j.accfor.2015.09.001

[21] L. T. H. Tran *et al.,* "Annual report narrative disclosures, information asymmetry and future firm performance: Evidence from Vietnam," *Int. J. Emerg. Mark.*, vol. 18, no. 2, pp. 351–375, 2023, doi: https://doi.org/10.1108/IJOEM-08-2020-0925

[22] Q. Deng, M. J. Hine, S. Ji, and S. Sur, "Inside the black box of dictionary building for text analytics: A design science approach," *J. Int. Technol. Inf. Manage.*, vol. 27, no. 3, pp. 119–159, 2019, doi: https://doi.org/10.58729/1941-6679.1376

[23] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychol. Bull.*, vol. 76, no. 5, pp. 378–382, 1971, doi: https://doi.org/10.1037/h0031619

[24] E. Brynjolfsson, D. Rock, and C. Syverson, "The productivity J-curve: How intangibles complement general purpose technologies," Nat. Bureau Econ. Res., Cambridge, MA, USA, Working Paper 25148, Oct. 2018, doi: https://doi.org/10.3386/w25148

[25] D. Q. Nguyen and A. T. Nguyen, "PhoBERT: Pre-trained language models for Vietnamese", in *Findings of the Assoc. Comput. Linguist.: EMNLP 2020*, Nov. 2020, pp. 1037–1042, doi: https://doi.org/10.18653/v1/2020.findings-emnlp.92

APPENDIX: SAMPLE INSURTECH DICTIONARY

This appendix releases the proposed sample InsurTech dictionary in accordance with the open-source commitment. Each entry comprises: Vietnamese keyword, equivalent English keyword, primary DC dimension, and notes.

No.	Vietnamese	English	DC Dimension	Notes
1	trí tuệ nhân tạo	artificial intelligence / AI	Sensing	Commonly used in bilingual form
2	phân tích dữ liệu lớn (*)	big data analytics (*)	Sensing	See dual-classification note
3	công nghệ bảo hiểm	InsurTech	Sensing	Retain as English term
4	chuỗi khối	blockchain	Sensing	Retain as English term
5	Internet vạn vật	Internet of Things / IoT	Sensing	Commonly used in bilingual form
6	điện toán đám mây	cloud computing	Sensing	
7	học máy	machine learning / ML	Sensing	Commonly used in bilingual form
8	học sâu	deep learning	Sensing	
9	xử lý ngôn ngữ tự nhiên	natural language processing / NLP	Sensing	
10	phân tích dữ liệu hành vi	behavioral data analytics	Sensing	
11	công nghệ nhận dạng	recognition technology / OCR	Sensing	

No.	Vietnamese	English	DC Dimension	Notes
12	giám sát từ xa	telematics	Sensing	Retain as English term
13	bảo hiểm theo hành vi	usage-based insurance / UBI	Sensing	
14	xu hướng số hóa	digitalization trend	Sensing	
15	nghiên cứu thị trường số	digital market research	Sensing	
16	nền tảng bảo hiểm số	digital insurance platform	Seizing	
17	tự động hóa quy trình	robotic process automation / RPA	Seizing	Commonly used in bilingual form
18	bancassurance trực tuyến	online bancassurance	Seizing	
19	chatbot / trợ lý ảo	chatbot / virtual assistant	Seizing	Retain as English term
20	bồi thường tự động	automated claims processing	Seizing	
21	đầu tư công nghệ	technology investment	Seizing	
22	triển khai nền tảng số	digital platform deployment	Seizing	
23	ứng dụng di động bảo hiểm	insurance mobile application	Seizing	
24	thanh toán điện tử	electronic payment	Seizing	
25	định giá rủi ro số	digital risk pricing	Seizing	
26	phân phối trực tuyến	online distribution	Seizing	
27	hợp đồng thông minh	smart contract	Seizing	
28	API mở	open API	Seizing	Retain as English term
29	hệ thống quản lý bảo hiểm số	digital policy management system	Seizing	
30	xác thực sinh trắc học	biometric authentication	Seizing	
31	chuyển đổi số toàn diện	comprehensive digital transformation	Transforming	
32	tái cấu trúc mô hình kinh doanh	business model restructuring	Transforming	
33	hệ sinh thái bảo hiểm số	digital insurance ecosystem	Transforming	
34	đổi mới sáng tạo tổ chức	organizational innovation	Transforming	

No.	Vietnamese	English	DC Dimension	Notes
35	hiện đại hóa hệ thống lõi	core system modernization	Transforming	
36	văn hóa số	digital culture	Transforming	
37	tái cơ cấu tổ chức	organizational restructuring	Transforming	
38	chuyển đổi mô hình vận hành	operating model transformation	Transforming	
39	đổi mới quy trình nghiệp vụ	business process innovation	Transforming	
40	chiến lược số hóa	digitalization strategy	Transforming	
41	năng lực số	digital capability	Transforming	See note on DC vs. digital capability
42	tái định hình chuỗi giá trị	value chain reshaping	Transforming	
43	quản trị dữ liệu	data governance	Transforming	
44	kiến trúc doanh nghiệp số	digital enterprise architecture	Transforming	
45	hợp tác hệ sinh thái	ecosystem collaboration	Transforming	

(*) "big data analytics" is classified under Sensing based on its primary functional role within the Teece framework [1], yet retains potential dual relevance to Seizing depending on the analytical context. See the discussion presented in Table 1 and Section 3.1.3.