



Content Analysis in the Big Data Era: A Systematic Review of Applications and Challenges in Corporate Communication

BOUCHEGOURA Soulef ^{1*}

¹ University of Tizi-Ouzou, soulef.bouchejoura@ummto.dz

Received: 14/07/2026

Accepted: 23/09/2026

Published: 30/09/2026

Doi: 10.53284/2120-013-003-022

Abstract: This study presents a systematic review of the applications of content analysis in corporate communication research in the era of digital data. It examines studies published between 2013 and 2025 retrieved from major academic databases. The review explores the methodological approaches, analytical tools, and technological challenges associated with digital content analysis. The findings indicate a significant expansion in the use of content analysis for reputation management, crisis communication, audience analysis, campaign evaluation, and corporate discourse analysis. They also reveal the growing adoption of automated content analysis, machine learning, natural language processing, and Big Data Analytics. Despite these advances, several challenges persist, including data quality, natural language complexity, algorithmic bias, sampling representativeness, and ethical concerns. The study concludes that integrating qualitative and quantitative approaches through hybrid analytical frameworks can enhance the effectiveness and rigor of corporate communication research in digital environments.

Keywords: Content Analysis; Corporate Communication; Big Data; Automated Content Analysis.



1. INTRODUCTION

The world has witnessed a rapid digital transformation that has led to the generation of enormous volumes of data, commonly referred to as **Big Data**. These data are characterized by the three fundamental dimensions of **Volume, Velocity, and Variety**, making their processing and analysis through traditional methods increasingly difficult and complex (Al-Shiakhli, 2019). In this context, **content analysis** has emerged as a crucial research methodology for understanding complex communication phenomena. Historically, content analysis has been recognized as a well-established and effective research method in the social sciences (Bardin, 2013). It has greatly benefited from advances in information and communication technologies as well as increased computing power, with computers proving highly efficient in routine tasks such as scanning texts for keywords. Nevertheless, the methodology remains "largely unknown" and continues to be subject to several methodological biases (Eriksson, 2013).

Research on digital content analysis within the context of institutional communication encompasses a multifaceted landscape involving methodological considerations, thematic explorations, and the influence of digital media on organizational communication strategies. The accurate measurement and interpretation of digital content have emerged as recurring methodological challenges in the literature. While digital data provide researchers with valuable opportunities to investigate complex social phenomena, they also present significant methodological difficulties. Consequently, content analysis, as a systematic approach to the analysis of textual and visual communication, has gained renewed attention as an essential tool for examining the vast amount of digital data generated in contemporary communication environments.

In the field of **corporate communication**, digital data have become a rich source of information for understanding corporate reputation, audience responses, campaign effectiveness, and crisis management. Content analysis is no longer limited to the manual examination of texts; it has evolved to incorporate more advanced techniques, including **Automated Content Analysis** and machine learning. Content analysis is defined as "a research technique for the objective, systematic, and quantitative description of the manifest content of communication" (Guédé, 2022). However, the evolution of content analysis in the digital era, particularly with the advent of the Internet, has transformed it into a widely practiced activity rather than one confined to academic research. It is employed not only by researchers conducting systematic analyses of texts but also by virtually anyone using a computer. Technological advances have fundamentally reshaped research practices by making data more

Content Analysis in the Era of Digital Data: Applications and Challenges

in Corporate Communication Research – A Systematic Review



accessible and substantially expanding the volume of information available for analysis (Eriksson, 2013).

Against this backdrop, conducting a **systematic literature review** is essential to synthesize the fragmented body of knowledge concerning the application of content analysis in corporate communication research in the digital era. Such a review can identify prevailing methodological approaches, reveal existing research gaps, and provide recommendations for future studies aimed at addressing the methodological and technological challenges associated with the rapidly evolving digital environment.

2. Objective of the Systematic Review

The primary objective of this systematic review is to provide a comprehensive analysis of the applications of content analysis in corporate communication research in the era of digital data, with particular emphasis on the major methodological and technological challenges encountered by researchers in this field. Specifically, this review aims to:

- Identify and examine how content analysis, both manual and automated, is employed to investigate corporate communication phenomena in digital environments.
- Explore the various tools and techniques, including machine learning and big data analytics, used to analyze digital content.
- Identify the key challenges associated with the use of content analysis in the context of large-scale digital data, including issues related to data quality and methodological complexity.
- Provide a critical comparison of qualitative and quantitative approaches to digital content analysis, highlighting the strengths and limitations of each approach.

3. Literature Review Process

Figure 1 illustrates the literature review process adopted in this study, which consisted of the following methodological steps:

- **Identifying relevant databases and information sources:** The literature search was conducted using a wide range of academic databases and scholarly resources, including **IEEE Xplore, Emerald, EBSCO, Web of Science, Google Scholar, Springer, and Elsevier**. After retrieving the relevant publications, the studies were screened according to predefined inclusion criteria, including publication year, abstract, and research



methodology. Priority was given to high-quality and relevant studies. The search was performed using the keywords "**Big Data**," "**Content Analysis**," and "**Automated Content Analysis**", covering the publication period from **2013 to 2025**.

- **Searching, reading, and analytically reviewing the literature:** This stage involved an in-depth examination of the selected studies to gain a comprehensive understanding of their objectives, methodologies, and principal findings. The process consisted of three successive phases. First, the **preparatory reading**, during which each article was read multiple times and preliminary notes were recorded. Second, the **analytical reading**, aimed at developing a deeper understanding of the study through careful examination of its theoretical and methodological foundations. Finally, the **critical evaluation**, in which each article was systematically assessed to determine its methodological rigor, relevance, and contribution to the field.

Table 1. Summary of the Studies Included in the Systematic Review

Study Title	Methodology	Year of Publication	Author(s)	Key Findings
Big Data Analytics: A Literature Review Perspective	Systematic Literature Review(	2019	Al-Shiakhli, S.	The study provides a comprehensive overview of current Big Data Analytics concepts and their role in decision-making, analyzes the available techniques and tools, and identifies their challenges and applications across various sectors..
Using Digital Content Analysis for Online Research: Online News Media Depictions of Older Adults.	) Digital Content Analysis (	2021	Amundsen, D.	The findings revealed that older adults are predominantly portrayed through negative

Content Analysis in the Era of Digital Data: Applications and Challenges

in Corporate Communication Research – A Systematic Review



				stereotypes in digital news articles, highlighting the need for more respectful and age-inclusive language.
Content Analysis in the Digital Age: Tools, Functions, and Implications for Security	) Content Analysis (	2013	Eriksson, J& ,. Giacomello, G.	The findings highlight the growing application of content analysis in the digital era, particularly in cybersecurity, and demonstrate how advances in computing power have enhanced its effectiveness.
From Content Analysis to Content Analysis of Digital Social Networks: Specificities and Limits of a Particular Content	Theoretical and Empirical Review	2023	Guédé, B. J.-Y.	The findings indicate that the web and digital social networks have reshaped research in the humanities and social sciences, expanding the use of content analysis while introducing new methodological challenges and limitations.



Content Analysis in the Research Field of Corporate Communication	) Content Analysis (	2023	Lischka, J. A.	The findings review the main research topics and methodological designs in corporate communication content analysis, highlighting the role of communication in building intangible organizational resources.
The application of content analysis as a research method in management sciences	) Computer-Aided Text Analysis - CATA (	2023	Miązek, A & „ Światowiec-Szczepańska, J.	The findings demonstrate that content analysis is an effective approach for examining the psychological characteristics of corporate leaders, particularly chief executive officers (CEOs), through the analysis of shareholder letters.
Big Data Analytics Capabilities: A Systematic Literature Review on Necessary Skills to Succeed in Big Data Analytics	) Structured Literature Review (	2022	Richly, M. A.	The findings identify the capabilities required for Big Data Analytics (BDA) in organizations and distinguish them from information technology and digital capabilities.
A new theoretical	A systematic	2021	Sabharwal, R.,	The findings

Content Analysis in the Era of Digital Data: Applications and Challenges

in Corporate Communication Research – A Systematic Review



understanding of big data analytics capabilities in organizations a thematic analysis	literature review following the PRISMA guidelines, combined with thematic analysis using NVivo 12.		&Miah, S. J.	advance a new theoretical understanding of Big Data Analytics (BDA) capabilities in organizations and identify five key themes that highlight limitations in the existing literature.
---	--	--	--------------	---

Source: Prepared by the author.

3.1 Analysis of the Findings of Previous Studies

3.1.1. Applications of Content Analysis in Corporate Communication Research in the Digital Data Era

The reviewed studies demonstrate that the applications of content analysis in corporate communication have expanded considerably in response to the rapid growth of digital data. Rather than being limited to the manual examination of communication messages, content analysis is increasingly integrated with automated techniques, machine learning, and Big Data Analytics to process large volumes of digital content (Al-Shiakhli, 2019).

A common finding across the reviewed literature is the widespread use of content analysis for reputation and crisis management. Several studies reported that organizations employ sentiment analysis to monitor online conversations, detect emerging crises, and evaluate public reactions to brands through social media platforms and digital news content.

Another recurring theme concerns audience understanding and stakeholder identification. The reviewed studies indicate that analyzing user-generated content enables organizations to identify key stakeholders, understand their expectations, and assess public perceptions, thereby supporting more effective communication strategies (Schäfer, 2023).

The findings also reveal that communication campaign evaluation has become one of the most prominent applications of digital content analysis. Researchers highlighted its usefulness in measuring audience engagement, message diffusion, and emotional responses, allowing organizations to assess campaign effectiveness and identify best communication practices (Patil, 2023).



Several studies further demonstrate the growing importance of corporate and leadership discourse analysis. Content analysis has been used to examine CEOs' speeches, shareholder letters, and corporate reports to identify communication strategies, rhetorical framing, and leadership characteristics associated with organizational performance and reputation (Miązek, 2023).

Another emerging application concerns internal communication and organizational culture. The reviewed studies suggest that digital content analysis can reveal communication patterns within organizations, employee engagement, digital skill gaps, and organizational cohesion (Poltoratska, 2025).

Finally, a number of studies emphasize the integration of Big Data Analytics into corporate communication. Their findings indicate that organizations increasingly combine content analysis with big data techniques to support strategic decision-making, improve operational efficiency, and enhance organizational planning (Miah, 2021).

3.1.2. Methodological and Technological Challenges of Digital Content Analysis

Despite these expanding applications, the reviewed studies consistently identify several methodological and technological challenges.

The most frequently reported challenge concerns data volume and quality. Researchers agree that the massive amount of digital data makes manual coding impractical, while automated analysis requires substantial computational resources. Furthermore, digital data are often unstructured, incomplete, or noisy, reducing analytical accuracy (Amundsen, 2022).

Another major challenge involves natural language complexity. Several studies emphasize that sarcasm, negation, ambiguity, and contextual meanings remain difficult for automated content analysis systems to interpret accurately, limiting the reliability of Natural Language Processing (Lischka, 2023).

The reviewed literature also highlights concerns regarding validity and reliability. Although automated methods substantially increase analytical efficiency, many authors argue that their results should be validated through human coding to ensure methodological rigor (Guédé, 2022).

In addition, algorithmic bias emerged as a recurring concern. The reviewed studies warn that algorithms used for data collection and analysis may unintentionally introduce systematic biases that influence research findings and compromise objectivity (Eriksson, 2013).

Ethical considerations represent another important challenge. Several studies underline the need to protect users' privacy, ensure data security, and comply with ethical standards when collecting and analyzing digital data (Amundsen, 2022).



Finally, many studies identify a shortage of researchers with expertise in big data analytics, machine learning, and automated content analysis, highlighting the growing need for interdisciplinary competencies in corporate communication research (Richly, 2022).

Overall, the reviewed literature indicates that content analysis has become an indispensable methodological approach for corporate communication research in the digital era. However, its effective application increasingly depends on integrating computational techniques while addressing methodological, ethical, and technological challenges.

3.1.3. Comparison of the Tools and Techniques Used

The reviewed studies indicate a clear evolution in content analysis techniques, moving from traditional manual approaches toward automated and hybrid methods capable of processing large-scale digital data. Although both approaches remain widely used, the findings suggest that researchers increasingly favor automated techniques due to their scalability, speed, and ability to analyze massive datasets.

Manual content analysis continues to play an important role, particularly in studies requiring contextual interpretation and nuanced understanding of language. The reviewed literature highlights its effectiveness in identifying sarcasm, irony, implicit meanings, and cultural references that remain difficult for computational systems to recognize. However, manual coding is consistently reported as being time-consuming, resource-intensive, and susceptible to inter-coder reliability issues (Amundsen, 2022).

By contrast, automated content analysis has become the dominant approach in digital communication research because it enables researchers to process large volumes of data rapidly, at lower cost, and with greater reproducibility (Lischka, 2023). The reviewed studies identify several major automated techniques.

The first group includes dictionary- and rule-based approaches, which are effective for keyword counting and predefined classifications but show limited ability to capture contextual meanings, negation, or sarcasm.

A second group consists of supervised machine learning methods, which rely on manually coded training datasets to classify new content. The reviewed studies emphasize their high predictive performance when large, high-quality training datasets are available.

The literature also reports increasing use of unsupervised machine learning, particularly topic modeling, to discover hidden themes and patterns without prior labeling. Although these methods are valuable for exploratory research, several studies note limitations regarding model stability and result interpretation.

Another widely adopted technique is sentiment analysis, which combines machine learning



and lexicon-based approaches to classify texts according to their emotional orientation. Despite continuous improvements, the reviewed studies indicate that automated sentiment analysis still performs slightly below expert human coders, especially when interpreting figurative language, irony, and complex contextual expressions.

Several studies further highlight the growing importance of Big Data Analytics, integrating data mining, statistical analysis, and predictive modeling to support organizational decision-making and strategic communication (Miah, 2021).

Finally, the reviewed literature consistently identifies mixed-methods approaches as the most promising direction. By combining manual coding with automated analysis, researchers are able to exploit the contextual accuracy of human interpretation while benefiting from the efficiency and scalability of computational techniques. This integrated approach is increasingly applied in corporate communication, political communication, and digital news analysis (Amundsen, 2022).

Overall, the reviewed studies agree that no single technique is sufficient to address all research objectives. Although automated methods have substantially improved the efficiency of digital content analysis, they continue to face limitations in interpreting linguistic complexity and contextual meaning. Consequently, many authors recommend hybrid analytical frameworks, improved language-processing models, larger annotated datasets, and the integration of complementary data sources, such as web-tracking data, to enhance the validity and reliability of digital content analysis (Al-Shiakhli, 2019).

4. Critical Analytical Review

The reviewed studies provide valuable insights into the use of content analysis in the era of digital data. Nevertheless, a critical examination reveals both significant methodological strengths and persistent limitations, highlighting several research gaps that warrant further investigation.

4.1 Strengths of the Reviewed Studies

The most prominent strength identified across the reviewed literature is the ability of automated content analysis to process large-scale digital datasets efficiently. The integration of machine learning and computational techniques has enabled researchers to analyze vast amounts of textual, visual, and multimedia data that would be impractical to examine manually, thereby expanding the scope of corporate communication research (Amundsen, 2022).

Another notable strength is the capacity of these approaches to generate new insights from unstructured digital data. Several studies demonstrate that automated content analysis facilitates the identification of hidden patterns, communication trends, and audience behaviors, providing valuable evidence for understanding corporate reputation, communication strategies, and stakeholder engagement (Miah, 2021).

The reviewed literature also highlights improvements in methodological transparency and reproducibility. When analytical procedures are systematically documented and standardized, automated techniques enhance research reliability by allowing studies to be replicated under



similar conditions (Amundsen, 2022).

Furthermore, many studies emphasize the value of methodological integration. Combining qualitative and quantitative content analysis, as well as integrating content analysis with survey data or digital trace data, provides a more comprehensive understanding of communication processes and strengthens the validity of research findings (Lischka, 2023).

4.2. Limitations of the Reviewed Studies

Despite these strengths, several methodological limitations remain evident across the reviewed studies.

One of the most frequently reported limitations concerns the interpretation of natural language. Automated techniques continue to experience difficulties in identifying sarcasm, irony, metaphors, negation, and contextual meanings, reducing their accuracy in complex communication settings, particularly during corporate crises (Lischka, 2023).

Another major limitation relates to data quality. The effectiveness of automated content analysis largely depends on the quality of the input data. Social media data are often incomplete, noisy, or contaminated with misleading information, which may compromise analytical accuracy and research validity.

Algorithmic bias represents another recurring concern. Several studies argue that machine-learning models may reproduce biases embedded in their training data or algorithmic design, potentially affecting the objectivity and fairness of analytical outcomes (Eriksson, 2013).

The reviewed literature also reveals the absence of a unified theoretical framework for digital content analysis. Although the use of computational methods has expanded considerably, theoretical development has not progressed at the same pace, limiting the comparability of studies and slowing cumulative knowledge development (Lischka, 2023).

Moreover, many studies acknowledge challenges related to representativeness and sampling. Digital datasets do not necessarily reflect the characteristics of the entire target population, raising concerns about the generalizability of research findings (Guédé, 2022).

Finally, ethical and privacy issues remain insufficiently addressed in several studies. The collection and analysis of personal digital data require greater attention to informed consent, data protection, transparency, and responsible research practices, particularly in studies involving social media platforms.

Overall, the reviewed literature demonstrates that while automated content analysis has significantly advanced corporate communication research, its effective application depends on addressing methodological, theoretical, and ethical limitations. Future research should therefore focus on developing more robust theoretical frameworks, improving computational language-



processing techniques, reducing algorithmic bias, and promoting hybrid analytical approaches that combine human expertise with artificial intelligence.

5. CONCLUSION

Expansion of Content Analysis Applications: The systematic review revealed a substantial expansion in the application of content analysis within corporate communication research in digital environments. Beyond the analysis of traditional communication messages, content analysis is increasingly employed for corporate reputation management, crisis monitoring, audience analysis, communication campaign evaluation, and the examination of corporate discourse and organizational culture through the analysis of leadership communications. These findings underscore the growing strategic role of content analysis in supporting corporate communication and organizational decision-making. (Poltoratska, 2025)

Growing Adoption of Automated Content Analysis and Artificial Intelligence: The reviewed studies indicate that automated content analysis has become the predominant approach for analyzing digital content by leveraging machine learning, lexicon-based methods, natural language processing (NLP), and Big Data Analytics. These technologies have significantly enhanced analytical efficiency by enabling researchers to process large volumes of textual, visual, and audiovisual data more rapidly and accurately than traditional manual approaches. (Richly, 2022)

Persistent Methodological and Technological Challenges: Despite the considerable advances in digital content analysis, the reviewed studies consistently identify several methodological and technological challenges. These include the complexity of natural language, difficulties in interpreting sarcasm and implicit meanings, issues related to data quality and representativeness, the limited accuracy of some computational models, and the risk of algorithmic bias. Ethical concerns surrounding data privacy and the responsible use of digital data also remain significant challenges requiring further attention. (Al-Shiakhli, 2019)

Need for Greater Methodological Integration: The findings demonstrate an increasing interest in mixed-methods approaches that combine qualitative and quantitative content analysis. However, this integration remains limited. The reviewed studies further recommend combining content analysis with complementary digital data sources, such as web-tracking and behavioral data, to provide more comprehensive and accurate insights into communication phenomena in digital environments. (Lischka, 2023)

6. Bibliography List :

1. Miązek. A .J. & Światowiec-Szczepańska. . (2023)The application of content analysis as a research method in management sciences. Eds: Wioletta Mierzejewska. Anna Michałkiewicz ‘The Application of Content Analysis as a Research Method (159–172). University of Łódź.

Content Analysis in the Era of Digital Data: Applications and Challenges

in Corporate Communication Research – A Systematic Review



2. A., Garmash, O., Marchenko, I., Garmatiuk, O., & Stefinin, V. Poltoratska. (2025). Exploring the Role of Digital Communication Tools in Shaping Organisational Culture and Enhancing Internal Communication. *Journal of Information Systems Engineering and Management* ,10 (2) ,367-374.
3. B. J.-Y. (2023). . In D.(Eds.), (pp.). Guédé. (2022). From Content Analysis to Content Analysis of Digital Social Networks: Specificities and Limits of a Particular Content. *Proceedings of the 2022 International Conference on International Studies in Social Sciences and Humanities (CISOC 2022)* (235–248). Atlantis Press.
4. Diana Amundsen. (2022). Using Digital Content Analysis for Online Research: Online News Media Depictions of Older Adults. Eds: Sage Research Methods: Doing Research Online. SAGE Publications Ltd.
5. Eriksson, J. &. (2013). Content Analysis in the Digital Age: Tools, Functions, and Implications for Security. Dans J. K. al., *The Secure Information Society* (pp. 137–148). Springer-Verlag London.
6. J.A. Lischka. (2023). Content Analysis in the Research Field of Corporate Communication. Eds: F., Kessler, S.H., Humprecht, E., Sommer, K., Castro, L. Oehmer-Pedrazzi ,*Standardisierte Inhaltsanalyse in der Kommunikationswissenschaft – Standardized Content Analysis in Communication Research* (349–361). Springer VS Wiesbaden.
7. L. Bardin. (2013). *L’analyse de contenu*. Presses Universitaires de France.
8. Renu Sabharwal and Shah Jahan Miah. (2021). A new theoretical understanding of big data analytics capabilities in organizations: a thematic analysis. *Journal of Big Data* ,8 (1) ,159.
9. Richly, M. A. (2022). Big Data Analytics Capabilities: A Systematic Literature Review on Necessary Skills to Succeed in Big Data Analytics. *Junior Management Science* , 7 (5), 1224-1241.
10. S., Ingle, T., Anand, A., Nagulwar, V., & Mankar, F. Patil. (2023). Digital Marketing Emerging Trends & Best Practices: A Comprehensive Analysis. *International Journal of Research Publication and Reviews* ,4 (10) ,587-590.
11. Sabrina Heike Kessler und Mike S. Schäfer. (2023). Content Analysis in the Research Field of Science Coverage. Eds: Sabrina Heike Kessler, Edda Humprecht, Katharina Sommer, Laia Castro Franziska Oehmer-Pedrazzi ,*Standardisierte Inhaltsanalyse in der Kommunikationswissenschaft – Standardized Content Analysis in Communication Research* (167-178). Springer VS Wiesbaden.
12. Sarah Al-Shiakhli. (2019). *Big Data Analytics : A Literature Review Perspective*. Luleå University of Technology .