Revista Catarinense da CIÊNCIA CONTÁBIL
ISSN 2237-7662, v. 25, 1-19, e3641, 2026
https://doi.org/10.16930/2237-7662202636412
Available at http://revista.crcsc.org.br

DIGITAL TRANSFORMATION DISCLOSURE IN COMPANIES LISTED ON B3: A TEXT MINING ANALYSIS

MICHELE NOBRE DO NASCIMENTO[1]
Federal Rural University of the Semi-Arid Region, Accounting Sciences Program, Department of Applied Social Sciences, Mossoró, Rio Grande do Norte, Brazil.
https://orcid.org/0009-0002-9496-585X
michele.nascimento@alunos.ufersa.edu.br

GIANINNI MARTINS PEREIRA CIRNE
Federal Rural University of the Semi-Arid Region, Accounting Sciences Program, Department of Applied Social Sciences, Mossoró, Rio Grande do Norte, Brazil.
https://orcid.org/0000-0002-9355-2749
gianinni.martins@gmail.com

YURI GOMES PAIVA AZEVEDO
Federal Rural University of the Semi-Arid Region, Accounting Sciences Program, Department of Applied Social Sciences, Mossoró, Rio Grande do Norte, Brazil.
https://orcid.org/0000-0002-0830-0214
yuri.azevedo@ufersa.edu.br

ALEXSANDRO GONÇALVES DA SILVA PRADO
Federal Rural University of the Semi-Arid Region, Accounting Sciences Program, Department of Applied Social Sciences, Mossoró, Rio Grande do Norte, Brazil.
https://orcid.org/0000-0002-7072-3621
alexsandro.prado@ufersa.edu.br

KALLYSE PRISCILA SOARES DE OLIVEIRA FREIRE
Federal Rural University of the Semi-Arid Region, Accounting Sciences Program, Department of Applied Social Sciences, Mossoró, Rio Grande do Norte, Brazil.
https://orcid.org/0000-0003-1693-0109
kallyse.oliveira@ufersa.edu.br

ABSTRACT

This study aimed to analyze the disclosure of digital transformation terms among publicly traded companies listed on B3 during the period from 2010 to 2023. Grounded in Disclosure Theory, the research examined 3,061 management reports from 281 companies, resulting in a final sample of 914 reports containing the keywords investigated in the study. The data were processed in RStudio using text mining techniques. The results reveal that the disclosure of specific information related to digital transformation remains limited. The most frequently reported terms were broad concepts such as digital, automation, and digitalization. The inverse frequency analysis suggests that blockchain and fintech are indeed representative of technological disruption; however, their low frequency indicates an early stage of disclosure. The co-occurrence analysis demonstrated that the terms digital and digitalization were the most strongly associated, likely because the two concepts are closely interconnected. The terms automation and cloud also showed a strong correlation with digital. Furthermore, the limited presence of more advanced technologies suggests that, although these technologies are considered important, they have not yet been prioritized or widely implemented in publicly disclosed digital transformation strategies. These findings have practical implications for investors by highlighting information gaps that hinder the assessment of companies' technological resilience and long-term strategic viability. This study contributes to the literature by addressing a research gap through the application of natural language processing techniques to accounting narratives in the Brazilian capital market.

Keywords: digital transformation, disclosure, accounting.

Edited in Portuguese and English. Original version in Portuguese.

[1] **Correspondence address:** Avenida Francisco Mota, 572 | Costa e Silva | 59625-900 | Mossoró/RN | Brasil.

Received on September 11, 2025. **Final version received on** April 29, 2026. Accepted on May 19, 2026, by Rogério João Lunkes. **Published on** June 30, 2026.

(cc) BY

1 INTRODUCTION

Digital transformation influences organizational decision-making processes by integrating information flows and strategically restructuring business processes. According to Vial (2019), this evolution extends beyond technology itself and encompasses the creation of new business models that depend on transparent communication. In response to technological advances, organizations seek operational adjustments aimed at increasing productivity, improving service quality, and supporting business decisions. This phenomenon is driven by developments in information technology that enable the integration of data generated throughout the organization, making such information available for management purposes in a timely manner (Ruschel *et al.*, 2011).

Accounting, as an applied social science, is responsible for measuring, interpreting, and disclosing events that affect organizational assets, providing qualitative and quantitative information that offers users a reliable basis for decision-making (Iudícibus, 2021). The relationship between digital transformation and accounting information quality was examined by Chen *et al.* (2024), whose findings indicate that digitalization affects business operations by mitigating agency problems, with stronger effects observed in companies characterized by lower media coverage or reduced market competition. In this regard, Yang *et al.* (2024) support this perspective by demonstrating that digital transformation enhances the comparability of accounting information. This occurs because the use of information technologies standardizes processes and reduces managerial discretion, allowing investors to more accurately identify economic similarities and differences among firms within the same industry, particularly those with weaker governance structures.

Within the Brazilian context, the literature has explored professionals' perceptions regarding the usefulness of digital tools, as evidenced in the studies conducted by Merlugo *et al.* (2021), Fredo (2021), and Franco (2021). However, these investigations primarily focus on operational aspects or individual perceptions, leaving a gap concerning the informational behavior of companies in the capital market. According to Falcão *et al.* (2021), the analysis of corporate narratives through automated techniques is essential for understanding how firms signal their level of technological maturity. It is important to emphasize that digital transformation is not viewed here merely as the adoption of isolated technological tools but rather as a strategic shift in the way organizations create and communicate value. From the perspective of Disclosure Theory, the analysis of term frequency in management reports therefore serves as a proxy for management's commitment to this agenda, revealing the extent to which firms choose to signal their modernization efforts to the market in order to mitigate agency problems and reduce the cost of capital (Wang & He, 2024).

Additionally, the analysis of technological disclosure must consider industry-specific characteristics and the financial condition of organizations. Industries more exposed to digital disruption may exhibit more proactive disclosure practices in response to competitive pressures. Conversely, financial constraints may also influence disclosure strategies, as firms facing greater resource scarcity may use digital transformation disclosure as a signaling mechanism to attract capital.

Against this backdrop, the research question guiding this study is: How are digital transformation-related terms presented in the management reports of publicly traded companies listed on B3? Accordingly, the objective of this study is to analyze the presence and correlation of terms associated with digital transformation in publicly traded companies listed on B3 during the period from 2010 to 2023. The study is justified by the need to investigate how the adoption of new technologies is communicated to external stakeholders, since transparency regarding innovation and competitive advantages assists managers and investors in risk management and business model evaluation (Nemoto *et al.*, 2018; Schappo & Martins, 2022).

By mapping corporate discourse, this study employs Disclosure Theory as a theoretical foundation for understanding how transparency regarding technological innovation affects the relationship between firms and the capital market. According to this theory, organizations engage in voluntary disclosure when the benefits derived from reducing information asymmetry outweigh the preparation and agency costs associated with disclosure. In this context, the presence of terms related to digital transformation in management reports is analyzed as a strategic transparency choice.

However, as noted by Zhang and Yang (2024), previous studies have predominantly focused on binary measures of technological adoption, failing to capture the semantics and depth of corporate communication. Consequently, this study addresses this gap by responding to the recommendations of Wang and He (2024), who emphasize the need to apply natural language processing techniques to reduce subjectivity in report analysis. The taxonomy employed allows for the distinction, through term frequency analysis, between the use of generic terminology and terminology describing disruptive technologies, thereby providing evidence regarding which elements of digital transformation are effectively prioritized in corporate communication. These findings support the discussions presented in this study concerning the level of specificity and usefulness of the information disclosed in the Brazilian market context.

Understanding and disclosing technological innovations within the digital transformation process enables companies to achieve greater operational efficiency, lower costs, and an enhanced ability to meet investors' demands. In this regard, the present study advances the Brazilian literature by providing an empirical analysis based on unstructured data, thereby overcoming the limitations associated with purely qualitative approaches.

2 THEORETICAL BACKGROUND

Accounting, as an applied social science, is founded on the need to reduce information asymmetry among market participants (Verrecchia, 2001). Within this context, Disclosure Theory posits that the voluntary communication of strategic information, such as digital transformation initiatives, functions as a signaling mechanism of an organization's adaptability and long-term sustainability (Spence, 1973). From this perspective, digital transformation is no longer viewed merely as a technological support tool but rather as a structural process that reshapes value creation dynamics. Consequently, the manner in which this phenomenon is reported to external stakeholders becomes an indicator of strategic resilience, enabling the market to assess how management responds to technological obsolescence and pursues efficiency gains in increasingly digitalized business models (Bushman & Smith, 2003).

From this standpoint, technological disclosure extends beyond technical description and becomes an indicator of operational efficiency and organizational risk management. Because digital transformation requires substantial investments and modifies firms' cost and revenue structures, transparency regarding these processes allows investors to evaluate a company's strategic resilience in the face of technological obsolescence. Consequently, signaling digital advancements contributes to reducing perceived risk and, by extension, the cost of equity capital, as firms provide the market with information that facilitates the valuation of intangible assets and dynamic capabilities under lower levels of uncertainty (Bushman & Smith, 2003). Therefore, integrating digital transformation into corporate reporting serves as a communication channel that validates the long-term viability of business models operating within an increasingly digitalized economy.

In this sense, Disclosure Theory encompasses the economic incentives that drive corporate transparency, functioning as a mechanism for reducing adverse selection costs and attracting potential investors. According to Verrecchia (2001), this theoretical framework examines the determinants that lead organizations to disclose information beyond mandatory

requirements, characterizing voluntary disclosure as a strategic choice. In the context of this study, the disclosure of digital transformation initiatives falls within this voluntary category, whereby management chooses to communicate technological progress in order to strengthen its image of modernization and competitiveness in the capital market.

Although extensively discussed in the literature, Disclosure Theory lacks a universally accepted definition. Verrecchia (2001) focused on understanding the economic incentives underlying firms' decisions to disclose or withhold information. His study proposed a taxonomy comprising three broad categories of disclosure research: association-based disclosure, discretionary disclosure, and efficiency-based disclosure. The latter is considered a promising starting point for the development of a comprehensive disclosure theory.

Conversely, Dye (2001) argues that there is no theory of mandatory disclosure, but rather a theory of voluntary disclosure, in which firms benefit from disclosing only information that portrays them favorably, as they have little incentive to reveal information that may negatively affect their interests. Accordingly, the author highlights the absence of a general disclosure theory within accounting research.

Disclosure Theory emphasizes transparency as a means of attracting and retaining investors. Within this context, organizational processes and value chains have increasingly embraced digital transformation, a trend widely valued by investors. Consequently, reporting investments in digital innovation serves as a signal of firms' technological viability, becoming an important analytical parameter for market participants seeking to identify long-term organizational value (Bushman & Smith, 2003).

The relevance of adopting new technologies for stakeholders lies in the ability of these tools to redefine operational efficiency and organizational transparency. For investors and creditors, technologies such as Artificial Intelligence and Big Data represent not only technical modernization but also indicators of reduced risk and lower transaction costs (Bushman & Smith, 2003). Furthermore, disclosure regarding technological infrastructure allows the market to value intangible assets more accurately, thereby reducing information asymmetry. As a result, digital transformation becomes a decision-making criterion by signaling a firm's adaptability to the digital economy and its capacity to generate long-term value.

Several studies have examined the relationship between accounting and digital transformation. Chen *et al.* (2022) analyzed publicly listed firms between 2007 and 2020 to investigate the impact of digital transformation on accounting information quality based on financing constraint and agency theories. Their findings indicate that digital transformation is not merely symbolic but constitutes a mechanism capable of significantly influencing business operations by mitigating agency problems, such as information asymmetry. These effects are particularly relevant among firms with limited media coverage, lower market competition, and reduced exposure to cyberattacks.

Similarly, Yang *et al.* (2024) examined the impact of digital transformation on accounting information quality among companies listed in Chinese capital markets between 2013 and 2019. Their results demonstrate that digitalization creates new opportunities for accounting by improving accounting information comparability and effectively reducing information asymmetry within firms.

Borrero-Domínguez *et al.* (2024) combined text mining and panel data techniques to investigate current digital transformation disclosure practices among European companies. Their findings indicate that firms voluntarily disclose information related to digital transformation; however, no standardized reporting pattern currently exists. Moreover, the study revealed that firm size, industry sector, the COVID-19 pandemic, and CEO gender significantly influence voluntary digital transformation disclosure.

Although the literature does not provide a universally accepted definition of digital transformation (DT), it is generally understood as the integration of technological processes and operations throughout an organization in pursuit of innovative and strategic improvements to its activities. According to Schallmo *et al.* (2017), digital transformation requires capabilities to capture, select, analyze, and communicate data in ways that generate useful information.

Digital transformation is fundamentally centered on technologies that may be implemented individually or combined with other technological solutions to improve organizational performance (Facin *et al.*, 2022). According to Saarikko *et al.* (2020), it is a process that enables firms to adapt to new conditions, organizational forms, and competencies required to remain viable within a digital environment.

The implementation of specific technologies such as blockchain illustrates how digital transformation redefines accounting reliability. Wang and Kogan (2018) proposed transaction-processing frameworks based on cryptographic mechanisms capable of providing continuous monitoring and fraud prevention. This perspective of immutable, real-time data aligns with Schallmo *et al.* (2017), who emphasize that the effectiveness of digital transformation depends on an organization's ability to convert raw data into useful and secure information for decision-making purposes. However, implementing such technologies requires substantial financial investment, making disclosure regarding these projects a competitive differentiator in the eyes of investors.

The impact of such communication directly influences organizational value perception. Guo and Huang (2024) examined the relationship between digital transformation tone and the cost of equity, concluding that positive signaling effectively reduces equity financing costs. This finding is consistent with the arguments advanced by Bushman and Smith (2003), who suggest that transparency regarding technological infrastructure allows markets to value intangible assets more accurately, thereby reducing information asymmetry. Consequently, voluntary disclosure not only strengthens corporate communication but also serves as a determinant of financial decisions by mitigating perceived risk among external stakeholders.

According to Wang and He (2024), the integration of technologies such as Big Data and the Internet of Things (IoT) improves resource allocation and enables agile responses to market demands. This view is reinforced by Facin *et al.* (2022), who argue that such technologies generate benefits that extend beyond operational improvements, fostering value creation through product and service innovation. Therefore, as advocated by Wang and He (2024), digital transformation functions as a differentiation mechanism that provides the accuracy and speed required in today's economic environment.

These technologies generate benefits associated not only with improving existing operational performance but also with creating value through innovation in products, services, and business models (Facin *et al.*, 2022). Nevertheless, studies indicate that this evolution represents a significant challenge for both organizations and professionals, as successful digital transformation requires more than simple technological adoption. It demands strategic restructuring to ensure effective implementation (Saarikko *et al.*, 2020).

Wang and He (2024) further argue that the precision, accuracy, and speed associated with these technologies can effectively generate competitive advantages. Consequently, they view digital transformation as an important channel through which firms can articulate strategies of efficient integration and innovation in order to achieve market differentiation.

Within the accounting profession, the application of these technologies remains in a consolidation phase, resulting in substantial changes to operational routines and service delivery models (Souza *et al.*, 2023). In this context, Schiavi (2020) emphasizes that accounting professionals must develop new competencies, including data analytics and strategic consulting skills, in order to maintain competitiveness. This need for adaptation extends across several areas

of accounting, influencing the strengthening of internal controls (Manita *et al.*, 2020), the efficiency of financial audits (Ferreira & Oliveira, 2024), strategic governance, and the promotion of sustainable growth (Esposito *et al.*, 2023; Guandalini, 2022).

Accordingly, this study adopts a strategic perspective by providing empirical evidence on digital transformation disclosure in the Brazilian capital market. By mapping corporate discourse, it contributes to the literature on technological transparency and offers an analytical foundation for future investigations into the disclosure of technological innovation and its implications for capital market participants.

3 METHODOLOGICAL PROCEDURES

This study is characterized as descriptive and quantitative (Gil, 2022). The research approach is based on identifying patterns of frequency and relevance of terms in management reports, drawing upon the methodological framework adopted in recent studies on digital transformation in capital markets (Chen *et al.*, 2024; Yang *et al.*, 2024; Zhang & Yang, 2024).

In this context, the study seeks to identify how digital transformation-related terms are disclosed by publicly traded companies.

The sample consisted of 281 publicly traded companies listed on B3 with data available on the website of the Brazilian Securities and Exchange Commission (CVM). The analysis period spans from 2010 to 2023, a timeframe that coincides with the full adoption of International Financial Reporting Standards (IFRS) in Brazil. The process used to construct the final sample followed the stages described in Table 1:

Table 1
Research Sample Composition

Sample Selection Stages	Number of Reports
Initial Population of Reports (2010-2023)	3.061
(-) Reports with No Identified Search Terms	2.147
Final Sample Analyzed	**914**

Source: Research Data (2025).

In this context, the sample investigated was obtained from the annual management reports of publicly traded companies available on the website of the Brazilian Securities and Exchange Commission (CVM), covering the period from 2010 to 2023. This timeframe is justified by the full mandatory adoption of the International Financial Reporting Standards (IFRS) in Brazil in 2010. The initial sample comprised 3,061 reports, of which 913 contained the keywords of interest, representing approximately 29% of the total sample and including only those reports that provided the information necessary for the analysis.

The data analyzed in this study, relating to digital transformation, were obtained through the examination of management reports using the R programming language within the RStudio environment. Data collection was conducted through a keyword-based search strategy. The search terms were selected based on the international literature, ensuring that the lexical framework employed adequately reflected contemporary dimensions of digital transformation. Given the lack of standardization in Brazilian corporate reports, the search included terms in both Portuguese and English (Table 2).

Table 2

Search Terms/Keywords Used in the Study and Their Theoretical References

Technological Dimension	Keywords	Reference Authors
Data and Information Technologies	*Artificial intelligence, blockchain, cloud computing and big data.*	Chen *et al.* (2024); Yang *et al.* (2024).
Artificial Intelligence and Machine Learning	*Artificial intelligence; Machine learning; Deep learning; Neural networks; Big data analytics.*	Zhang & Yang (2024).
Digital Maturity and Operations	*Automation, Fintech, Digitization*, Digital.	Prepared by the authors based on *et al.* (2022).

Source: Research Data (2025).

The definition of the search lexicon followed a three-step criterion. First, terms well established in the international accounting and technology literature were adopted (Chen *et al.*, 2024; Zhang & Yang, 2024). Second, a technical translation and adaptation to the Brazilian context were performed, while retaining English terms that are commonly used without translation in corporate reports, such as *blockchain* and *big data*. Finally, a sensitivity test was applied to the initial sample to identify terminological variations, including *automation* and *digitalization*, based on their alignment with the framework proposed by Facin *et al.* (2022), which conceptualizes digital transformation as a process that extends beyond the adoption of isolated technological tools. Accordingly, the final set of keywords was designed to be sufficiently comprehensive to capture the phenomenon across different stages of corporate digital maturity.

After defining the terms to be identified in the companies' management reports, the frequency of digital transformation-related content was assessed through the occurrence of specific keywords. The search was conducted using both Portuguese and English terms, given the absence of standardized terminology in corporate reports and the possibility that these expressions may appear in either language.

Table 3

Definition of the Search Terms/Keywords Used in the Study.

Keywords	Definition
Artificial intelligence – Inteligência artificial	A set of technologies that enables computers to perform a wide range of advanced functions, including the ability to perceive, understand, and translate spoken and written language, analyze data, and generate recommendations.
Automation – Automação	The application of technology, software, robotics, or processes to achieve outcomes with minimal human intervention.
Big data analytics – Análise de big data	A technology that enables the processing of both structured and unstructured data.
Big data – Megadados ou grandes volumes de dados	A collection of extremely large and complex datasets.
Blockchain – Cadeia de blocos	A technology that organizes and stores information in interconnected blocks secured through cryptographic techniques.
Cloud – Nuvem	A vast network of remote servers distributed across the globe that are interconnected and operate as a unified ecosystem.

Deep learning – Aprendizado profundo	An artificial intelligence technique that uses deep neural networks to automatically learn and extract complex patterns from large volumes of data.
Digital – Digital	The predominant use of the latest computer-based technologies to improve organizational processes.
Digitization – Digitalização	The use of digital technologies to transform business models and create new opportunities for revenue generation and value creation.
Fintech – Tecnologia financeira	Technology and innovation applied to the provision of financial services and solutions.
Machine learning – Aprendizado de máquina	A data analysis method that automates the development of analytical models.
Neural networks – Redes neurais	An artificial intelligence method that enables computers to process data in a manner inspired by the functioning of the human brain

Source: Research Data (2025).

The analysis was conducted using RStudio and employed a Text Mining methodology. According to Aranha and Passos (2006), this field enables the automated extraction of knowledge from unstructured textual databases. The process followed five rigorous stages: (i) selection of PDF documents; (ii) data cleaning and preparation, including tokenization and stopword removal; (iii) indexing and normalization; (iv) calculation of term relevance and weighting; and (v) post-processing for the generation of word clouds and co-occurrence graphs.

Text mining, as an extension of data mining, was applied to automate the extraction of knowledge from management reports (Aranha & Passos, 2006; Feldman & Dagan, 1995). The suitability of this technique for the present study is justified by the need to process a large volume of unstructured data comprising 3,061 documents, thereby ensuring objectivity in identifying patterns of technological disclosure.

Term relevance was operationalized using the TF-IDF (Term Frequency-Inverse Document Frequency) metric, which measures word importance at both local and global levels. The rationale behind this metric is based on the assumption that terms with different frequencies carry different informational weights. Words that occur frequently within a specific report but are rare throughout the remainder of the corpus receive higher scores because they possess greater signaling power (Morais & Ambrósio, 2007; Wives, 2002). In the context of this study, such a distinction is essential for mitigating the influence of common operational terms while highlighting specific disruptive innovations. Thus, the calculation balances the individual frequency of a term with its statistical rarity across the sample, ensuring that the final weight reflects the company's actual technological disclosure strategy rather than the mere repetition of generic terminology.

The relevance metrics employed for data processing follow established standards in the text mining and natural language processing literature. Equations 1 through 5, corresponding to the calculations of Absolute Frequency, Relative Frequency, Inverse Frequency, TF-IDF weighting, and Co-occurrence Analysis, are based on the methodologies proposed by Aranha and Passos (2006) and Morais and Ambrósio (2007). These mathematical models ensure that the final weight assigned to each term accurately reflects its statistical and informational importance within the analyzed corpus, eliminating biases associated with document length or the repetition of generic terms.

The first measure refers to absolute frequency, as presented in Equation 1:

$$TF = N.TF\ in\ NTF \quad (1)$$

Where:
TF = Term Frequency
NTF = Number of times the term appears in the document.

It represents the number of times a term appears within a specific document. According to the reference literature, this metric is useful for indicating the importance of a term within an individual document; however, it may not be sufficient to distinguish terms that are widely used across multiple documents.

Relative frequency takes document length into account and is expressed as shown in Equation 2:

$$F_{rel}(x) = \frac{F_{abs\,(x)}}{N} \quad (2)$$

Where:

F_{abs} = Absolute frequency
N = Total number of words in the same document

This frequency represents the proportion of a term relative to the total number of processed words, thereby providing a comparative measure of the relevance of each term within the analyzed corpus.

The inverse document frequency is based on the informational importance of a term and is expressed as shown in Equation 3:

$$[Weight]_{td} = \frac{Freq_{td}}{DocFreq_{td}} \quad (3)$$

Where:
$Weight]_{td}$: degree of association between term t and document d;
$Freq_{td}$: number of times term t appears in document d;
$DocFreq_{td}$: number of documents in which term *t* appears.

This measure refers to the calculation of a term's importance relative to the total number of documents in the corpus. It is essential for preventing terms that occur frequently across all documents from being incorrectly classified as highly relevant. In other words, it increases the importance of terms that appear in relatively few documents while reducing the importance of terms that appear in many documents, based on the premise that low-frequency terms are generally more discriminative.

To complement the analysis of term relevance, a binary co-occurrence model was implemented, as represented in Equation 4:

A formula that measures the degree of association between a term and the document in which it appears.

$$d_{ij} = tf_{ij} \times log\frac{N}{df_i} \quad (4)$$

Where:

d_{ij} = combined weight of term *j* in document *i*;
tf_{ij} = frequency of term *j* in document *i*;

N = total number of documents in the corpus;
df_j = document frequency of term j, that is, the number of documents in which term j appear).

A formula used to analyze the co-occurrence of terms within the texts, based on the results generated by the previous formula).

$$\text{Co-occurrence} = \frac{\sum_{i=1}^{n} d_{ijk}}{\sum_{i=1}^{n} d_{ij}}, \text{ para } \quad d_{ijk} = tf_{ijk} \times log \frac{N}{df_{ij}} \tag{5}$$

Where:

tf_{ij}: number of co-occurrences of terms j and k in document i, calculated as the minimum frequency observed between the two terms within that document);
df_{ij}: number of documents, within the corpus N, in which terms j and k occur simultaneously.

The analysis of term co-occurrence indicates the simultaneous presence of pairs of terms within documents. This approach makes it possible to identify patterns of association among keywords and to explore semantic relationships within the analyzed corpus.

Finally, following the description of the methodology employed, the results are presented and examined through the visualization and analysis of keywords using a word cloud, as well as through absolute, relative, and inverse frequency measures and co-occurrence analysis, with the aim of addressing the research objective proposed in this study.

4 RESULTS

Through the analysis of the annual management reports of the publicly traded companies included in the sample, 5,240 keywords related to digital transformation were identified. Among these, the most prominent terms were fintech, cloud, digitalization, digital, blockchain, automation, and cloud computing. Figure 1 presents the study's keywords through a word cloud visualization.

Figure 1
Keyword Cloud



Source: Research Data (2025).

The relatively small size of the word cloud does not indicate an analytical limitation but rather a relevant finding regarding the maturity stage of companies in disclosing digital transformation initiatives. This pattern suggests that firms may prioritize generic terminology rather than explicitly referring to disruptive technologies. The word cloud reflects the prominence of terms associated with digital transformation and their disclosure within management reports. The emphasis on terms such as *fintech* and *cloud* suggests a focus on emerging technologies and cloud-based solutions, whereas *automation*, *digital*, and *digitalization* indicate concerns related to modernization and operational efficiency.

Table 4 presents the absolute and relative frequencies of the terms comprising the word cloud.

Table 4
Frequency distribution of digital transformation terms

Terms	Absolute frequency	Relative frequency
Aprendizado de Máquina	0	0
Aprendizado Profundo	0	0
Inteligência Artificial	0	0
Automação	509	0.097137405
Big Data	0	0
Blockchain	19	0.003625954
Nuvem	116	0.022137405
Computação em Nuvem	0	0
Computação em Nuvem	0	0
Inteligência de Dados	0	0
Aprendizado Profundo	0	0
Digital	3988	0.761068702
Digitalização	476	0.090839695
Fintech	67	0.01278626
Big Data	0	0
Inteligência Artificial	0	0
Inteligência de Dados	0	0
Aprendizado de Máquina	0	0
Redes Neurais	0	0
Nuvem	65	0.01240458
Redes Neurais	0	0

Source: Research Data (2025).

Table 4 presents the frequency distribution of terms associated with digital transformation. The results indicate that *digital* is by far the most recurrent term, accounting for 76% of the relative frequency, followed by *automation* (10%) and *digitalization* (9%). Terms associated with more advanced technologies showed low incidence, including *fintech* (2%), *cloud* and *cloud computing* (1%), and *blockchain* (0.36%), while the remaining terms recorded no occurrences.

The findings reveal that the most frequently disclosed digital transformation-related terms in management reports are *digital*, *automation*, and *digitalization*. These terms broadly signal that companies are employing technological solutions to support the development and improvement of their processes. In contrast, terms associated with more advanced technologies, such as *artificial intelligence*, *machine learning*, *deep learning*, *big data*, *data intelligence*, and *neural networks*, either recorded no occurrences or appeared only marginally, as observed for *blockchain* and *fintech*. This pattern suggests that many companies are either not investing in these technologies or are not disclosing their adoption. These findings are largely consistent with those reported by Borrero-Domínguez *et al.* (2024), differing only with respect to a limited number of specific terms.

However, regarding the focus of the technologies disclosed, the results of the present study diverge from those reported in previous research. While earlier studies indicate that major European companies included in the EuroStoxx50 frequently report *Artificial Intelligence* among

the four most common terms in their annual reports (Borrero-Domínguez *et al.*, 2024), no occurrence of this term was identified in the present study. This finding suggests a distinct stage of technological disclosure maturity among Brazilian companies.

The limited disclosure of advanced technologies is also observed in the Chinese context, where only 24.3% of the sampled firms reported the use of digital systems during the analyzed period (Chen *et al.*, 2024). Furthermore, the predominance of generic terms such as *digital* and *digitalization* among B3-listed companies indicates that disclosure practices remain concentrated on broad concepts rather than specific technological applications. In contrast, European publicly traded firms, although also using *digitalization* as their most frequent term, complement their disclosures with a substantial number of technology-specific and tool-related codes, including *big data*, *blockchain*, and *cloud computing* (Borrero-Domínguez *et al.*, 2024). A similar pattern has been observed in Asia, where Yang *et al.* (2023) found that companies frequently employ generic terms such as *digital* and *automation*, while increasingly highlighting emerging technologies such as *artificial intelligence* and *big data*, albeit at lower frequencies than those reported in Europe. Collectively, these findings suggest that the adoption and disclosure of advanced technologies evolve at different rates across countries, reinforcing the notion of heterogeneous levels of maturity in corporate digital transformation.

Table 5
Inverse frequency distribution

Terms	**Inverse frequency**
Automation	1.109625616
Blockchain	4.044147158
Nuvem	2.689601496
Digital	0.31095182
Digitalization	1.414058499
Fintech	3.382748676
Cloud	3.079066262

Source: Research Data (2025).

To improve the quality of the information and prevent highly frequent terms from being erroneously considered relevant, weights are assigned to words in order to increase or decrease their importance. Accordingly, Table 5 presents the inverse frequency of the terms, which reflects the relative importance of each term within the document corpus. This measure increases the importance of terms that appear in relatively few documents while reducing the importance of those that occur across many documents, based on the premise that low-frequency terms are generally more discriminative. Therefore, the higher the inverse frequency value, the more relevant and informative the term is considered to be.

Within this context, the terms *blockchain* and *fintech*, which are more specific, relevant, and informative, may be interpreted as examples of strategic voluntary disclosure, consistent with Disclosure Theory (Verrecchia, 2001; Dye, 2001). These terms signal technological innovation and competitive advantage, standing out because they exhibit the highest inverse frequency values. Yang *et al.* (2023) support this interpretation by demonstrating that more advanced technological terms tend to receive greater weight in corporate documents, indicating that firms strategically emphasize such technologies to signal value creation to stakeholders.

Accordingly, these terms exhibit the highest weights among the digital transformation-related keywords, suggesting that companies still disclose these technologies at a relatively early stage. In contrast, the term *digital* presents a low weight, indicating that its disclosure is weakly

discriminative and already widely expected by stakeholders. This finding aligns with the notion that the disclosure of common information generates limited incremental value (Verrecchia, 2001; Dye, 2001).

The term *blockchain* signals the use of advanced technologies in contemporary markets. As an emerging technology that gained prominence through the rise of the cryptocurrency Bitcoin, blockchain has increasingly attracted attention due to its ability to enhance transparency, reliability, and efficiency (Falcão *et al.*, 2021). It is considered a disruptive technology for organizations and remains in an evolutionary stage of adoption. This characteristic helps explain its low disclosure frequency while reinforcing the logic of selective disclosure aimed at maximizing stakeholders' perceptions of organizational value.

The term *fintech* also exhibited a relatively high weight, which may be explained by its direct association with the use of technology to innovate financial services. Given that firms from the financial sector were excluded from this study, it can be inferred that the companies disclosing this term in their reports are signaling the adoption of new digital transformation tools and initiatives.

These findings are consistent with the arguments presented by Facin *et al.* (2022), who contend that such technologies enhance operational performance and foster value creation. For this reason, companies disclose information regarding these technologies in their reports, highlighting their technological advancements and reinforcing their market positioning.

Conversely, the term *digital* exhibited a lower weight across the analyzed documents, likely because it has become a widely disseminated concept appearing in most reports. From the perspective of Disclosure Theory, this pattern suggests the disclosure of information with limited strategic value. Consequently, the term may be considered to have a relatively low informational impact, as it forms part of standard corporate discourse and contributes little novelty when reported by organizations.

Furthermore, the terms *cloud*, *digitalization*, and *automation* displayed intermediate weights. These terms indicate process modernization and technological advancement; however, they do not necessarily characterize disruptive technologies. This finding supports the argument that corporate disclosure is selective and oriented toward maximizing the informational relevance of disclosed content for stakeholders (Verrecchia, 2001; Dye, 2001; Facin *et al.*, 2022).

Table 6
Co-occurrence of terms across simultaneous sets of documents

Term 1	Term 2	Co-occurrence
Automation	Cloud	10
Blockchain	Automation	5
Cloud	Digitalization	33
Cloud	Digital	54
Cloud	Blockchain	6
Cloud	Automation	21
Cloud	Cloud Computing	23
Digital	Blockchain	13
Digital	Automation	98
Digital	Cloud	33
Digitalization	Digital	186
Digitalization	Blockchain	9

Digitalization	Automation	46
Digitalization	Cloud Computing	23
Fintech	Cloud	7
Fintech	Digitalization	18
Fintech	Digital	28
Fintech	Blockchain	3
Fintech	Automation	8
Fintech	Cloud	3

Source: Research Data (2025).

Table 6 illustrates the co-occurrence relationships among the analyzed terms. This measure captures the simultaneous presence of pairs of terms within documents and allows the identification of association patterns among keywords, that is, the number of times these terms appear together across the analyzed reports.

Within this context, the terms *digital* and *digitalization* exhibited a strong association, likely due to the close relationship between these concepts from both linguistic and conceptual perspectives.

Additionally, *automation* and *cloud* showed strong correlations with *digital*. This relationship is consistent with the notion that digital evolution often drives the automation of organizational processes. Similarly, cloud computing provides the essential infrastructure required to support digital transformation, since companies adopting digital solutions typically depend on cloud-based platforms for data storage, processing capabilities, and remote access.

In contrast, the terms *fintech* and *blockchain* exhibited low frequencies and weak associations with other terms. This finding suggests that, despite their recognized importance within the context of technological innovation, companies are still not emphasizing these technologies in their corporate reports. The limited disclosure of these terms may indicate that, although such technologies are considered important, they have not yet been prioritized or widely implemented within publicly disclosed digital transformation strategies.

5 CONCLUSION

Considering the growing importance of digital transformation and the scarcity of studies dedicated to this topic, the present study examined how digital transformation disclosure is presented by publicly traded companies listed on B3 during the period from 2010 to 2023. This analysis was conducted through text mining techniques, employing absolute frequency, relative frequency, inverse frequency, and co-occurrence analyses in order to address the proposed research objective.

The findings indicate that relatively few Brazilian companies disclose information regarding digital transformation in their reports. The initial sample comprised 3,061 reports; however, due to the absence of relevant keywords, the final sample was reduced to 913 reports, representing approximately 29% of the total. Moreover, among the companies that do disclose digital transformation-related information, disclosure tends to remain superficial, with limited discussion of the specific implementation and use of advanced digital technologies.

The results further reveal that the most recurrent terms are broad concepts such as *digital*, *automation*, and *digitalization*. Terms associated with more advanced technologies appear at considerably lower frequencies, with *blockchain*, for example, accounting for only 0.36% of occurrences. To identify the most specific, relevant, and informative terms, inverse frequency analysis was employed, assigning greater weight to keywords indicative of advanced technologies. In addition, the co-occurrence analysis showed that the most strongly associated

terms tend to represent closely related concepts. The limited presence of more advanced technologies suggests that, although such technologies are important, they have not yet been prioritized or widely implemented within publicly disclosed digital transformation strategies.

From the perspective of Disclosure Theory, the predominance of generic terms such as *digital* and *automation*, rather than more disruptive technologies such as *blockchain* and *fintech*, suggests that digital transformation signaling in Brazil remains at an early stage of development. This evidence indicates that, although technology is recognized as a source of value creation, companies tend to provide less informative disclosures regarding advanced technological tools, possibly to avoid revealing proprietary strategies or because their level of technological implementation has not yet reached a stage that justifies more detailed reporting.

The primary contribution of this study lies in integrating the empirical metrics of text mining with the theoretical foundations of accounting research. From a scholarly perspective, the study advances the literature by mapping the digital transformation lexicon within the Brazilian capital market and providing evidence that corporate discourse does not evolve at the same pace as disruptive technological innovation. Understanding and emphasizing technological innovation within the digital transformation process can improve operational efficiency, reduce costs, and enhance a company's ability to meet investors' expectations. Furthermore, the study contributes to the mapping of corporate discourse related to digital transformation, supporting future research in this area while also extending the empirical application of Disclosure Theory by demonstrating how companies communicate the adoption of digital transformation tools and initiatives.

Additionally, the findings suggest that relying solely on keyword searches within corporate reports is insufficient to determine whether a company is truly undergoing digital transformation. In other words, textual descriptions alone do not provide enough evidence to confirm the existence or extent of digital transformation initiatives. Consequently, future analyses should incorporate the classification of terms and industry-specific contexts, as well as direct examination of report content, to better assess the scope and depth of digital transformation efforts.

Indeed, the disclosure of information related to digital transformation, innovation, corporate strategy, and long-term vision has become increasingly important to stakeholders. Investors and other stakeholders are paying closer attention to such information when making decisions and no longer rely exclusively on financial data. Instead, they seek a broader understanding of companies' future prospects and strategic positioning. Therefore, strengthening the disclosure of digital transformation initiatives is essential for understanding its impact on corporate communication and financial decision-making, as this aspect has become an increasingly relevant criterion in stakeholder evaluations.

It is important to acknowledge that this study has limitations, particularly its focus on individual words rather than the broader contextual meaning embedded in corporate reports. To address this limitation, future research should further investigate how financial constraints and industry-specific characteristics moderate digital transformation disclosure. Additionally, the application of more advanced Natural Language Processing (NLP) techniques is recommended to enable richer contextual analyses and provide deeper insights into the nature and quality of digital transformation disclosures.

REFERENCES

Aranha, C., & Passos, E. (2006). A tecnologia de mineração de textos. *Revista Eletrônica de Sistemas de Informação*, *5*(2). https://doi.org/10.21529/RESI.2006.0502001

Borrero Domínguez, C., Cortijo Gallego, V., & Escobar Rodríguez, T. (2024). Digital transformation voluntary disclosure: Insights from leading European companies. *International Journal of Accounting Information Systems*, *55*, 100711. https://doi.org/10.1016/j.accinf.2024.100711

Bushman, R. M., & Smith, A. J. (2003). Transparency, financial accounting information, and corporate governance. *Economic Policy Review*, *9*(1), 65–87.

Chen, W., Zhu, J., & Zhang, Y. (2024). Gimmick or revolution: Can corporate digital transformation improve accounting information quality? *International Journal of Emerging Markets*, *19*(10), 2966–2990. https://doi.org/10.1108/IJOEM-03-2023-0417

Dye, R. A. (2001). An evaluation of "Essays on disclosure" and the disclosure literature in accounting. *Journal of Accounting and Economics*, *32*(1–3), 181–235. https://doi.org/10.1016/S0165-4101(01)00029-0

Esposito, P., Perrone, G., Ricci, P., & Verde, F. (2023). The strategic governance of the digital accounting environment: Insights from virtual museums. *Meditari Accountancy Research*, *31*(2), 366–380. https://doi.org/10.1108/MEDAR-03-2020-0837

Facin, A. L. F., Silva, D., Oliveira, M. P. V. de, & Ibanes, J. (2022). Temas de destaque na pesquisa em transformação digital: Evidências de estudo bibliométrico e análise de conteúdo. *Revista de Administração de Empresas*, *62*(6), e2021-0112. https://doi.org/10.1590/S0034-759020220607

Falcão, A. I. L., Oliveira, T. F. A. de, & Farias, R. S. de. (2021). Blockchain: Tendência para a contabilidade digital. *Revista Liceu On-Line*, *11*(2), 6–27.

Feldman, R., & Dagan, I. (1995). Knowledge discovery in textual databases (KDT). In *Proceedings of the First International Conference on Knowledge Discovery and Data Mining* (pp. 112–117).

Ferreira, J. S., & Oliveira, M. C. de. (2024). Impactos da inteligência artificial na auditoria contábil: Explorando transformações e desafios. *Revista Ibero-Americana de Humanidades, Ciências e Educação*, *10*(6), 2742–2754. https://doi.org/10.51891/rease.v10i6.14591

Franco, G., Lima, M. V., & Ferreira, J. (2021). Contabilidade 4.0: Análise dos avanços dos sistemas de tecnologia da informação no ambiente contábil. *CAFI*, *4*(1), 55–73. https://doi.org/10.23925/cafi.v4i1.53860

Fredo, A. R. (2021). *Transformação digital: A percepção dos profissionais de contabilidade* (Dissertação de mestrado). Universidade de Caxias do Sul.

Gil, A. C. (2022). *Como elaborar projetos de pesquisa* (7ª ed.). Atlas.

Guandalini, I. (2022). Sustainability through digital transformation: A systematic literature review for research guidance. *Journal of Business Research*, *148*, 456–471. https://doi.org/10.1016/j.jbusres.2022.05.015

Guo, B., & Huang, X. (2024). Digital transformation tone signal and the cost of equity: Evidence from Chinese listed companies. *Finance Research Letters*, *59*, 104720. https://doi.org/10.1016/j.frl.2023.104720

Iudícibus, S. de. (2021). *Teoria da contabilidade*. Grupo GEN.

Manita, R., Elommal, N., Baalousha, M., & Hussainey, K. (2020). The digital transformation of external audit and its impact on corporate governance. *Technological Forecasting and Social Change*, *150*, 119751. https://doi.org/10.1016/j.techfore.2019.119751

Merlugo, W. Z., Carraro, W. B. W. H., & Pinheiro, A. B. (2021). Transformação digital na contabilidade: Os contadores estão preparados? *Revista Pensamento Contemporâneo em Administração*, *15*(1), 180–196. https://doi.org/10.12712/rpca.v15i1.48122

Morais, E. A. M., & Ambrósio, A. P. L. (2007). *Mineração de textos* (Relatório técnico). Instituto de Informática, Universidade Federal de Goiás.

Nemoto, M. C. M. O., Santos, G. Z. V., & Pinochet, L. H. C. (2018). Adoção de inovação: Internet das Coisas para melhoria de desempenho de sustentabilidade na Klabin. *Revista Gestão & Tecnologia*, *18*(1), 197–224. https://doi.org/10.20397/2177-6652/2018.v18i1.1215

Ruschel, M. E., Frezza, R., & Utzig, M. J. S. (2011). O impacto do SPED na contabilidade: Desafios e perspectivas do profissional contábil. *Revista Catarinense da Ciência Contábil*, *10*(29), 9–26. https://doi.org/10.16930/2237-7662/rccc.v10n29p9-26

Saarikko, T., Westergren, U. H., & Blomquist, T. (2020). Digital transformation: Five recommendations for the digitally conscious firm. *Business Horizons*, *63*(6), 825–839. https://doi.org/10.1016/j.bushor.2020.06.009

Schallmo, D., Williams, C. A., & Boardman, L. (2017). Digital transformation of business models: Best practice, enablers, and roadmap. *International Journal of Innovation Management*, *21*(8), 1740014. https://doi.org/10.1142/S1363919617400142

Schappo, B. H., & Martins, Z. B. (2022). A utilização de tecnologia na contabilidade: Uma percepção de profissionais contábeis do estado de Santa Catarina. *ConTexto – Contabilidade em Texto*, *22*(50), 2–15. https://seer.ufrgs.br/index.php/ConTexto/article/view/118089

Schiavi, G. S., Beal, A., & Lopes, J. E. G. (2020). No caminho da inovação: Análise das capacidades de inovação de empresas contábeis diante das tecnologias digitais. *Revista Brasileira de Gestão de Negócios*, *22*(2), 381–405. https://doi.org/10.7819/rbgn.v22i2.4051

Souza, P. L., Pacheco, W. O., Ferreira, R. C., & Andrade, A. M. de. (2023). Inteligência artificial e contabilidade: Uma aliança estratégica para o futuro profissional no Brasil. *Revista Contemporânea*, *3*(9), 14928–14951. https://doi.org/10.56083/RCV3N9-079

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics, 87*(3), 355–374. https://doi.org/10.2307/1882010

Verrecchia, R. E. (2001). Essays on disclosure. *Journal of Accounting and Economics*, *32*(1–3), 97–180. https://doi.org/10.1016/S0165-4101(01)00024-1

Vial, G. (2019). Understanding digital transformation: A review and research agenda. *The Journal of Strategic Information Systems*, *28*(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003

Wang, Y., & Kogan, A. (2018). Designing confidentiality-preserving blockchain-based transaction processing systems. *International Journal of Accounting Information Systems*, *30*, 1–18. https://doi.org/10.1016/j.accinf.2018.07.001

Wang, Y., & He, P. (2024). Enterprise digital transformation, financial information disclosure and innovation efficiency. *Finance Research Letters*, *59*, 104707. https://doi.org/10.1016/j.frl.2023.104707

Wives, L. K. (2002). *Tecnologias de descoberta de conhecimento em textos aplicadas à inteligência competitiva* [Exame de qualificação de mestrado, Universidade Federal do Rio Grande do Sul].

Yang, J., Ying, L., & Xu, X. (2024). Digital transformation and accounting information comparability. *Finance Research Letters*, *61*, 104835. https://doi.org/10.1016/j.frl.2024.104835

Zhang, C., & Yang, J. (2024). Artificial intelligence and corporate ESG performance. *International Review of Economics & Finance*, *96*, 10371. https://doi.org/10.1016/j.iref.2024.10371

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interest regarding this submitted work.

DATA AVAILABILITY

The dataset supporting the results of this study is not publicly available.

AUTHOR CONTRIBUTIONS

Roles	1st author	2nd author	3rd author	4th author	5th author
Conceptualization	♦	♦			
Data Curation	♦	♦	♦		
Formal Analysis	♦	♦	♦	♦	
Funding Acquisition	Not Applicable				
Investigation	♦	♦			
Methodology	♦	♦	♦		
Project Administration	♦	♦			
Resources	♦	♦			
Software	♦	♦	♦		
Supervision	♦	♦			
Validation		♦	♦	♦	♦
Writing – Original Draft Preparation	♦	♦			
Writing – Review & Editing		♦			♦

COPYRIGHTS

The authors retain the copyright relating to their articles and grant the *Revista Catarinense da Ciência Contábil* (RCCC) the right of first publication. The work is simultaneously licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).