University of Basel
Center for Philanthropy Studies

ceps

CEPS Working Paper Series
No. 23, August 2026

How do Nonprofits Report About Impact? Applying Retrieval-Augmented Generation (RAG) for Content Analysis of Annual Reports

Georg von Schnurbein
Center for Philanthropy Studies (CEPS), University of Basel
georg.vonschnurbein@unibas.ch

Omar Hussein
Center for Philanthropy Studies (CEPS), University of Basel
omar.hussein@unibas.ch

***** Working Paper – not peer reviewed *****

Citation:
von Schnurbein, G. & Hussein, O. (2026). How do Nonprofits Report About Impact? Applying Retrieval-Augmented Generation (RAG) for Content Analysis of Annual Reports. CEPS Working Paper Series, No. 23, Basel: CEPS.

ISSN: 2296-7516 (Print)
2296-7524 (Online)

Abstract

Impact reporting is a growing expectation for nonprofit organizations, yet existing research faces methodological limitations in analyzing large volumes of qualitative data across organizations. This study introduces an innovative application of Retrieval-Augmented Generation (RAG) to conduct a large-scale content analysis of impact-related information in 476 annual reports from Swiss nonprofits. RAG integrates large language models (LLMs) with real-time document retrieval, offering a hybrid approach that combines the depth of qualitative interpretation with the scale of automated analysis. To address the constraints of context length and document retrieval capacity, we present an Ensemble RAG architecture—assigning a dedicated RAG model to each organization and consolidating responses across a uniform set of 18 prompts.

Our findings reveal a substantial gap between implicit and explicit communication of impact. While few organizations explicitly use frameworks like Theory of Change or logic models, a majority convey core components (e.g., inputs, outcomes, impacts) without naming them. This underreporting, especially at the outcome and impact levels, reflects a disconnect between strategic intent and reporting practice. The study contributes a replicable method for applying LLMs to nonprofit texts and offers new insights into how impact is framed in sector-wide communication. Methodologically, it demonstrates that ensemble RAG can overcome the limitations of conventional qualitative and quantitative approaches, paving the way for more scalable and nuanced analysis of nonprofit documents. For practitioners, the results underscore the need for clearer articulation of impact and institutionalized responsibility for impact reporting in nonprofit governance.

Keywords: impact reporting, nonprofit management, retrieval augmented generation (RAG), nonprofit acountability

Disclosure Statement: The author(s) declare that there are no conflicts of interest that relate to the research, authorship, or publication of this article

Content

1 Introduction 4
2 Literature review 4
2.1 The use of annual reports 4
2.2 Impact reporting 5
2.3 Overcoming limits of qualitative and quantitative content analysis 5
3 Methodology 6
3.1 Dataset 6
3.2 Large Language Models and Retrieval Augmented Generation 6
3.3 Ensemble Retrieval Augmented Generation and Experimental Setup 7
4 Findings 9
5 Discussion 10
5.1 Application of the RAG model 11
6 Conclusion 12
References 12
Annex 1: Prompt example 16

1 Introduction

Annual reports are among the most important documents for nonprofit organizations. They serve multiple purposes: documenting organizational history, attracting and retaining funders, communicating successes, and demonstrating accountability (Cordery & McConville, 2022). Typically, these reports combine standardized financial statements with a more flexible section on activities and output or outcome. When well-crafted, annual reports are powerful tools to showcase impact, build trust with stakeholders, and highlight mission-driven achievements. Through impact reporting – using data, stories, and key performance indicators – nonprofits can illustrate how their programs contribute to social change (Gordon et al. 2016).

However, effective reporting impact presents challenges (McConville, 2017). Many nonprofits struggle to measure long-term outcomes, collect reliable data, and directly link changes to their efforts, especially in complex social or environmental contexts. Limited resources and staff capacity often hinder comprehensive analysis. Balancing transparent reporting with compelling storytelling is also difficult. Although impact has become a core goal for nonprofits (von Schnurbein, 2016), reporting it remains complicated due to its qualitative nature. Despite new models for evaluation and reporting, practical implementation has seen little change (Benjamin et al., 2023).

Previous research on annual reports typically relies on either large-scale, regulator-provided quantitative data or small-scale qualitative analysis, as the latter demands extensive time to read and classify content (Cordery & McConville, 2022). This study proposes a new approach using a Retrieval-Augmented Generation Model (RAG model) to analyze large volumes of annual reports more efficiently. We argue that impact reporting is an ideal use case for this method, offering both time savings and practical insights. Our research explores two questions:

(1) How do nonprofits report on impact in their annual reports?
(2) What are the strengths and limitations of using a RAG model to analyze these documents?

We begin by reviewing literature on the use of annual reports for impact reporting, then explain the RAG model and its operationalization. Afterwards, we present findings and discuss their implications for research and practice.

2 Literature review

2.1 The use of annual reports

Nonprofits face diverse accountability demands due to their reliance on public and private support. Annual reports—combining financial and performance information—play a key role in addressing these. They offer transparency and aid decision-making through both functional and symbolic roles (Flack, 2007). From a resource dependency perspective, they help secure continued support (Parsons, 2003). Signaling theory also suggests they demonstrate responsible stewardship and accountability to key stakeholders. Nevertheless, some critics point out significant shortcomings in annual reporting, such as the use of promotional language (Patrut et al., 2011), lack of transparency (Skaerbaek, 2005), and the dissemination of potentially misleading information (Krishnan et al., 2006). These issues are often explained through the lens of positive stakeholder theory, which posits that organizations tailor their reporting to meet anticipated expectations of donors and other stakeholders (Cordery & McConville, 2022). As a result, nonprofits may fall into a "starvation cycle," where excessive spending on overhead and fundraising efforts compromises their ability to sustain core operations (Lecy & Searing, 2015).

2.2 Impact reporting

Impact reporting has become a crucial practice for nonprofits, enhancing transparency, accountability, and organizational identity. By clearly articulating the outcomes of their programs, nonprofits show how donations and public support contribute to meaningful change. This builds trust with funders, beneficiaries, and the public, reinforcing credibility. In a competitive funding environment, strong impact reporting helps organizations distinguish themselves by highlighting their effectiveness and results, strengthening their case for continued support.

Despite its growing importance, impact reporting faces several challenges (von Schnurbein, 2016). Outcomes in the nonprofit sector are often multidimensional, long-term, and difficult to quantify—particularly in areas involving social or behavioral change (Maier et al., 2024). The absence of standardized frameworks leads to inconsistencies across organizations, complicating comparison and interpretation. Many nonprofits also lack the resources or capacity to collect reliable data, especially in diverse or underserved communities. These issues can limit the depth and credibility of reported outcomes. From a social constructivist perspective, impact is shaped by the stakeholder's viewpoint (Benjamin et al., 2022; Herman & Renz, 2008). Donors, beneficiaries, and board members may interpret the same outcomes differently (Mauer, 2014), posing a challenge for creating a single report that satisfies all groups. Another lens, institutional isomorphism (DiMaggio & Powell, 1983), suggests that impact reporting is shaped by external pressures—legal requirements or third-party certifications (coercive isomorphism), or by mimicking peer organizations in the face of uncertainty (mimetic isomorphism) (Hersberger et al., 2020; AbouAssi & Bies, 2017).

To support impact reporting, key frameworks have been developed. The Theory of Change (ToC), popularized in the 1990s by philanthropic organizations like the Aspen Institute, outlines how and why change occurs, mapping assumptions, inputs, activities, and outcomes (Anderson, 2004). The Logical Framework (LogFrame), created in the 1970s for USAID, links project resources, objectives, and indicators in a structured matrix (Prinsen & Nijhof, 2015). The broader concept of social impact draws on development economics and systems thinking to connect specific initiatives to wider societal change (Benjamin et al., 2023). These frameworks are widely used by NGOs, governments, and development agencies to guide planning, evaluation, and communication (von Schnurbein, 2016).

2.3 Overcoming limits of qualitative and quantitative content analysis

Cordery and McConville (2022) argue that annual reports are a rich yet underutilized source of data for understanding nonprofit organizations, especially registered charities. They use the term "content analysis" broadly to include both qualitative and quantitative approaches. Qualitative content analysis can reveal important themes and narratives but is often labor-intensive, limited in scale, and subject to researcher bias (Harwood & Garry, 2003). These issues can affect the validity and generalizability of findings, especially for smaller research teams. Quantitative content analysis, while more scalable and systematic, can oversimplify complex information. By converting rich text into numerical codes, it may miss deeper meanings and reduce contextual nuance (Polizzi, 2022). Standardized coding frameworks can also limit flexibility, making it harder to capture unique organizational expressions.

To address these limitations, emerging methods using large language models (LLMs) and natural language processing (NLP) offer a promising middle ground (Cordery & McConville, 2022). These tools combine the scale and consistency of quantitative methods with the interpretive depth of qualitative analysis. LLMs can quickly process large volumes of text, identify themes, and even detect sentiment with human-like understanding (Chew et al., 2023). They are already being applied to nonprofit texts like mission statements

(Ma, 2021; Meier & von Schnurbein, 2024) and legal documents (Litofcenko et al., 2020). Unlike traditional methods, LLMs are more flexible, reducing the need for rigid coding while maintaining contextual sensitivity. As a result, they enhance both the insight and scalability of content analysis in nonprofit research.

3 Methodology

This section introduces the technical methodologies used in this paper. As stated before, the primary aim of the paper is a content analysis on impact reporting in nonprofit annual reports. The second aim is the introduction of a RAG-model (Lewis et al. 2020) to nonprofit research.

3.1 Dataset

The dataset used in our study and experiment contains 476 annual reports from Swiss nonprofits. The format of these reports is Portable Document Format (PDF) and most of the reports are written in either German (80.9%) or French (15.0%) languages. Further descriptive information on the sample of organizations and their annual reports are presented in table 1. The sample size excesses previous qualitative content analysis sample by far (Cordery and McConville 2022). However, for a LLM and a RAG-model the sample size is moderate, but for the purpose of this paper sufficient. The nonprofits in the sample show a high level of homogeneity in reporting standards, as all of the nonprofits included in the study are or have been certified at least once before by Zewo foundation. Zewo foundation is an independent Swiss foundation that audits charities and other nonprofits in Switzerland (von Schnurbein & Fritz 2017). NPOs that pass the Zewo audit indicate to donors that their donation is well-managed.

Table 1: Descriptive data on the sample (n=476)

Category				
field of activity			N	in %
	1: Domestic health		80	17.6 %
	2: International humanitarian affairs		119	26.2 %
	3: Domestic social affairs		209	46.0 %
	4: Environment		22	4.8 %
	5: Domestic care homes		46	10.1 %
Size in revenues (CHF)	min 0.0	max 360,741,000	average 9,939,682	std.dv. 27,136,069.42
Size of annual reports in pages	min 2	max 68	average 19.5	std.dv. 11.2
Languages ratios in annual reports	German 389 (81.72 %)	French 72 (15.13 %)	Italian 10 (2.1 %)	English 5 (1.05 %)

3.2 Large Language Models and Retrieval Augmented Generation

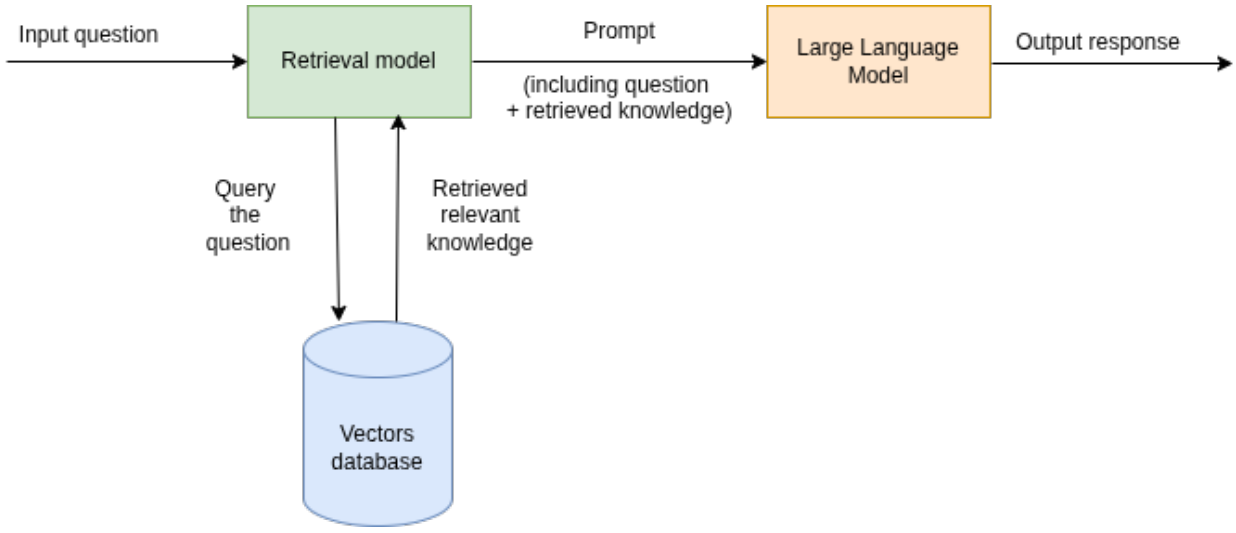
Large Language Models (LLMs) are advanced AI systems based on neural networks, often using transformer architectures (Vaswani et al., 2017). They are used for language tasks like writing, translating, and coding, and can handle multiple languages. Their effectiveness depends heavily on prompt quality –

well-crafted prompts improve accuracy and reduce errors like hallucinations. However, LLMs have limitations, such as fixed context windows and outdated knowledge due to training cutoffs.

When analyzing 476 annual reports, an LLM can be a helpful support tool, because a manual review would be too time-consuming. However, nonprofit annual reports are a very specific type of document and a conventional LLM might lack effective training on these documents. To enhance reliability, we thus use a RAG model. These combine LLMs with external document retrieval, grounding responses in real-time data (Arslan et al., 2024). RAG models improve factual accuracy, reduce hallucinations, and provide transparency by referencing source texts. They can also stay current as the knowledge base updates, making them especially useful for tasks requiring up-to-date and verifiable information – such as yearly annual reports.

The RAG architecture consists of three components: a vector database, a retrieval model, and a LLM. The vector database stores documents in vector form. When a user submits a query, the retrieval model identifies and selects the most relevant documents. These documents are then used by the LLM to generate a response. In essence, the LLM integrates these retrieved documents as an external knowledge source to inform its answer. This enables the system to provide accurate, contextual responses even on topics beyond the original training data.

Figure 1: RAG-Model



3.3 Ensemble Retrieval Augmented Generation and Experimental Setup

The goal in this study is to use a RAG model to answer questions about every organization in the dataset. There are two main challenges in using a single RAG model to cover the full dataset at once. The first challenge is the retrieval model as it can only retrieve a certain number of relevant texts, and it does not have the ability to retrieve relevant text from every organization's annual report at once. The second challenge is the LLMs context length, the context length may be smaller than the length of retrieved texts from the 476 organizations in the dataset. In order to get plausible answers, the text retrieved should not be

too long (e.g. one page in the annual report), but if it is too short (e.g. one sentence), the information might be incomplete.

Accordingly, we propose the idea of implementing an Ensemble RAG model as a solution for the existing challenges. A single vector database was created for each nonprofit, containing only the data from the annual report of a single organization. Then, a RAG model was created for each organization. Therefore, each question is asked once per organization through a prompt, and the retriever model retrieves from the vector database that contains only the organization's data. After implementing the previous steps for all the organizations and the prompts, the results are ensembled together to create a full added answer for the full sample for each prompt.

For the experiment, the programming codes were written in python and langchain python library was used for generating retrievals (Harrison, H. 2022). The LLM used is GPT-4o (OpenAI 2024) which was the highest performing LLM according to the global benchmarks at the time of the study. Several other LLMs were considered at the beginning of the study, the other models considered were Mistral-7B-Instruct-v0.3 (Mistral AI 2024), Mixtral-8x22B-Instruct-v0.1 (Mistral AI 2024), Llama-3.1-8B-Instruct (Meta 2024), and Llama-3.3-70B-Instruct (Meta 2024). On reviewing their results, they did not outstand GPT-4o, their results were inconsistent accompanied by occasional hallucinations and misunderstanding to the prompt. A total number of 18 prompts were asked in the Ensemble RAG model, a prompt for each question, each prompt was including clear descriptions of the concepts based on literature in order to lower the LLM's bias as much as possible (Benjamin et al. 2023, Helmig et al. 2014 ; Herman & Renz 2008; Hyndman & McConville 2016; McConville 2017 ; Meier 2023 ; Prinsen & Nijhof 2015; von Schnurbein 2016). Table 2 lists the core questions of the prompts, and an example of a full prompt is provided in the annex.

In order to assess the performance of GPT-4o, a small random validation dataset (24 reports ≈ 5%) was sampled from the original dataset and reviewed based on human for every prompt. The random sample resembled the full sample in terms of language (87,5% German) and average length (23.4%). The RAG-model accuracy was calculated as follows:

$$Accuracy(\%) = \frac{Numberofcorrectpredictions}{Totalnumberofpredictions} x100$$

Accuracy scores per each question were calculated separately along with the overall accuracy of the model as shown in table 2.

Table 2: Accuracy scores

Question	Accuracy (%)
Does the organization speak of an impact report?	100.0 %
Are the organizations' goals stated?	54.17 %
Do the organizations talk about a theory of change?	79.17 %
Do the organizations talk about a logic framework?	95.83 %
Do the organizations talk about social impact?	70.83 %
Do the organizations refer to the SDGs?	50.0 %

Do the organizations report on output?	87.5 %
Do the organizations report on outcome?	58.33 %
Do the organizations report on impact?	83.33 %
Do the organizations report on results?	66.67 %
Do the organizations report on concrete impact goals?	83.33 %
Is output reported quantitatively or qualitatively?	83.33 %
Is reporting done at the project level or organizational level?	70.83 %
Are actions derived from the impact findings?	95.83 %
Is someone in the organization responsible for impact?	70.83 %
Is impact measurement conducted externally?	95.83 %
Is impact measurement conducted for external funders?	100.0 %
Do the organizations report on failure?	91.67 %
Overall Accuracy	**79.86 %**

The model's overall accuracy seems plausible, since the questions asked are not always easy to find or clear in the reports. Some other studies in different research domains showed similar accuracy results for their RAG-models (Xiong, G. et al 2024; Ke, Y. et al 2024).

4 Findings

The following section presents the findings on impact reporting based on the 18 questions included in the RAG-model. First, we present general findings on the application of general concepts of impact and performance in nonprofits. Second, we will analyze more in-depth the content of the reports. The section concludes with findings on organizational issues related to impact reporting.

There is no regulation on the name of an annual report in Switzerland. Hence, the most common designation is the "annual report" (66.3%). However, nonprofits may use the title to point to their favored direction of reporting. Because of increased efficiency perspective, other designations such as "progress report" (17.3%) or "activity report" (8.6%) have emerged. Only few organizations call their annual report a "impact report" (0.3%).

Every nonprofit follows a charitable mission. An annual report is a perfect place to present the organization's primary goals. In our sample, only 43.7% of the organizations make use of this opportunity. Again, only few (2.3%) share more precise impact goals of the organization. On the contrary, many nonprofits (72.3%) refer to the sustainable development goals (SDGs) as a reference. This finding is in line with previous research connecting nonprofit missions with the SDGs (Meier 2023).

Reporting on impact should be based on a systematical approach that facilitates the understanding of inputs, outputs, and outcomes or impacts. Hence, the application and presentation of the organization's activities following a clear concept such as social impact, ToC or logic framework allows us to better understand the organization's mission. Social impact as the broadest concept is more common (25%) than the two other concepts. Only few report their theory of change (2.9%) and no organization uses the logic framework in its

annual report. These findings do not accord with the general trend on impact orientation in nonprofits (Benjamin et al. 2023). Hence, we used the RAG model to test whether the information for the components of a theory of change (input, output, outcome, impact) can be found in the annual reports. The result of this prompt was completely different: 92.9% of the organizations implicitly share their theory of change in the annual report. This way of communicating leaves it to the readers to make the connections between the different components and to understand the theory of change on their own. A more explicit communication about the theory of change (or a different concept) would ensure that readers get informed by the organization itself what its own understanding of impact is.

Without reference to a given concept, more than half of the nonprofits (56.3%) in the sample report about impact. The majority (96.3%) of NPOs use qualitative and quantitative information to report on their activities and performance at output level. The remaining 3.1% report only quantitatively on output. More specifically, 83.8% report on output measures such as number of people served or amount of service hours provided. 18,1% report concrete outcome measures such as satisfaction of clients or increase of wellbeing of clients. Finally, only 2.9% report on impact results on a societal scale. These more detailed findings align with the findings on the general use of impact concepts.

Impact measurement and reporting do pay off if the results lead to learning and improvement (Benjamin 2013, von Schnurbein 2016). As nonprofits are dependent on external funders that they do not want to lose, reports about learning and failure are not popular among nonprofits (Helmig et al. 2014). This does also account for the Swiss nonprofits analyzed in this survey. Only 1.5% report about consequences initiated out of impact measurement. At least 8.4% report about failures in a more general manner, e.g. project freeze, lack of funds, or problems with employees.

A final aspect of the analysis was the organizational placement of impact measurement. Only 14.5% of the reports mention a responsible person in the organization for impact measurement or the like. Highlighting responsibilities indicates the importance of a task to internal and external stakeholders.

5 Discussion

The findings of this study reveal critical insights into how Swiss nonprofits approach impact in their annual reports, and how these practices align—or not—with established concepts in the field of nonprofit accountability and impact orientation. While most organizations remain in the conventional setting of an "annual report," there is a growing trend to move towards efficiency (e.g., "progress report" or "activity report") or even effectiveness to signal specific priorities. However, a disconnect between the broader discourse on nonprofit impact and the practical communication strategies used by organizations has to be stated.

As a general perception one can conclude that nonprofits take too much information for granted at their stakeholders. In many reports, basic information such as the nonprofits' mission and primary goals or major impact achievements are not mentioned. A particularly striking finding is the limited use of the annual report to communicate the organization's primary and impact objectives. This suggests that the opportunity to communicate organizational purpose and intended societal value remains underutilized. Referring to constructivist theory, the nonprofits might not even attempt to present their understanding as impact is always socially constructed. At the same time, referencing the Sustainable Development Goals (SDGs) appears widespread, reflecting a growing trend in aligning nonprofit missions with globally recognized frameworks (Meier & von Schnurbein, 2024). This form of isomorphism may serve both to enhance legitimacy and to offer a common language for stakeholders, even if it does not necessarily translate into concrete reporting on impact.

The adoption of structured impact concepts such as social impact, theory of change, or logic framework remains limited. These findings contradict the growing academic and practitioner emphasis on structured, theory-driven impact reporting (Benjamin et al., 2023). A reason may be the complex nature of a full-scale theory of change or logic framework, and nonprofits might fear overstraining the readers, especially funders and donors. However, breaking down a theory of change to in a digestible complexity is also a good exercise to crystallize the core values of the nonprofit's mission. In this context, the RAG model has demonstrated its advantages. Prompting implicit information on core components of a theory of change, the results proved a dominant existence of the impact orientation in the organizations. This gap between implicit and explicit reporting suggests that while organizations may engage in outcome thinking internally, they lack either the capacity or the will to articulate these connections clearly for external audiences. Making these theories more explicit would not only clarify organizational intent and impact logic but also facilitate better understanding and engagement from stakeholders.

Furthermore, while over half of the organizations report on impact in some form, their focus remains heavily skewed toward outputs. Outcome-level reporting, which reflects changes experienced by beneficiaries, is far less common, and only a marginal proportion reports societal-level impact. This imbalance reveals a prevalent issue in nonprofit reporting: the dominance of activity and service metrics at the expense of meaningful change indicators. One explanation might be the sufficiency of output information for regulatory standards. Only recently the most important standard setter for Swiss charities, the Zewo foundation, has clarified its expectations towards impact reporting (Zewo 2024). Although qualitative and quantitative data are used extensively, the substance often falls short of true impact measurement, which ideally should capture not just what was done, but what was achieved in terms of long-term change.

Finally, the organizational embedding of impact measurement is weak. Only 14.5% of reports name a person or unit responsible for this function, indicating a lack of institutionalization. This undermines both the credibility and effectiveness of impact practices. When responsibility for impact is vague or absent, it signals that the task is not yet central to organizational strategy or governance.

Overall, these findings suggest that while Swiss nonprofits are beginning to align rhetorically with impact-oriented frameworks (e.g., SDGs), there is a substantial lag in the systematic and transparent application of impact orientation and reporting. To close this gap, nonprofits must move toward more explicit communication of their theories of change, strengthen their outcome and impact metrics, and institutionalize responsibility for impact within their organizational structures. Doing so would not only enhance transparency and stakeholder engagement but also support strategic learning and long-term effectiveness.

5.1 Application of the RAG model

The RAG model, as a methodology, provides the ability to navigate large volumes of data that are time-consuming to read and often private or not yet accessible to LLMs. It demonstrated the effectiveness of leveraging LLMs' capacity to understand and process texts to perform a variety of tasks. These models can be adapted across domains, with vector store databases incorporating parsed data from various sources and data types. Since the concept of data retrieval remains constant, the RAG model offers fast and efficient responses. Moreover, achieving precise results is possible by selecting an appropriate model structure and ensuring well-crafted prompts in complex task scenarios (Lewis et al. 2020).
As discussed in the methodology section, the basic architecture of RAG models is limited by the context length and the number of retrieved knowledge chunks. These constraints can pose limitations on the volume of data that can be considered when answering a single query. Additionally, these factors may impact the model's accuracy. Therefore, a solid understanding of

both the data and the architecture of RAG models is essential to ensure optimal performance and to explore alternative architecture or configurations when needed. The choice of the LLM within a RAG model can significantly influence its performance, as different LLMs behave differently depending on their architectural complexity and the training data used.
Multilingual capabilities are also important. In our case, it was essential to use a multilingual LLM to handle reports in various languages and to understand prompts across linguistic contexts. These requirements made GPT-4o the ideal choice, given its multilingual strengths and state-of-the-art performance at the time of the study.
However, all LLMs developed to date still exhibit limitations, including occasional misunderstandings and hallucinations. As a mitigating strategy and a direction for future work, a Mixture of Experts RAG (MoE-RAG) model can be developed and utilized (Xiong and Zheng, 2024). An MoE-RAG, which integrates multiple retrieval techniques and diverse, complex LLMs, may offer a more accurate RAG model. Although it would entail higher computational complexity, it holds the potential to significantly reduce bias and improve reliability.

6 Conclusion

The purpose of this study was to apply a RAG model to analyze the impact related content of nonprofit annual reports. Previous research was limited by factors such as time restraints for in-depth analysis of large amounts of annual reports or lack of possibilities to include qualitative data in quantitative methods such as financial health analysis. The distinctiveness of a RAG model compared to more generic AI-based analysis is the clear definition of the analyzed documents. This hinders typical problems of LLMs such as hallucinations or reliance on outdated sources. However, prompting proved to be a challenge as the RAG-model needs very distinct and clear commands to lead to satisfactory results. Especially in the context of nonprofits where usually qualitative results are more important than numbers, this method offers many areas to apply whenever many similar texts needs to be analyzed to find connections or overlaps. For theory and practice, the application of a RAG-model might be beneficial for the analysis of project reports, grant submissions, or mission statements. Our analysis revealed that Swiss nonprofits annual reports lack explicit mentions on impact, although the information is available. Improving clarity and visibility of this information can help to increase awareness among donors and funders and to better proof the good results of the organization's activities.

References

AbouAssi, K., & Bies, A. (2017). Relationships and resources: the isomorphism of nonprofit organizations' (NPO) self-regulation. *Public Management Review*, *20*(11), 1581–1601. https://doi.org/10.1080/14719037.2017.1400583

Anderson, A. 2004. Theory of Change as a Tool for Strategic Planning. A Report on Early Experiences. Washington, DC: The Aspen Institute.

Arslan, M., Ghanem, H., Munawar, S., & Cruz, C. (2024). A Survey on RAG with LLMs. *Procedia Computer Science*, *246*, 3781-3790

Benjamin, L. M. (2013). The potential of outcome measurement for strengthening nonprofits' accountability to beneficiaries. *Nonprofit and Voluntary Sector Quarterly*, *42*(6), 1224-1244.

Benjamin, L. M., Ebrahim, A., & Gugerty, M. K. (2023). Nonprofit organizations and the evaluation of social impact: A research program to advance theory and practice. *Nonprofit and Voluntary Sector Quarterly*, *52*(1_suppl), 313S-352S.

Chew, R., Bollenbacher, J., Wenger, M., Speer, J., & Kim, A. (2023). LLM-assisted content analysis: Using large language models to support deductive coding. *arXiv preprint arXiv:2306.14924*.

Cordery, C., & McConville, D. (2022). Annual reporting in voluntary organisations: opportunities for content analysis research. In *Researching Voluntary Action* (pp. 110-121). Policy Press.

Coy, D., Fischer, M., & Gordon, T. (2001). Public accountability: A new paradigm for college and university annual reports. Critical Perspectives on Accounting, 12(1), 1–31. DOI: 10.1006/cpac.2000.0416

Flack, E. D. (2007). *The role of annual reports in a system of accountability for public fundraising charities* (Doctoral dissertation, Queensland University of Technology).

Gordon, T.P., Khumawala, S.B., Kraut, M. and Neely, D.G. (2010), Five dimensions of effectiveness for nonprofit annual reports. Nonprofit Management and Leadership, 21: 209-228. https://doi.org/10.1002/nml.20021

Greenfield, J. M. (2001). Fundraising management. In T. D. Connors (Ed.), The nonprofit handbook: Management. New York: Wiley.

Harrison, H. (2022) *LangChain*. GitHub. Available at: https://github.com/langchain-ai/langchain (Accessed: 12 May 2025).

Harwood, T. G., & Garry, T. (2003). An overview of content analysis. *The marketing review*, *3*(4), 479-498.

Helmig, B., Ingerfurth, S. and Pinz, A. (2014) 'Success and failure of nonprofit organizations: Theoretical foundations, empirical evidence, and future research', Voluntas: International Journal of Voluntary and Nonprofit Organizations, 25(6): 1509–1538.

Herman R. D., Renz D. O. (2008). Advancing nonprofit organizational effectiveness research and theory: Nine theses. *Nonprofit Management and Leadership*, 18(4), 399–415.

Hersberger-Langloh, S. E., Stühlinger, S., & von Schnurbein, G. (2021). Institutional isomorphism and nonprofit managerialism: For better or worse?. *Nonprofit Management and Leadership*, *31*(3), 461-480.

Hyndman, N. and McConville, D. (2016). Transparency in reporting on charities' efficiency: A framework for analysis. Nonprofit and Voluntary Sector Quarterly, 45(4): 844–865.

Krishnan, R., Yetman, M. H., & Yetman, R. J. (2006). Expense misreporting in nonprofit organizations. Accounting Review, 81(2), 399–420. DOI: 10.2308/accr.2006.81.2.399

Lecy, J. D., & Searing, E. A. (2015). Anatomy of the nonprofit starvation cycle: An analysis of falling overhead ratios in the nonprofit sector. *Nonprofit and Voluntary Sector Quarterly*, *44*(3), 539-563.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.T., Rocktäschel, T. and Riedel, S., 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, *33*, pp.9459-9474.

Litofcenko, J., Karner, D. & Maier, F. Methods for Classifying Nonprofit Organizations According to their Field of Activity: A Report on Semi-automated Methods Based on Text. *Voluntas* **31**, 227–237 (2020). https://doi.org/10.1007/s11266-019-00181-w

Ma, J. (2021). Automated coding using machine learning and remapping the U.S. nonprofit sector: A guide and benchmark. *Nonprofit and Voluntary Sector Quarterly, 50*(3), 662–687. https://doi.org/10.1177/0899764020968153

Maier, F., Meyer, M., Burkart, C., & Terzieva, B. (2025). Societal roles of nonprofit organizations: Parsonian echoes and Luhmannian reframing of the organization–society interface. *Nonprofit and Voluntary Sector Quarterly*, *54*(2), 327-349.

Maurer, L. L. (2014). *Board member perceptions of nonprofit organization effectiveness*. Walden University.

McConville, D. (2017) New development: Transparent impact reporting in charity annual reports – benefits, challenges and areas for development. Public Money & Management, 37(3): 211-216.

McCulloch, W.S. & Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), pp.115–133.

Meier, D. S. (2023). The evolution of SDG-related third sector and public administration literature: an analysis and call for more SDG-related research. *Sustainability: Science, Practice and Policy*, *19*(1). https://doi.org/10.1080/15487733.2023.2236501

Meier, D. S., & von Schnurbein, G. (2024). From Mission to Market: Assessing Sector Overlap Between Nonprofits and For-Profits. *Nonprofit and Voluntary Sector Quarterly*, *0*(0). https://doi.org/10.1177/08997640241300509

Meta (2024) *Llama-3.1-8B-Instruct*. Hugging Face. Available at: https://huggingface.co/meta-llama/Llama-3.1-8B-Instruct (Accessed: 12 May 2025).

Meta (2024) *Llama-3.3-70B-Instruct*. Hugging Face. Available at: https://huggingface.co/meta-llama/Llama-3.3-70B-Instruct (Accessed: 12 May 2025).

Mistral AI (2024) *Mistral-7B-Instruct-v0.3*. Hugging Face. Available at: https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.3 (Accessed: 12 May 2025).

Mistral AI (2024) *Mixtral-8x22B-Instruct-v0.1*. Hugging Face. Available at: https://huggingface.co/mistralai/Mixtral-8x22B-Instruct-v0.1 (Accessed: 12 May 2025).

OpenAI (2024) *Hello GPT-4o*. Available at: https://openai.com/index/hello-gpt-4o/ (Accessed: 12 May 2025).

Parsons, L. M. (2003). Is accounting information from nonprofit organizations useful to donors? A review of charitable giving and value-relevance. Journal of Accounting Literature, 22, 104–129.

Patrut, M., Cmeciu, C., & Miron, L. (2011). NGO annual reports as a device to frame education in Romanian public communication campaigns. Public Relations Review, 37, 432–434. DOI: 10.1016/j.pubrev.2011.06.006

Polizzi, S. (2022). *Risk disclosure in the European Banking Industry: Qualitative and quantitative content analysis methodologies*. Springer Nature.

Prinsen, G., & Nijhof, S. (2015). Between logframes and theory of change: reviewing debates and a practical experience. *Development in Practice*, *25*(2), 234-246.

Ryan, C., Dunstan, K., & Brown, J. (2002). The value of public sector annual reports and annual reporting awards in organisational legitimacy. Accounting, Accountability & Performance, 8(1), 61-76.

Skaerbaek, P. (2005). Annual reports as interaction devices: The hidden constructions of mediated communication. Financial Accountability and Management, 21(4), 385–411. DOI: 10.1111/j.0267-4424.2005.00226.x

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. *Advances in neural information processing systems*, *30*.

von Schnurbein, G. (2016). Measuring impact and recognizing success. In *The Routledge companion to philanthropy* (pp. 468-481). Routledge.

von Schnurbein, G., & Fritz, T. M. (2017). Benefits and Drivers of Nonprofit Revenue Concentration. *Nonprofit and Voluntary Sector Quarterly*, *46*(5), 922-943. https://doi.org/10.1177/0899764017713876

Xiong, X. and Zheng, M., 2024. Merging mixture of experts and retrieval augmented generation for enhanced information retrieval and reasoning.

Zewo Foundation. (n.d.). *Zewo – Your donation in good hands*. Retrieved May 12, 2025, from https://zewo.ch/en/

Zewo, S. (2024). Die 21 Zewo-Standards. *Stiftung Zewo, Zürich. https://zewo.ch/wp-content/uploads/2024/02/21-Zewo-Standards-DE-2024_02.pdf. Retrieved May 12, 2025*.

Xiong, G., Jin, Q., Lu, Z. and Zhang, A., 2024, August. Benchmarking retrieval-augmented generation for medicine. In *Findings of the Association for Computational Linguistics ACL 2024* (pp. 6233-6251).

Ke, Y., Jin, L., Elangovan, K., Abdullah, H.R., Liu, N., Sia, A.T.H., Soh, C.R., Tung, J.Y.M., Ong, J.C.L. and Ting, D.S.W., 2024. Development and Testing of Retrieval Augmented Generation in Large Language Models--A Case Study Report. *arXiv preprint arXiv:2402.01733*.

Annex 1: Prompt example

*«An **impact report** is a report that presents the impact and success of an organization or project. It describes how and to what extent the goals have been achieved, which measures were taken, and which measurable results were obtained. The report is particularly important for non-profit organizations in order to ensure transparency and document goal achievement.*

*There are significant differences between **Output, Outcome,** and **Impact** that should not be confused with an impact report:*
***Output** refers to the immediate results or products of an activity (e.g., the number of trainings conducted).*
***Outcome** refers to the short- or medium-term effects of the outputs (e.g., the participants' skills after the training).*
***Impact** describes the long-term, comprehensive changes caused by the activities (e.g., an improved standard of living in the community).*
*Carefully review the document and the following context: **{context}**. Then answer the question: **'Does an impact report exist for the organizations?'***

Answer format:

Provide a brief answer with 'Yes' or 'No'. Only answer 'Yes' if an impact report is explicitly mentioned in the document.
In the second part, explain why you chose this answer. If an impact report is mentioned in the document, describe the context and quote relevant passages from the document. Make sure not to incorrectly interpret outputs, outcomes, or impacts as an impact report.»