

Integration of GRI standards in the sustainability reporting of Borussia Dortmund: a longitudinal RAG analysis

Sport, Business
and Management:
An International
Journal

Marcus Becker

*Department of Business and Management, Management Center Innsbruck,
Innsbruck, Austria*

Timo Zimmermann

*Business Administration, Sports and Events industry,
Westfälische Hochschule, Gelsenkirchen, Germany, and*

Judith Maria Beermann

*Department of Economics and Quantitative Methods,
International School of Management, Dortmund, Germany*

Received 27 March 2026

Revised 20 July 2026

7 August 2026

Accepted 7 August 2026

Abstract

Purpose – This study examines how the public sustainability reporting of Borussia Dortmund GmbH & Co. KGaA (BVB), a listed German professional football club, can be assessed through a Global Reporting Initiative (GRI)-oriented retrieval-augmented generation (RAG) workflow. The paper has a dual, case-specific purpose: to describe longitudinal disclosure patterns in BVB reports and to evaluate the methodological usefulness and boundaries of RAG-assisted report analysis.

Design/methodology/approach – The study uses a longitudinal single-case design based on eight public BVB sustainability reports covering the 2016/17 to 2023/24 reporting periods. Public PDF reports were converted into machine-readable text, segmented into overlapping chunks, embedded, retrieved through a Facebook AI Similarity Search (FAISS) vector store and analysed with standardised prompts aligned with the eight GRI reporting principles. Ratings were interpreted as RAG-derived evidence scores and manually cross-checked against retrieved report passages.

Findings – The RAG-derived scores indicate moderate and fluctuating evidence for GRI-oriented reporting quality rather than a monotonic improvement. Accuracy, balance, timeliness, and verifiability mostly receive medium-to-high scores, while completeness, sustainability context, and the clarity of later reporting remain more uneven. The results should not be read as legal compliance findings or assurance-type verification, but rather as a usable framework to semi-automate information extraction from sustainability reports of professional sport organisations.

Research limitations/implications – The study is limited to one club, public textual report content, and one RAG configuration. It demonstrates how RAG can structure disclosure review, but also shows that expert judgement remains necessary.

Practical implications – The paper provides a transparent workflow and concrete reporting recommendations for clubs preparing more traceable, comparable, and standard-oriented sustainability disclosures.

Social implications – Sustainability has become one of the most dominant issues shaping the strategic orientation of sport organizations in the twenty-first century. Professional football clubs, as globally visible and economically powerful entities, are increasingly expected to demonstrate environmental, social, and governance (ESG) responsibility. Their influence extends beyond sporting performance, reaching fans, communities, sponsors, and policymakers. Consequently, stakeholders now demand transparent reporting of sustainability commitments. Emerging regulatory governance further transforms sustainability from a voluntary communication practice into a key component of organizational legitimacy.

© Marcus Becker, Timo Zimmermann and Judith Maria Beermann. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) license. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this license may be seen at <http://creativecommons.org/licenses/by/4.0/>



Sport, Business and Management: An
International Journal
Emerald Publishing Limited
e-ISSN: 2042-6798
p-ISSN: 2042-678X
DOI 10.1108/SBM-03-2026-0100

Originality/value – The study connects sport sustainability reporting, GRI-based reporting principles and RAG-assisted document analysis in a longitudinal football case.

Keywords Sustainability reporting, Professional football, Borussia Dortmund, Global reporting initiative, GRI standards, Retrieval-augmented generation, Artificial intelligence, Sport management, ESG, Accountability

Paper type Research article

1. Introduction

Professional sport organisations increasingly operate in an environment in which financial performance, sporting success, and social accountability are observed simultaneously. This does not mean that every club faces the same sustainability obligations or stakeholder pressures. It does, however, mean that clubs with large audiences, commercial partners and capital-market visibility are increasingly asked to explain how they manage environmental, social and governance matters (Adam *et al.*, 2025; Yu and Jeong, 2025). Earlier sport management research has shown that corporate social responsibility (CSR) in professional team sport organisations is closely connected to stakeholder relationships, community expectations, and legitimacy concerns (Anagnostopoulos and Shilbury, 2013; Anagnostopoulos *et al.*, 2014; Walzel *et al.*, 2018). More recent work has broadened this discussion by examining sustainability management in the football industry, the environmental footprint of professional football events, and supporters' perceptions of CSR initiatives (Tavares *et al.*, 2023; Daddi *et al.*, 2026; L'Abate *et al.*, 2026).

The present study focuses on sustainability reporting rather than on sustainability performance itself. This distinction is important. Public sustainability reports are documents through which organisations select, structure and communicate sustainability-related information. They can improve transparency when they provide specific, balanced, and comparable information, but they can also leave information gaps when topics, boundaries, methods, or adverse developments are not disclosed in a sufficiently traceable way. In professional football, reporting is especially complex because clubs combine characteristics of listed or privately owned companies, cultural institutions, event organisers, employers, community actors, and media brands. These characteristics make football a relevant setting for analysing how sustainability communication is formalised over time.

Borussia Dortmund GmbH & Co. KGaA (BVB) provides an information-rich case for such an analysis. BVB is a prominent German professional football club and a capital-market-oriented organisation that has published sustainability or non-financial reports across several consecutive reporting periods. The available reporting sequence from 2016/17 to 2023/24, at the time of our investigation, enables a longitudinal view of how disclosure structures and evidence change over eight years. The case is not used to infer how all football clubs report. Rather, it is used to investigate how an AI-assisted, GRI-oriented workflow performs when applied to a comparatively data-rich football case.

The regulatory setting requires careful specification. The empirical analysis in this paper is anchored in the GRI Standards and their reporting principles, as the BVB's reports for the reporting sequence from 2016/17 to 2023/24 were prepared in accordance with these standards. The European Non-Financial Reporting Directive (NFRD), the German implementation of non-financial reporting requirements, the Corporate Sustainability Reporting Directive (CSRD), the European Sustainability Reporting Standards (ESRS), the ISSB Standards and the Voluntary Sustainability Reporting Standard for non-listed SMEs (VSME) are relevant contextual developments. They are not, however, treated as if they had governed all reporting years in the sample. The CSRD and ESRS form part of the current and future reporting environment, whereas the analysed BVB reports are evaluated against GRI-oriented quality principles. This narrower framing avoids treating a document analysis as a legal compliance assessment.

A second motivation is methodological. Sustainability reports are lengthy, heterogeneous, and often combine narrative sections, indicators, tables, and references to

external frameworks. Traditional manual content analysis can generate detailed interpretations, but it is difficult to scale and reproduce over multiple reports without a substantial coding protocol. Quantitative disclosure indices improve comparability but may abstract from contextual meaning. Retrieval-Augmented Generation (RAG) offers a middle position: it retrieves relevant text passages from the source documents before generating answers to pre-specified questions. This can make the use of large language models more traceable than unconstrained prompting, although it does not eliminate the need for human validation (Lewis *et al.*, 2020). Furthermore, RAG addresses one of the most significant limitations of large language models (LLMs) by reducing the risk of hallucinated or otherwise unverifiable responses (Lewis *et al.*, 2020). However, we can never completely rule out hallucinations, RAG forces LLMs to retrieve their knowledge from pre-selected sources of information.

Artificial intelligence (AI) has also become more visible in sports business and management research. Recent work published in Sport, Business and Management: An International Journal (SBM) has positioned AI as a research agenda for sport, while football-related studies have used interpretable machine learning for predictive analytics (Abeza, 2026; Li *et al.*, 2026). In adjacent sport management scholarship, AI-supported text analysis has been proposed for CSR-through-sport research and for methodological innovation in sport business studies (Anagnostopoulos *et al.*, 2025; Wang *et al.*, 2025). These studies indicate that AI is not confined to athlete performance analytics; however, governance-oriented and disclosure-oriented uses of AI remain less developed than commercial and performance applications. The present paper contributes to this emerging methodological conversation by applying RAG to public sustainability reports in a professional football case.

Our study therefore asks three focused questions (including subquestions):

- RQ1. How did the RAG-derived GRI-principle ratings for BVB's public sustainability reports develop between 2016/17 and 2023/24?
- RQ2a. Which disclosure features explain the strongest and weakest RAG-derived assessments?
- RQ2b. Which thematic emphases are associated with these assessments?
- RQ3a. What methodological contribution does the RAG workflow provide?
- RQ3b. What limitations in the application of RAG remain?

These questions are deliberately case-specific. The purpose is not to produce a sector-wide benchmark, but to examine what can and cannot be learned from an AI-assisted analysis of one longitudinal reporting corpus.

The paper makes three contributions. First, it offers a longitudinal analysis of a listed football club's sustainability reporting using the GRI reporting principles as an organising framework. Second, it operationalises the GRI quality principles in a transparent RAG protocol and distinguishes RAG-derived evidence scores from "assurance-type" evaluations of reporting quality. Third, it positions AI-assisted disclosure analysis within sport business and management by linking computational text analysis to accountability, stakeholder communication and reporting standardisation. The remainder of the paper reviews the regulatory and scholarly background, explains the RAG workflow, presents the results, and discusses implications for BVB, professional football clubs, and future sport management research.

2. Literature review and regulatory basis

2.1 Sustainability reporting frameworks and regulation

Sustainability reporting frameworks differ in legal status, audience, and conceptual basis. The GRI Standards are a global, stakeholder-oriented reporting framework that asks organisations

to report their impacts on the economy, environment and people. GRI 1: Foundation 2021 explains the purpose and system of the standards and specifies reporting principles for high-quality sustainability reporting. For this study, the relevant principles are accuracy, balance, clarity, comparability, completeness, sustainability context, timeliness, and verifiability. They are suitable for a disclosure analysis because they concern the way information is prepared and presented rather than only the underlying sustainability performance.

GRI is not identical to the European legal reporting framework. In Europe, the NFRD first introduced mandatory non-financial disclosure requirements for certain large public-interest entities ([European Union, 2022](#)). The CSRD expands sustainability reporting requirements and introduces the ESRS as mandatory standards for in-scope undertakings ([Commission Delegated Regulation, 2023/2772, 2023](#)). The first companies subject to the new CSRD rules had to apply them for the 2024 financial year, with reports published in 2025. This timing matters for the present study because BVB's reports from 2016/17 to 2023/24 were not produced under CSRD/ESRS obligations, as these were not mandatory in Germany at this time. The ESRS therefore provide a current regulatory reference point, but they are not the benchmark used to score the historical reports.

The relationship between GRI and ESRS is nevertheless important. European Financial Reporting Advisory Group (EFRAG), which provides technical advice to the European Commission and developed the draft ESRS, has worked with GRI on interoperability materials to reduce double reporting and to help companies understand commonalities between ESRS datapoints and GRI disclosures ([Global Reporting Initiative and European Financial Reporting Advisory Group, 2024](#)). This is relevant for football clubs that have already developed GRI-oriented reporting practices and may need to adapt to ESRS or related requirements in the future. Interoperability does not mean that a GRI report automatically satisfies all ESRS requirements, but it does imply that a GRI-based analysis can provide insight into reporting capabilities that may support future regulatory adaptation.

Other frameworks are also part of the reporting environment. The German Sustainability Code (DNK) has been important in the German context because it offers a structured and accessible reporting template. The ISSB Standards, especially IFRS S1 and IFRS S2, focus on investor-oriented sustainability-related financial disclosures and climate-related disclosures. The VSME standard aims to provide a proportionate voluntary framework for non-listed SMEs. These frameworks are discussed here only briefly to define the broader reporting and standard-setting landscape in which GRI-based sustainability reporting is situated.

For the period examined in this study (2016/17–2023/24), sustainability reporting by Borussia Dortmund evolved within a regulatory environment that combined mandatory non-financial disclosure requirements with voluntary reporting practices. BVB published its first standalone sustainability report in the 2016/17 financial year. The NFRD and its implementation into German law through the “CSR-Richtlinie-Umsetzungsgesetz” introduced mandatory non-financial reporting requirements for certain large public-interest entities from the 2017 reporting year onwards. These requirements represented the main regulatory framework applicable during most of the period under investigation. However, the reporting obligations were less extensive and less standardised than those introduced later under the CSRD and the ESRS.

The consistent application of the GRI Standards across all reports in the sample provides an important methodological advantage. Since BVB explicitly adopted GRI as its reporting framework throughout the entire observation period, GRI offers a stable and comparable basis for evaluating disclosure quality over time. For this reason, the present study uses the GRI reporting principles as its analytical framework. This approach ensures alignment between the assessment criteria and the framework that guided the preparation of the reports themselves, while avoiding the retrospective application of ESRS requirements that were not yet applicable during the period under investigation.

2.2 Sustainability, CSR and accountability in professional football

[Anagnostopoulos and Shilbury \(2013\)](#) show that CSR implementation in football cannot be reduced to a single theoretical explanation. It combines strategic, institutional, and relational elements. [Walzel et al. \(2018\)](#) similarly note that CSR research in professional team sport has often focused on community programmes, stakeholder relations and qualitative case evidence.

Recent studies suggest that sustainability in football has become more operational and data-oriented, but not necessarily uniform. [L'Abate et al. \(2026\)](#) analyse sustainable sports management in the football industry and show that sustainability has become part of broader management and governance concerns. [Daddi et al. \(2026\)](#) calculate the environmental footprint of a professional football match using life cycle assessment and thereby demonstrate that football sustainability issues can be measured in operational terms. [Tavares et al. \(2023\)](#) show that supporters perceive CSR activities of sports clubs and that such perceptions are connected to reputational and behavioural outcomes. These studies support the relevance of sustainability reporting while also indicating that disclosure quality cannot be assumed from the mere existence of sustainability initiatives.

Professional football also faces sector-specific materiality issues. Unlike ordinary service organizations, clubs are embedded in supporter communities and often function as symbolic carriers of identity, emotional attachment, and place-based belonging ([Heere and James, 2007](#)). As a result, material sustainability topics may include not only stadium operations, supporter mobility, merchandising, employee practices, governance structures, and commercial partnerships, but also youth development, diversity, anti-discrimination, safety, and community programmes. This specific combination of event-based operations and symbolic social relevance may not be fully reflected in general corporate reporting frameworks ([Babiak and Wolfe, 2009](#); [Smith and Westerbeek, 2007](#)). However, it also creates a risk that clubs emphasize positive sport-specific narratives and community engagement while providing less systematic information on environmental performance, reporting boundaries, or negative impacts, consistent with broader evidence on selective and symbolic sustainability disclosure ([Cho and Patten, 2007](#); [Boiral, 2013](#); [Michelon et al., 2015](#)).

BVB is a relevant case within this context because it combines a large supporter base, a major stadium, capital-market visibility, and a continuous public reporting sequence. The club is not presented as representative of all German or European clubs. Rather, its reporting record creates an opportunity to study a club for which longitudinal disclosure evidence is available. The case can therefore inform sport management research on how sustainability reporting becomes routinised and how AI-supported methods can be used to examine such routinisation.

2.3 AI-assisted disclosure analysis and RAG in sport management

[Abeza \(2026\)](#) identifies AI in sport as a topic that requires attention to applications, political economy and future research agendas. This agenda includes athlete performance, fan engagement, media production, governance, commercial decision-making, and organisational strategy. [Li et al. \(2026\)](#) use interpretable machine learning for soccer player valuation, illustrating how computational methods can support data-intensive football decisions. [Wang et al. \(2025\)](#) argue that AI can reshape sport business research and open new methodological possibilities. These studies provide a basis for positioning AI-assisted disclosure analysis as a further, governance-oriented application.

Prior work has applied Natural Language Processing (NLP) to identify climate-relevant sections in corporate reports ([Luccioni et al., 2020](#)), to automatically detect GRI topics and disclosures in sustainability reports ([Polignano et al., 2022](#)), and to assess the similarity or “affinity” of sustainability reports with the GRI framework using text-mining and semantic similarity methods ([Gutierrez-Bustamante and Espinosa-Leal, 2022](#)). More recent NLP applications also use BERT-based multi-label classification to match sustainability-report passages to GRI requirements ([Hillebrand et al., 2023](#)). RAG is particularly relevant where the

task requires answers grounded in source documents. In a RAG workflow, a user query first retrieves semantically related passages from a document corpus, and a language model then generates an answer conditioned on the retrieved context (Lewis *et al.*, 2020). The method is attractive for sustainability reports because it can make generated answers traceable to document passages. However, RAG does not solve all problems of content analysis. Retrieval can miss relevant passages, extracted PDF text may omit tables or figures, prompts can overstate what can be inferred, and generated ratings still depend on human-designed criteria.

This distinction is central to the present paper. A RAG system can support the review of public disclosures by identifying whether a report contains evidence related to a GRI principle. It cannot, on the basis of public PDF text alone, verify whether the disclosed data are materially correct, complete in an assurance sense, or legally compliant. Claims about accuracy, completeness, and verifiability must therefore be interpreted as evidence-based disclosure assessments, not as factual confirmation of BVB's internal systems or performance. This limitation responds directly to concerns in sustainability reporting research that disclosure quality does not necessarily reflect underlying sustainability performance or assurance quality, since sustainability reports may be used selectively or symbolically to manage legitimacy rather than to provide complete and verifiable accountability (Cho and Patten, 2007; Boiral, 2013; Michelon *et al.*, 2015).

The methodological contribution of this study is thus not that RAG replaces expert analysis. Rather, it is that RAG can help structure a reproducible first-pass review, keep the analysis close to source passages, and generate comparable evidence scores across reporting years. For sport management, this is useful because many clubs publish narrative reports that are difficult to compare manually. A transparent RAG protocol can support researchers, leagues and clubs in identifying where disclosures provide sufficient evidence and where expert follow-up is mandatory. This positioning is consistent with emerging management research that treats LLM-based methods as part of a qualitative analytical process rather than as autonomous substitutes for researcher interpretation. Garcia Quevedo *et al.* (2026), for example, show how LLM-based NLP and information retrieval can extend researchers' capacity to explore and select textual evidence while preserving human engagement with the data and leaving substantive interpretation to the researcher.

2.4 Conceptual lens: accountability, legitimacy, and reporting evidence

The study draws on stakeholder-oriented accountability and legitimacy reasoning without claiming that these theories exhaust the field. Sustainability reporting can be understood as a mechanism through which organisations account for their social and environmental impacts, responsibilities, and responses to stakeholders (Gray *et al.*, 1995; Deegan, 2002; O'Dwyer, 2005). In sport, this is especially relevant because clubs operate in dense stakeholder environments and depend on supporter trust, community acceptance, sponsor relationships, and regulatory approval (Breitbarth and Harris, 2008; Walters and Tacon, 2010; Babiak and Wolfe, 2009). Reports may therefore serve both accountability and legitimacy functions. They can provide information for stakeholder assessment, while also demonstrating that the organisation is responsive to social expectations and legitimacy pressures (Suchman, 1995; Cho and Patten, 2007; Michelon *et al.*, 2015).

A GRI-oriented RAG analysis offers a document-level lens on these functions. The analysis asks whether the public reports contain evidence that enables a reader to evaluate BVB's sustainability communication. This includes whether quantitative data are specific, whether challenges are acknowledged, whether information is comparable over time, and whether methods and sources are sufficiently traceable. The focus is deliberately on reporting evidence. It does not infer organisational motives, does not assess the effectiveness of sustainability programmes, and does not establish legal compliance.

This conceptual boundary strengthens the study. It connects the substantive case analysis with the methodological RAG analysis by treating public reports as the shared object of inquiry. BVB's reports are not only sources of sustainability information; they are also test cases for the limits of AI-assisted disclosure analysis. The study can therefore answer both what the RAG workflow identifies in the BVB reporting sequence and what this identification process reveals about the use of AI in sport management research.

3. Methodology

3.1 Research design and case selection

The study uses a longitudinal single-case design combined with AI-assisted document analysis. A single case is appropriate because the research questions concern an information-rich reporting sequence and the methodological performance of a RAG workflow in a concrete organisational context. The design follows a qualitative-computational logic. The RAG system supports retrieval, summarisation and scoring, while the interpretation remains grounded in a human-defined framework and manual plausibility checks. This division of analytical roles follows the broader "text-as-data" principle that computational text analysis should be embedded in an explicit social-science research design (Grimmer *et al.*, 2022), and with emerging LLM-assisted qualitative research in which computational methods support rather than replace researcher interpretation (Garcia Quevedo *et al.*, 2026; Than *et al.*, 2025).

BVB was selected for four reasons. First, it is one of the few professional football clubs with a continuous sequence of public sustainability reports over the examined period. Second, its legal form and capital-market visibility make transparency and governance a relevant context for the case. Third, during the period under consideration, the club explicitly refers to established reporting frameworks, particularly GRI, which permits a standards-oriented analysis. Fourth, the case is substantively meaningful for sport management because it combines professional sport, public attention, social engagement, and reporting formalisation.

The case-selection logic is purposive rather than probabilistic (Creswell, 2009). Although the study does not use BVB to inductively build theory, it follows the case-study principle that cases can be chosen for their theoretical and informational usefulness rather than random representativeness (Eisenhardt, 1989). This is consistent with sport-management case research, where purposeful selection is appropriate when a bounded organisational setting offers particularly rich evidence for the research question (Morse and McEvoy, 2014). Here, the case is bounded by one organisation and the reporting periods 2016/17–2023/24. Within this boundary, all eight public sustainability reports constituting the analysed reporting sequence were included; no report years within the defined sequence were selectively omitted. The case-selection and corpus-construction logic therefore support an in-depth longitudinal analysis of reporting development and of the RAG workflow applicability. It does not claim to provide statistical generalisation to professional football clubs as a population.

3.2 Data corpus and preprocessing

The empirical corpus consists of eight public BVB sustainability reports covering the reporting periods 2016/17 to 2023/24. The documents were collected as PDF files from public corporate reporting sources. The analysis used textual report content. Images, videos, and design elements were not independently analysed. Tables were included only to the extent that their text was extracted by the PDF processing tool. This restriction is important because sustainability reports often contain relevant information in figures, tables, or layout structures that may not be fully captured by text extraction [1].

The RAG pipeline converted the PDFs into machine-readable text using PyPDFLoader. In this context, machine-readable text means that the textual content of each PDF page was extracted into strings that can be processed computationally. Non-breaking spaces and line breaks were normalised, repeated whitespace was removed, and each text unit retained

metadata including the source file, reporting period, page number, and chunk identifier. No private, confidential, or internal BVB documents were used. The analysis therefore reflects the information available to external readers of the public reports.

The cleaned page text was segmented into chunks of 500 characters with an overlap of 50 characters. Chunking is standard in RAG workflows because language models and vector retrieval systems generally operate more effectively on smaller passages than on entire long documents, as shorter retrieval units can improve the specificity and traceability of retrieved evidence (Pinecone, 2025; Unstract, 2025). The 50-character overlap, corresponding to 10% of the chunk size, was used to reduce the risk that relevant information located at a chunk boundary is split in a way that impairs retrieval. At the same time, the chunking design introduces limitations. Very short chunks may fragment context, while longer chunks may reduce retrieval precision; accordingly, there is no universally optimal chunk size, as appropriate settings depend on the document structure, retrieval model, and downstream task (Li et al., 2024; Bhat et al., 2025). The selected parameters therefore represent a transparent and reproducible design choice that balances granularity, traceability, and computational efficiency for the analysis of sustainability reports.

For each reporting year, the pipeline created a separate vector store. This design prevented evidence from one year being retrieved when the question concerned another year. It also enabled longitudinal comparison because identical GRI-oriented prompts were applied to each reporting period. The model’s outputs were exported to an Excel workbook containing detailed answers and a numerical ranking sheet. The present article uses the ranking sheet as the basis for the aggregated longitudinal results and interprets the scores together with the methodological limitations of the pipeline.

3.3 RAG architecture and prompt design

The RAG workflow followed five sequential steps (see Figure 1). First, the reports were pre-processed and converted into machine-readable text. Second, the cleaned text was segmented into overlapping chunks and embedded into a vector space. Third, standardised GRI-oriented questions were submitted separately for each reporting year. Fourth, the retriever identified the most relevant text passages for each question, and the language model generated an answer constrained by these passages. Fifth, the generated answers, retrieved passages, and ratings were exported to create an audit trail for manual plausibility checks and longitudinal comparison.

Methodologically, we treat this pipeline as an LLM-assisted qualitative-computational form of document analysis rather than only as a technical retrieval system. Emerging research on LLM-assisted content analysis and qualitative coding treats LLMs as coding or annotation aids whose outputs depend on explicit instructions, transparent model and prompt choices, and human validation (Carlson and Burbano, 2026; Farjam et al., 2026; Than et al., 2025). In the present study, the GRI principles provide the deductive analytical framework; retrieval identifies criterion-relevant passages; the first prompt produces an evidence-focused

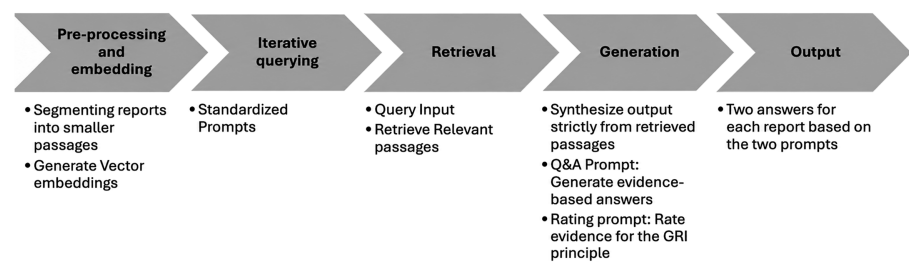


Figure 1. RAG workflow for GRI-oriented sustainability report analysis. Source: Created by the authors

qualitative summary; the second prompt maps that summary to a predefined ordinal score; and the research team reviews the resulting evidence and ratings for traceability and plausibility. Thus, the LLM performs a structured “first-pass” analysis, while criterion definition and final interpretation remain researcher-controlled.

The implemented pipeline used LangChain components for document loading, splitting, retrieval, and chain construction. The language model was Meta’s open-source model Llama 3.2:3 b accessed through Ollama. Embeddings were generated with OpenAI’s text-embedding-3-small model. The vector store was FAISS, and the retriever used similarity search with the top 10 retrieved chunks for each query. The RAG system therefore had two stages: retrieval of the ten most similar text passages for a GRI-oriented question and generation of an answer constrained by those passages.

The Q&A prompt instructed the model to answer each numbered question using only the retrieved context, to include page numbers if relevant, to include direct quotes and numerical data when available, to avoid general summaries, and to answer “I don’t know” when the answer was not present in the context. The template for the Q&A plot was as follows:

Answer the numbered question using only the given context. Do not rate the answer.

If relevant, include the page number. Use direct quotes if available and include numbers and concrete data.

Avoid general summaries.

If the answer is not in the context, reply: “I don’t know.”

Context: {context}

Question: {question}

A second prompt then asked the model to assign a score from 1 to 5 and provide a short justification. The rating scale defined 1 as no or minimal evidence, 2 as weak evidence, 3 as moderate evidence, 4 as largely fulfilled, and 5 as fully fulfilled. The second rating prompt was given by:

You are given an answer to a question.

Evaluate how well the answer fulfills the criterion of the question.

Question: {question}

Answer: {answer}

Assign a rating on a 5-point Likert scale and justify your decision.

Likert scale definitions:

1 = Not fulfilled (no or minimal evidence, major omissions)

2 = Weak (some mention but many requirements missing)

3 = Moderate (several requirements addressed but important gaps remain)

4 = Largely fulfilled (mostly consistent and complete, only minor issues)

5 = Fully fulfilled (comprehensive, consistent, verifiable, fully aligned with GRI principle)

Output format:

Rating: [1–5]

Justification: [short explanation]

The exported total rating equals the sum of the eight numeric ratings divided by the maximum possible score of 40, multiplied by 100.

The separation between the Q&A prompt and the rating prompt was a deliberate design choice. The first prompt performed an evidence-extraction function: it asked the model to identify and summarise only information contained in the retrieved passages. The second prompt performed an evaluative function: it assessed the generated answer against the respective GRI principle. This procedural separation reduced the risk that the model would assign ratings before the relevant evidence had been made explicit, and it made the basis for each score more transparent for subsequent human review.

Using identical prompts across the eight reporting years was essential for longitudinal comparability because differences in scores should reflect differences in retrieved report evidence rather than changes in the questions posed to the model. The prompts were revised conceptually in this manuscript to avoid implying that the model can determine factual correctness in an assurance-like sense. For example, the accuracy criterion is interpreted as the presence of specific, internally coherent, and sufficiently detailed disclosed information, not as proof that the underlying data are correct. Similarly, completeness is interpreted as the breadth of public report coverage relative to disclosed material topics and boundaries, not as proof that no undisclosed material issue exists.

3.4 Operationalisation of the GRI reporting principles

[Table 1](#) summarises how the eight GRI reporting principles were translated into RAG-readable criteria. The operationalisation deliberately distinguishes what can be assessed from public report text and what remains outside the scope of the method.

3.5 Scoring, validation, and interpretation boundaries

The output scores are labelled RAG-derived evidence scores. This label is used to avoid conflating three different objects: the quality of BVB's underlying sustainability management, the quality of BVB's public reporting, and the quality of the RAG-generated output. The score indicates how much evidence the RAG system retrieved and generated for a particular reporting principle in a particular report. It is therefore a structured disclosure indicator, not a legal judgement, audit opinion, or full reporting-quality index.

Manual validation was conducted in two ways. First, the research team reviewed the RAG prompts, the extracted ratings, and the aggregated ranking sheet for consistency with the operationalisation in [Table 1](#). Second, selected high and low ratings were cross-checked against the retrieved passages and the original reports where needed. This process was used to identify overly general model justifications, missing context, and cases where the model might overstate the implications of textual evidence. The validation procedure should therefore be understood as a plausibility and traceability check rather than as external assurance or full intercoder validation. The review examined whether model-generated answers were specific, relevant, clear, and sufficiently grounded in the retrieved passages, and whether the assigned ratings were consistent with the operational definitions in [Table 1](#). Where outputs were overly general, insufficiently supported, or potentially misleading, they were interpreted conservatively in the results section.

This validation strategy deliberately stops short of the stronger benchmarking designs proposed in the emerging LLM-methods literature. [Carlson and Burbano \(2026\)](#) show that prompt design can materially affect annotations and downstream analyses and therefore stress documentation, sensitivity analysis, and multi-stage validation. [Farjam et al. \(2026\)](#) likewise validate automated coding against human-coded material, while [Than et al. \(2025\)](#) recommend a "researcher-LLM-researcher" workflow in which the model performs an initial pass and humans validate or refine the output. Our procedure follows the latter division of labour through manual plausibility and source-grounding checks, but it does not implement a multi-model, multi-prompt sensitivity analysis or a full independent human-coded

Table 1. Translation of GRI reporting principles into RAG-readable evidence criteria and interpretation boundaries

GRI principle	RAG-readable evidence criterion	Examples of evidence sought	Boundary of interpretation
Accuracy	Specificity and internal coherence of disclosed qualitative and quantitative information	Units, dates, measurement descriptions, numerical indicators, consistency between related statements	Does not verify the correctness of underlying measurement systems or internal calculations
Balance	Presence of both favourable information and disclosed challenges, limitations, or adverse developments	References to risks, trade-offs, incidents, missed targets, and critical stakeholder concerns	Does not establish whether all negative impacts were disclosed
Clarity	Accessibility and understandability of the public report text	Definitions, structure, non-technical language, tables, or summaries captured as text	Does not evaluate visual design beyond text extraction
Comparability	Evidence that readers can compare information across periods or against frameworks	Prior-year figures, recurring KPIs, stable metrics, explanations of methodological changes	Does not validate comparability of underlying data collection methods unless disclosed
Completeness	Breadth of public coverage of material topics, boundaries and stakeholder-relevant areas	Materiality descriptions, topic boundaries, environmental, social, and governance coverage	Does not prove that all material impacts were included
Sustainability context	Connection between club disclosures and broader environmental, social, or sector contexts	SDGs, climate targets, league or regional benchmarks, sector-specific challenges	Does not assess substantive contribution to sustainable development
Timeliness	Evidence that information relates to a defined and recent reporting period	Reporting period, publication cycle, recent updates, explanation of delays	Does not evaluate whether internal data collection was timely
Verifiability	Traceability of disclosed information to sources, methods, or external references	GRI index, assurance references, methodologies, page-level evidence, public sources	Does not substitute for external assurance or audit procedures

Source(s): Edited by the authors based on GRI 1: Foundation 2021 and the implemented RAG protocol

benchmark. Accordingly, the validation supports the traceability and cautious interpretation of the RAG-derived results, but does not establish equivalence with independent human coding or robustness across alternative models and prompt configurations.

The results are reported conservatively. A low score for completeness, for example, is phrased as limited retrieved evidence for public report completeness relative to the operationalised criterion. It is not phrased as proof that BVB omitted legally required information. Likewise, a high score for verifiability means that the report text provided more traceable evidence or references for the model to retrieve, not that the information was independently audited in full. This wording is essential because several GRI principles are difficult to evaluate on the basis of public documents alone.

Potential hallucinations were addressed through three safeguards. First, the Q&A prompt restricted the model to retrieved report passages and required an “I don’t know” response when the relevant information was not available in the context. Second, generated answers were checked against the retrieved passages and, where necessary, against the original reports to identify unsupported statements, overly broad inferences, or inconsistencies between the justification and the cited evidence. Third, the results were interpreted conservatively: outputs that were insufficiently grounded, overly general, or potentially misleading were not treated as

independent findings but as indicators requiring expert judgement. These safeguards reduce, but cannot eliminate, the risk of hallucinated or unsupported outputs.

3.6 Research ethics and AI-use disclosure

The study used only publicly available corporate reports and did not involve human participants, personal data collection, or confidential organisational data. The RAG pipeline itself is the object and instrument of the study. The code used for preprocessing, retrieval, generation, scoring, and Excel export is documented in the accompanying “RAG_final.py file”.

The empirical RAG system described in this section is part of the research object and analytical workflow. A separate declaration of generative AI and AI-assisted technologies has been added at the end of our article to disclose any AI-assisted manuscript-preparation support, distinguish it from our empirical RAG analysis, and clarify full author responsibility for our final submission manuscript.

4. Results

4.1 Descriptive score pattern

Table 2 and Figure 2 present the RAG-derived GRI-principle scores for the eight BVB reporting periods. The aggregate scores range from 60% in 2023/24 to 72.5% in 2017/18. The pattern does not support a simple narrative of continuous improvement. Instead, it suggests a fluctuating development in which some principles receive stable evidence across years while others remain uneven. This is consistent with the interpretation that sustainability reporting becomes more complex as standards, stakeholder expectations, and report structures evolve.

The strongest and most stable dimensions are accuracy, balance, timeliness, and verifiability, which mostly receive scores of 3 or 4. Although the present scoring scheme is tailored to GRI-principle assessment rather than to generic question-answering, the use of intermediate scores is consistent with RAG evaluation research, which commonly assesses retrieved-context relevance, answer relevance, and groundedness on graded or continuous scales rather than relying exclusively on binary classifications (Es et al., 2024; Saad-Faloon et al., 2024; Yang et al., 2024). Accordingly, scores of 3 or 4 should not be interpreted as factual verification of BVB’s sustainability performance, but as evidence-based assessments of how strongly the retrieved report passages support each GRI principle according to our operationalisation mapping in Table 1.

Comparability remains moderate, with a high score in 2019/20 but otherwise concentrates around 3. Completeness is the weakest dimension, with scores of 2 in most years and a temporary increase to 4 in 2022/23. Sustainability context is also moderate and drops to 2 in 2022/23. Clarity is high in earlier years but falls in 2022/23 and 2023/24, suggesting that the retrieved evidence became less accessible or less straightforward for the model to summarise under the clarity criterion.

4.2 Evidence by reporting principle

Accuracy receives scores of 3 or 4 across the period. In the operationalisation used here (see Table 1), this indicates that the reports often contain quantitative indicators, dates, units, or internally coherent explanations that can be retrieved by the model. The lower scores in 2019/20 and 2021/22 should be interpreted cautiously. They do not imply that the data are inaccurate. They indicate that, under the prompt and retrieval settings, the extracted evidence was less comprehensive or less specific than in other years.

Balance varies between 3 and 4. The model retrieved both positive statements and some references to challenges or limitations, but the evidence does not justify a broad claim that the reports are fully balanced. In football reporting, this distinction matters because clubs naturally have incentives to communicate achievements and community value

Table 2. RAG-derived GRI-principle ratings for BVB sustainability reports, 2016/17–2023/24

Year	Accuracy	Balance	Clarity	Comparability	Completeness	Sustainability context	Timeliness	Verifiability	Total (%)
2016/17	4	3	4	3	2	3	4	4	67.5
2017/18	4	4	4	3	3	3	4	4	72.5
2018/19	4	3	4	2	2	3	4	4	65
2019/20	3	4	4	4	2	3	4	4	70
2020/21	4	3	4	3	2	3	4	3	65
2021/22	3	4	4	3	2	3	2	4	62.5
2022/23	4	3	3	3	4	2	4	4	67.5
2023/24	4	4	2	3	2	3	3	3	60

Note(s): Ratings range from 1 (minimal retrieved evidence) to 5 (comprehensive retrieved evidence). Total rating is the sum of eight scores divided by 40

Source(s): Edited by the authors based on the RAG output workbook and Rankings sheet

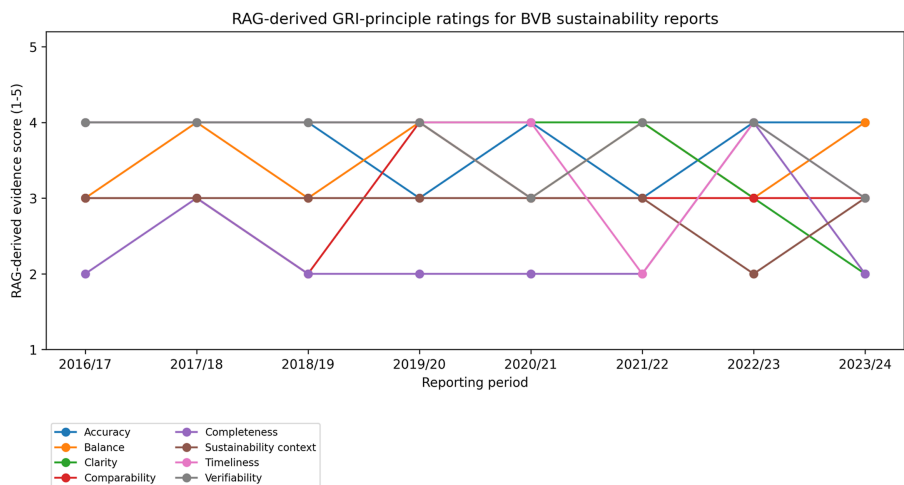


Figure 2. RAG-derived GRI-principle ratings across reporting periods. Source: Edited by the authors based on RAG evaluation outputs

(Breitbarth and Harris, 2008). A balanced report should also provide enough information about trade-offs, risks, negative impacts, or targets not achieved. The RAG evidence suggests that BVB’s public reports contain some such elements, but the extent differs across years.

Clarity is consistently strong from 2016/17 to 2021/22 but declines in the two most recent reports, reaching 2 in 2023/24. This result may reflect denser reporting structures, more technical language, or more fragmented retrieved passages. It may also reflect the limitations of text extraction from PDFs if tables and layout cues become more important in later reports. Therefore, the decline should not be read as a direct deterioration of the report as experienced by human readers. It indicates that the RAG system retrieved less clear textual evidence for the clarity criterion.

Comparability is moderate across the period. The score of 4 in 2019/20 suggests stronger retrieved evidence for recurring indicators, prior-year comparisons, or methodological continuity in that year. In other years, scores of 2 or 3 indicate that the retrieved report passages did not consistently provide a longitudinal structure across all relevant topics. This is important for stakeholders because comparability enables readers to distinguish one-off initiatives from sustained improvement. For clubs, a stable set of core indicators across years can support both internal management and external accountability.

Completeness is the most critical dimension in the RAG-derived results. The repeated score of 2 does not mean that BVB’s reports are legally incomplete. It means that the model often retrieved limited evidence that all material topics, boundaries, and stakeholder-relevant areas were covered comprehensively. The temporary score of 4 in 2022/23 suggests that the report for that year provided more retrievable evidence of breadth. The pattern points to a practical reporting issue: football clubs can strengthen public reporting by making materiality processes, topic boundaries and omissions more explicit.

The sustainability context is mostly moderate. The reports include references to broader frameworks such as sustainability goals or social themes, but the retrieved evidence often provides limited connection to sector-specific benchmarks, regional environmental challenges or scientifically grounded thresholds. This matters because GRI’s sustainability context principle asks organisations to place their impacts in broader systems. For football clubs, relevant context may include stadium-related emissions, fan mobility, supply-chain impacts, community conditions and league-level sustainability strategies.

Timeliness is strong in most years but falls to 2 in 2021/22 and 3 in 2023/24. The generally higher scores indicate that the reports typically identify reporting periods and recent activities. Lower scores should again be interpreted as limitations in retrieved evidence rather than as proof of delayed publication or outdated internal data. Future reports could strengthen timeliness by consistently distinguishing reporting-period data, post-balance-sheet developments, and multi-year targets.

Verifiability receives scores of 3 or 4. In the operationalisation used here (Table 1), verifiability refers to the traceability of public report information through references, indices, page-level evidence, methodologies, or external assurance statements. It does not mean that the RAG system verified the reported data. The results suggest that BVB's reports provide several traceable elements, but that methods, data sources, and assurance boundaries could be made more explicit to support external review.

4.3 Methodological performance of the RAG workflow

The RAG workflow added value in three ways. First, it standardised the questions asked of each reporting year, which makes longitudinal comparison more transparent than an unstructured reading of each report. Second, it linked generated answers to retrieved passages, which helped the research team inspect whether justifications were grounded in the source documents. Third, it produced numerical scores that can be evaluated in follow-up studies.

At the same time, the workflow showed important limitations. The model sometimes produced general justifications even when the prompt asked for concrete evidence. It can also confuse the presence of a page reference, number, or framework label with stronger evidence of quality. The PDF extraction process may underrepresent non-textual material such as charts, icons, tables, and visual hierarchies. The prompt-based rating approach is sensitive to wording, retrieval parameters, and model configuration. These limitations mean that the scores are best understood as screening indicators rather than final judgements.

The most important methodological lesson is that the object of assessment must be clearly named. When the model scores 'accuracy' or 'completeness', it is not certifying the report. It is evaluating whether the retrieved public text contains evidence that is relevant to the operationalised criterion in Table 1. This distinction makes RAG useful for sport management research because it supports reproducibility and source-grounded analysis, while preserving the role of expert interpretation.

4.4 Case-specific interpretation for BVB

The BVB case shows that a football club with continuous public sustainability reporting can still display uneven evidence across reporting principles. The RAG-derived pattern suggests that BVB's reports often provide accessible information on current activities, social engagement, employees, governance references, and reporting periods. The weaker dimensions concern breadth, contextualisation, and the continuity of comparable indicators across all material areas. This is not unusual for organisations moving from broad narrative, i.e. descriptive, sustainability communication toward more formalised sustainability reporting, as prior empirical research shows that standard adoption and reporting experience may increase the amount and structure of disclosed information without necessarily resolving weaknesses in transparency, comparability or disclosure quality (Vigneau *et al.*, 2015; van Oorschot *et al.*, 2024).

The findings also suggest that later reports may have become more complex. Increasing regulatory and stakeholder expectations can lead to longer, more technical, and more framework-oriented reporting. This may support compliance preparation but can reduce clarity if readers cannot easily identify what changed, which indicators are comparable, and how the club defines the boundaries of its impacts. BVB and comparable clubs could therefore benefit from combining narrative communication with a stable data appendix, a clear GRI/ESRS mapping table, and explicit explanations of methodological changes.

The results should not be framed as criticism of BVB's sustainability performance. The study only analyses public reports and the RAG system's evidence retrieval. Nonetheless, the findings provide constructive guidance for reporting practice: make material topics and boundaries explicit, report both achievements and challenges, maintain recurring KPIs, distinguish measured from estimated data, provide external assurance boundaries where applicable, and explain how football-specific impacts such as fan mobility, stadium operations and community engagement are contextualised.

5. Discussion

RQ1 asked how the RAG-derived GRI-principle ratings for BVB's public sustainability reports developed between 2016/17 and 2023/24. The answer is that the ratings fluctuate rather than improve linearly. The highest aggregate score appears in 2017/18, and the lowest appears in 2023/24. Several principles remain comparatively stable, while completeness and sustainability context are more uneven. This finding cautions against assuming that the mere accumulation of reporting experience automatically produces higher evidence scores across all GRI principles. Prior research similarly shows that more extensive or more formalized sustainability reporting does not necessarily translate into higher disclosure quality, completeness or credibility, as reports may become longer, more standardized or more symbolic without providing proportionately more transparent information ([Michelon et al., 2015](#); [Boiral, 2013](#); [van Oorschot et al., 2024](#)).

RQ2a-b asked which disclosure features and thematic emphases explain the strongest and weakest RAG-derived assessments. Stronger assessments are associated with report passages that contain specific data, reporting-period references, traceable sources, and clear descriptions of initiatives. Weaker assessments are associated with limited evidence of topic boundaries, incomplete longitudinal KPIs, insufficient contextualisation against sector or sustainability benchmarks, and less explicit discussion of limitations or omissions. In substantive terms, the BVB reports appear to provide more retrievable evidence for social and organisational themes than for fully contextualised environmental or boundary-related disclosures. This interpretation is consistent with sport CSR research showing that CSR in professional sport and football has often been framed around stakeholder management, community engagement, and the creation of social, cultural, and reassurance value ([Breitbarth and Harris, 2008](#); [Walters and Tacon, 2010](#); [Babiak and Wolfe, 2009](#)). However, we phrase this interpretation cautiously because our study does not benchmark BVB against other clubs.

RQ3a-b asked what methodological contributions and limitations the RAG workflow provides for AI-assisted analysis of public sustainability disclosure in professional football. The contribution lies in making document review more systematic, transparent, and reproducible. The same prompts can be applied across reporting years, and retrieved evidence can be inspected. The limitations are equally important. RAG does not verify internal data, cannot reliably interpret all non-textual content, depends on chunking and retrieval quality, and still requires expert judgement in the final interpretation. The study therefore positions RAG as a support tool for disclosure analysis rather than a substitute for assurance, legal review or qualitative expertise.

The present study extends current SBM conversations from performance analytics ([Li et al., 2026](#)) and broad AI agendas ([Abeza, 2026](#)) to public disclosure analysis. It also connects football sustainability research with reporting standards, showing that the future of sport sustainability research may require both substantive domain knowledge and transparent computational workflows. In this respect, the study also points toward agentic AI workflows, where specialised AI agents could sequentially retrieve relevant report passages, evaluate them against predefined reporting criteria, flag ambiguities or missing evidence, and produce traceable outputs for human validation rather than replacing expert judgement.

For reporting practice, the findings have several implications. First, clubs should distinguish communication-oriented narratives from reporting evidence. Stakeholders can

value stories and examples, but standard-oriented reporting also requires stable indicators, methods, boundaries, and links to material topics. Second, clubs preparing for ESRS or similar frameworks can use GRI-based analyses as a diagnostic starting point, especially because GRI and ESRS show substantial overlap in impact-related reporting. Third, AI-assisted tools can be used internally as a pre-review mechanism to identify missing evidence, unclear terminology, or weak comparability before a report is published.

For BVB specifically, the analysis, based on the reports through the exemplary season 2023/24, suggests six practical improvements. The club could provide a concise materiality table that links each material topic to boundaries, KPIs, targets, and responsible functions. It could maintain a time-series appendix for core football-specific indicators, including stadium energy, emissions, waste, fan mobility, employee diversity, youth development and community programmes. It could distinguish measured, estimated, and narrative information. It could explain changes in measurement methods. It could include a clearer mapping between GRI disclosures and future ESRS-relevant data points where appropriate. Finally, it could state the scope and limits of external assurance or internal verification in a way that allows readers to assess traceability without overinterpreting the report.

The regulatory environment remains dynamic. The EU sustainability reporting framework has expanded through the CSRD and ESRS, while the European Commission's 2025 Omnibus simplification package illustrates that the scope, timing and administrative burden of sustainability reporting obligations remain subject to political recalibration ([European Commission, 2025](#)). The Commission's Omnibus initiative explicitly aims to simplify sustainability reporting and due diligence requirements, including changes to CSRD scope and future ESRS simplification. This uncertainty strengthens rather than weakens the methodological relevance of our study. Even if reporting obligations are delayed, simplified or narrowed, clubs and other sport organisations will continue to face stakeholder requests for credible sustainability information. A RAG workflow can help organisations examine whether their public reports contain the evidence needed for such requests, provided the outputs are reviewed by experts and interpreted within the correct legal context.

The study's limitations should be explicitly stated. The paper analyses one case and cannot make general claims about all football clubs. It uses public PDF reports and therefore cannot assess undisclosed information, internal controls, or the correctness of source data. It focuses on textual extraction and may miss visual or tabular content. Future research could extend the present text-based RAG workflow into a multimodal RAG system as described in [Chen et al. \(2022\)](#). Such a system would treat sustainability reports not only as textual documents but also as visually structured artefacts that contain relevant information in tables, figures, charts, photographs, icons, and page layouts. One approach would be to extract each report page as an image, detect visual elements such as tables and figures, generate textual descriptions or structured representations of these elements using vision-language models, and index both the original text and the visual summaries in the retrieval database. Recent work on multimodal RAG and visual document retrieval provides first examples of this direction, showing how retrieval can be extended beyond text to include images, tables, charts, and document layout ([Chen et al., 2022](#); [Faysse et al., 2025](#); [Gao et al., 2023](#)). Related document-understanding models such as LayoutLM, LayoutLMv3, and Donut further show that layout and visual information can be incorporated into document analysis rather than relying solely on OCR-extracted text ([Xu et al., 2020](#); [Huang et al., 2022](#); [Kim et al., 2022](#)). Applied to sustainability reporting, such an extension could help assess whether environmental indicators, reporting boundaries, or negative impacts are disclosed in visual elements that are not fully captured by text extraction.

Another shortcoming of our research results in the mono-use of only one embedding model, one retrieval configuration, and one prompt family. The validation process checks plausibility but does not provide a fully independent human-coded benchmark.

Despite these limitations, the study demonstrates a feasible route for AI-assisted sport disclosure research. The key is methodological modesty. RAG can improve transparency in

how researchers read documents, but it should not be presented as a machine capable of making definitive sustainability judgements. Its value lies in disciplined retrieval, standardised questioning and a clearer audit trail for scholarly interpretation.

6. Conclusion

This study analyses BVB's sustainability reporting as a case-specific, GRI-oriented, and RAG-assisted disclosure study. The results show that the public reports provide moderate evidence for several GRI reporting principles, especially accuracy, balance, timeliness, and verifiability as operationalised in the RAG protocol. At the same time, completeness, sustainability context, and later-report clarity remain areas where retrieved evidence is more uneven. These findings are not legal or assurance conclusions; they are structured indicators of what the RAG system could retrieve and justify from public report text.

The broader contribution is methodological. A RAG workflow can support longitudinal analysis of sport sustainability reports by making the questions, evidence retrieval, and scoring logic more transparent. It can help researchers and practitioners identify where reports contain specific evidence and where follow-up review is required. The method is most valuable when combined with clear operational definitions and cautious interpretation. For professional football, where reporting is becoming more formalised but remains shaped by sport-specific stakeholder relations, such tools can support more comparable, traceable and accountable sustainability communication.

Beyond the present GRI-oriented application, the proposed RAG workflow provides a flexible foundation for future research. The methodology can be adapted to assess sustainability reporting against other established reporting frameworks and emerging regulatory standards, enabling consistent and transparent comparative analyses across different reporting regimes. In this way, the results of this study and the developed RAG model can serve as a starting point for broader AI-assisted analyses of sustainability disclosures in sport management and beyond.

Declaration of generative AI and AI-assisted technologies

During preparation of the revised manuscript, ChatGPT (OpenAI, GPT-5.5 Thinking, accessed 17 June 2026) was used as an editorial support tool to assist with language refinement, structural consistency and the drafting of possible response-letter wording. The tool was not credited as an author and was not used as an independent source of data, empirical evidence, citations, theoretical claims, or conclusions. The authors reviewed, edited, and verified all AI-assisted wording, checked the cited literature and empirical interpretations, and take full responsibility for the accuracy, integrity, and final content of the manuscript. The empirical RAG system described in the methods section is part of the research object and analytical workflow; its configuration, role, and limitations are reported in [Sections 3.2-3.6](#).

Note

1. A shortcoming that could be resolved by applying multimodal RAG systems (MURAGs) as introduced by [Chen et al. \(2022\)](#). We discuss this shortcoming and further enhancements in the limitation section.

References

- Abeza, G. (2026), "Artificial intelligence in sport: applications, political economy and research agendas", *Sport, Business and Management: An International Journal*, ahead-of-print, pp. 1-15, doi: [10.1108/SBM-03-2026-0092](https://doi.org/10.1108/SBM-03-2026-0092).
- Adam, S., Müller-Thümen, F., Kaiser, S. and Hovemann, G. (2025), "Kicking towards sustainability: the role of institutional pressures in German professional football", *Managing Sport and Leisure*, pp. 1-17, doi: [10.1080/23750472.2025.2459728](https://doi.org/10.1080/23750472.2025.2459728).

-
- Anagnostopoulos, C. and Shilbury, D. (2013), "Implementing corporate social responsibility in English football: towards multi-theoretical integration", *Sport, Business and Management: An International Journal*, Vol. 3 No. 4, pp. 268-284, doi: [10.1108/sbm-05-2013-0009](https://doi.org/10.1108/sbm-05-2013-0009).
- Anagnostopoulos, C., Byers, T. and Shilbury, D. (2014), "Corporate social responsibility in professional team sport organisations: towards a theory of decision-making", *European Sport Management Quarterly*, Vol. 14 No. 3, pp. 259-281, doi: [10.1080/16184742.2014.897736](https://doi.org/10.1080/16184742.2014.897736).
- Anagnostopoulos, C., Rahman, S.A., Rizvi, S.W.F. and Mamo, Y. (2025), "Leveraging artificial intelligence for data collection and analysis: the case of and application in CSR through sport research", *International Journal of Sports Marketing and Sponsorship*, Vol. 26 No. 4, pp. 803-822, doi: [10.1108/ijsms-09-2024-0261](https://doi.org/10.1108/ijsms-09-2024-0261).
- Babiak, K. and Wolfe, R. (2009), "Determinants of corporate social responsibility in professional sport: internal and external factors", *Journal of Sport Management*, Vol. 23 No. 6, pp. 717-742, doi: [10.1123/jsm.23.6.717](https://doi.org/10.1123/jsm.23.6.717).
- Bhat, S.R., Rudat, M., Spiekermann, J. and Flores-Herr, N. (2025), "Rethinking chunk size for long-document retrieval: a multi-dataset analysis", arXiv preprint, arXiv:2505.21700.
- Boiral, O. (2013), "Sustainability reports as simulacra? A counter-account of A and A+ GRI reports", *Accounting, Auditing and Accountability Journal*, Vol. 26 No. 7, pp. 1036-1071, doi: [10.1108/AAAJ-04-2012-00998](https://doi.org/10.1108/AAAJ-04-2012-00998).
- Breitbart, T. and Harris, P. (2008), "The role of corporate social responsibility in the football business: towards the development of a conceptual model", *European Sport Management Quarterly*, Vol. 8 No. 2, pp. 179-206, doi: [10.1080/16184740802024484](https://doi.org/10.1080/16184740802024484).
- Carlson, N.A. and Burbano, V. (2026), "The use of LLMs to annotate data in management research: foundational guidelines and warnings", *Strategic Management Journal*, Vol. 47 No. 3, pp. 699-725, doi: [10.1002/smj.70023](https://doi.org/10.1002/smj.70023).
- Chen, W., Hu, H., Chen, X., Verga, P. and Cohen, W.W. (2022), "MuRAG: multimodal retrieval-augmented generator for open question answering over images and text", *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 5558-5570.
- Cho, C.H. and Patten, D.M. (2007), "The role of environmental disclosures as tools of legitimacy: a research note", *Accounting, Organizations and Society*, Vol. 32 Nos 7/8, pp. 639-647, doi: [10.1016/j.aos.2006.09.009](https://doi.org/10.1016/j.aos.2006.09.009).
- Commission Delegated Regulation (EU) 2023/2772 (2023), "Commission delegated regulation (EU) 2023/2772 of 31 July 2023 supplementing Directive 2013/34/EU of the European Parliament and of the Council as regards sustainability reporting standards".
- Creswell, J.W. (2009), *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 3rd ed., Sage, Thousand Oaks, CA.
- Daddi, T., Marrucci, L., Todaro, N.M. and Iraldo, F. (2026), "How polluting is professional football? The environmental footprint of a football match calculated with a life cycle assessment", *Sport, Business and Management: An International Journal*, Vol. 16 No. 3, pp. 339-361, doi: [10.1108/SBM-11-2024-0167](https://doi.org/10.1108/SBM-11-2024-0167).
- Deegan, C. (2002), "Introduction: the legitimising effect of social and environmental disclosures – a theoretical foundation", *Accounting, Auditing and Accountability Journal*, Vol. 15 No. 3, pp. 282-311.
- Eisenhardt, K.M. (1989), "Building theories from case study research", *Academy of Management Review*, Vol. 14 No. 4, pp. 532-550, doi: [10.2307/258557](https://doi.org/10.2307/258557).
- Es, S., James, J., Espinosa-Anke, L. and Schockaert, S. (2024), "RAGAs: automated evaluation of retrieval augmented generation", *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pp. 150-158.
- European Commission (2025), *Proposal for a Directive of the European Parliament and of the Council amending Directives (EU) 2022/2464 and (EU) 2024/1760 as regards the dates from which Member States are to apply certain corporate sustainability reporting and due diligence requirements*, COM(2025) 80 final, 26 February 2025.
-

- European Union (2022), "Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 as regards corporate sustainability reporting", *Official Journal of the European Union*.
- Farjam, M., Meyer, H. and Lohkamp, M. (2026), "A practical guide and case study on how to instruct LLMs for automated coding during content analysis", *Social Science Computer Review*, Vol. 44 No. 3, pp. 488-502, doi: [10.1177/08944393251349541](https://doi.org/10.1177/08944393251349541).
- Faysse, M., Sibille, H., Wu, T., Omrani, B., Viaud, G., Hudelot, C. and Colombo, P. (2025), "ColPali: efficient document retrieval with vision language models", *International Conference on Learning Representations*.
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J. and Wang, H. (2023), "Retrieval-augmented generation for large language models: a survey", arXiv:2312.10997.
- Garcia Quevedo, D., Glaser, A. and Verzat, C. (2026), "Enhancing theorization using artificial intelligence: leveraging large language models for qualitative analysis of online data", *Organizational Research Methods*, Vol. 29 No. 1, pp. 92-112, doi: [10.1177/10944281251339144](https://doi.org/10.1177/10944281251339144).
- Global Reporting Initiative and European Financial Reporting Advisory Group (2024), "GRI-ESRS interoperability index", available at: <https://www.globalreporting.org/standards> (accessed 17 June 2026).
- Gray, R., Kouhy, R. and Lavers, S. (1995), "Corporate social and environmental reporting: a review of the literature and a longitudinal study of UK disclosure", *Accounting, Auditing and Accountability Journal*, Vol. 8 No. 2, pp. 47-77.
- Grimmer, J., Roberts, M.E. and Stewart, B.M. (2022), *Text as Data: A New Framework for Machine Learning and the Social Sciences*, Princeton University Press.
- Gutierrez-Bustamante, M. and Espinosa-Leal, L. (2022), "Natural language processing methods for scoring sustainability reports – a study of Nordic listed companies", *Sustainability*, Vol. 14 No. 15, p. 9165, doi: [10.3390/su14159165](https://doi.org/10.3390/su14159165).
- Heere, B. and James, J.D. (2007), "Sports teams and their communities: examining the influence of external group identities on team identity", *Journal of Sport Management*, Vol. 21 No. 3, pp. 319-337, doi: [10.1123/jsm.21.3.319](https://doi.org/10.1123/jsm.21.3.319).
- Hillebrand, L., Pielka, M., Leonhard, D., Deußer, T., Dilmaghani, T., Kliem, B., . . . and Sifa, R. (2023), "Sustain. AI: a recommender system to analyze sustainability reports", *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law*, pp. 412-416.
- Huang, Y., Lv, T., Cui, L., Lu, Y. and Wei, F. (2022), "LayoutLMv3: pre-training for document AI with unified text and image masking", *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 4083-4091.
- Kim, G., Hong, T., Yim, M., Nam, J., Park, J., Yim, J., Hwang, W., Yun, S., Han, D. and Park, S. (2022), "OCR-free document understanding transformer", *European Conference on Computer Vision*, pp. 498-517.
- L'Abate, V., Vitolla, F. and Raimo, N. (2026), "Scoring goals or missing the target? Exploring sustainability in the football industry", *Business Strategy and Development*, Vol. 9 No. 1, e70278, doi: [10.1002/bsd.70278](https://doi.org/10.1002/bsd.70278).
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Kuttler, H., Lewis, M., Yih, W., Rocktaschel, T., Riedel, S. and Kiela, D. (2020), "Retrieval-augmented generation for knowledge-intensive NLP tasks", *Advances in Neural Information Processing Systems*, Vol. 33, pp. 9459-9474.
- Li, Y., Ayanso, A., Yuan, S., Kusy, M. and Kerwin, S. (2026), "When interpretable machine learning meets the beautiful game: a predictive analytics approach to soccer player valuation in the transfer market", *Sport, Business and Management: An International Journal*, Vol. 16 No. 2, pp. 240-261, doi: [10.1108/SBM-05-2025-0108](https://doi.org/10.1108/SBM-05-2025-0108).
- Li, Z., Li, C., Zhang, M., Mei, Q. and Bendersky, M. (2024), "Retrieval augmented generation or long-context LLMs? A comprehensive study and hybrid approach", *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*.

- Luccioni, A., Baylor, E. and Duchene, N. (2020), "Analyzing sustainability reports using natural language processing", arXiv preprint, arXiv:2011.08073.
- Michelon, G., Pilonato, S. and Ricceri, F. (2015), "CSR reporting practices and the quality of disclosure: an empirical analysis", *Critical Perspectives on Accounting*, Vol. 33, pp. 59-78, doi: [10.1016/j.cpa.2014.10.003](https://doi.org/10.1016/j.cpa.2014.10.003).
- Morse, A.L. and McEvoy, C.D. (2014), "Qualitative research in sport management: case study as a methodological approach", *Qualitative Report*, Vol. 19, pp. 1-13, doi: [10.46743/2160-3715/2014.1032](https://doi.org/10.46743/2160-3715/2014.1032).
- O'Dwyer, B. (2005), "Stakeholder democracy: challenges and contributions from social accounting", *Business Ethics: A European Review*, Vol. 14 No. 1, pp. 28-41, doi: [10.1111/j.1467-8608.2005.00384.x](https://doi.org/10.1111/j.1467-8608.2005.00384.x).
- Pinecone (2025), "Chunking strategies for LLM applications", available at: <https://www.pinecone.io/learn/chunking-strategies/> (accessed 3 July 2026).
- Polignano, M., Basile, P., de Gemmis, M. and Semeraro, G. (2022), "An NLP approach for the analysis of global reporting initiative indexes from corporate sustainability reports", *Proceedings of the First Computing Social Responsibility Workshop within the 13th Language Resources and Evaluation Conference*, pp. 1-10.
- Saad-Falcon, J., Khatlab, O., Potts, C. and Zaharia, M. (2024), "ARES: an automated evaluation framework for retrieval-augmented generation systems", *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 338-354.
- Smith, A.C.T. and Westerbeek, H.M. (2007), "Sport as a vehicle for deploying corporate social responsibility", *The Journal of Corporate Citizenship*, Vol. 25, pp. 43-54, doi: [10.9774/GLEAF.4700.2007.sp.00007](https://doi.org/10.9774/GLEAF.4700.2007.sp.00007).
- Suchman, M.C. (1995), "Managing legitimacy: strategic and institutional approaches", *Academy of Management Review*, Vol. 20 No. 3, pp. 571-610, doi: [10.2307/258788](https://doi.org/10.2307/258788).
- Tavares, J., Borges, A.P., Vieira, B.M., Vieira, E. and Rodrigues, P. (2023), "Corporate social responsibility of sports clubs: what do supporters think?", *Sport, Business and Management: An International Journal*, Vol. 13 No. 6, pp. 752-768, doi: [10.1108/SBM-03-2023-0030](https://doi.org/10.1108/SBM-03-2023-0030).
- Than, N., Fan, L., Law, T., Nelson, L.K. and McCall, L. (2025), "Updating 'the future of coding': qualitative coding with generative large language models", *Sociological Methods and Research*, Vol. 54 No. 3, pp. 849-888, doi: [10.1177/00491241251339188](https://doi.org/10.1177/00491241251339188).
- Unstrat (2025), "Chunk size and overlap", available at: https://docs.unstrat.com/unstrat/unstrat_platform/user_guides/chunking/ (accessed 3 July 2026).
- van Oorschot, K.E., Johansen, V.A., Thorup, N.L. and Aspen, D.M. (2024), "Standardization cycles in sustainability reporting within the global reporting initiative", *European Management Journal*, Vol. 42 No. 4, pp. 492-502, doi: [10.1016/j.emj.2024.04.001](https://doi.org/10.1016/j.emj.2024.04.001).
- Vigneau, L., Humphreys, M. and Moon, J. (2015), "How do firms comply with international sustainability standards? Processes and consequences of adopting the Global Reporting Initiative", *Journal of Business Ethics*, Vol. 131 No. 2, pp. 469-486, doi: [10.1007/s10551-014-2278-5](https://doi.org/10.1007/s10551-014-2278-5).
- Walters, G. and Tacon, R. (2010), "Corporate social responsibility in sport: stakeholder management in the UK football industry", *Journal of Management and Organization*, Vol. 16 No. 4, pp. 566-586, doi: [10.5172/jmo.2010.16.4.566](https://doi.org/10.5172/jmo.2010.16.4.566).
- Walzel, S., Robertson, J. and Anagnostopoulos, C. (2018), "Corporate social responsibility in professional team sports organizations: an integrative review", *Journal of Sport Management*, Vol. 32 No. 6, pp. 511-530, doi: [10.1123/jsm.2017-0227](https://doi.org/10.1123/jsm.2017-0227).
- Wang, J.J., Mao, L.L. and Mastromartino, B. (2025), "Guest editorial: sport business research in the era of artificial intelligence", *International Journal of Sports Marketing and Sponsorship*, Vol. 26 No. 4, pp. 661-669, doi: [10.1108/ijsms-10-2025-376](https://doi.org/10.1108/ijsms-10-2025-376).

- Xu, Y., Li, M., Cui, L., Huang, S., Wei, F. and Zhou, M. (2020), "LayoutLM: pre-training of text and layout for document image understanding", *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1192-1200.
- Yang, X., Sun, K., Xin, H., Sun, Y., Bhalla, N., Chen, X., Choudhary, S., Gui, R., Jiang, Z.W., Jiang, Z., Kong, L., Moran, B., Wang, J., Xu, Y.E., Yan, A., Yang, C., Yuan, E., Zha, H., Tang, N., Chen, L., Scheffer, N., Liu, Y., Shah, N., Wanga, R., Kumar, A., Yih, W. and Dong, X.L. (2024), "CRAG – comprehensive RAG benchmark", *Advances in Neural Information Processing Systems, Datasets and Benchmarks Track*.
- Yu, J.G. and Jeong, Y. (2025), "Enhancing environmental, social, and governance practices in professional sports clubs: a Delphi study approach", *Sage Open*, Vol. 15 No. 4, doi: [10.1177/21582440251395587](https://doi.org/10.1177/21582440251395587).

Further reading

- Global Reporting Initiative (2021), *GRI 1: Foundation 2021*, Global Reporting Initiative, Amsterdam.
- ISSB (2023a), *IFRS S1 General Requirements for Disclosure of Sustainability-Related Financial Information*, IFRS Foundation, London.
- ISSB (2023b), *IFRS S2 Climate-Related Disclosures*, IFRS Foundation, London.

Corresponding author

Marcus Becker can be contacted at: Marcus.Becker@mci.edu