ETH zürich

Analysing Sustainability Reports Using Machine Learning

Report an die Digitale Verwaltung Schweiz (DVS) zur Leistungsvereinbarung

«IN8 Anwendung von NLP auf die nicht-finanzielle Berichterstattung von Schweizer Unternehmen»

Authors (in alphabetical order):
Fritz Brugger, PhD, David Etienne, Johannes Hool, Roman Küpper, David Weiss

Zürich, 31 January 2023

ETH Zürich
Fritz Brugger, PhD
Center for Development and Cooperation
Clausiusstrasse 37
8092 Zürich

bruggerf@ethz.ch
https://nadel.ethz.ch/

Table of contents

Executive summary

Starting with 2023, new legislation makes human rights due diligence and non-financial reporting mandatory for companies registered in Switzerland. This report explores the feasibility of using Natural Language Processing (NLP) for the automated analysis of sustainability reports. While efforts are under way for NLP-assisted analysis of environmental, and in particular, climate reporting, this is not yet the case for the social dimensions of sustainability.

We focus on the social dimensions of non-financial reporting, testing the potential of NLP-assisted analysis to reveal the content reported on human rights. However, we do not aim to assess whether a company meets its legal reporting requirements in this area.

At the conceptual level, the challenge includes defining what to look for, given that the legal provisions only vaguely define what companies must report. We analyse the scope and depth of reporting along the human rights management cycle.

Based on an analysis of existing NLP methods, we identify semantic search with sentence transformers[1] as best suited for this purpose. We build a sentence transformer-based text classifier[2] that searches sentences matching the structure of a set of query sentences. The query sentences are defined by sustainability experts and cover reporting on human rights along all steps of the management cycle.

The results show that our sentence transformer-based text classifier has a high potential to identify the relevant sections in a sustainability report. Yet, allocating the text blocks identified to the corresponding section of the management cycle proves more difficult. Following the example of ClimateBERT[3], a next step could be to build ethicaLM, a model trained specifically to target the language used to report on social dimensions of sustainability.

[1] A transformer is a deep learning model that adopts the mechanism of self-attention, differentially weighting the significance of each part of the input data.
[2] Text classification is a common NLP task that assigns a label or class to text.
[3] https://climatebert.ai/

Context

The Swiss Parliament has adopted new regulations on human rights due diligence and non-financial reporting for companies as of January 1, 2022. Starting with the 2023 financial year, the Swiss Code of Obligations[4] now requires Swiss companies to report on environmental, social and labour issues, as well as human rights and the fight against corruption. Such non-financial reporting is often referred to as “sustainability reporting” or as “environmental, social and governance (ESG) reporting”.

Larger companies have already been publishing such reports for quite some time. However, under the new regulation, all companies with more than 500 employees and an annual turnover of 40 million CHF in two subsequent years must publish sustainability reports. Estimations on the number of companies that will be affected range from several hundred to over a thousand companies.

This expected influx of reports raises the question of how to monitor compliance with the new law: manual review and verification of sustainability reports is time consuming and currently it is not clear whether the administration is ready to invest the human resources expected to be necessary for a systematic annual review of the sustainability reports. This information, however, is of significant public interest, as this new law was introduced as a counter-proposal to the Swiss Responsible Business Initiative, which was approved by a majority of Swiss voters.

We explore whether and how Natural Language Processing (NLP) can be used for the automated assessment of sustainability reports. The analysis includes the assessment of the scope of human rights reporting, i.e., whether it reports on all aspects of managing human rights issues (referred to as 'management cycle’), from establishing whether and how the company is exposed to human rights risks, the strategy and actions to address it, the method to measure progress, and the results achieved. The analysis further includes the depth of reporting, i.e., the level of detail in which a company reports on each step of the management cycle.

Ideally, NLP-supported automated analysis of reports would allow legislators and interested stakeholders to continuously evaluate the entire body of sustainability

[4] The legal basis of the non-financial reporting requirements is defined in the following documents:
- OR 964$^{\text{bis-septies}}$, https://fedlex.data.admin.ch/filestore/fedlex.data.admin.ch/eli/cc/27/317_321_377/20230101/de/pdf-a/fedlex-data-admin-ch-eli-cc-27-317_321_377-20230101-de-pdf-a-9.pdf
- Verordnung über Sorgfaltspflichten und Transparenz bezüglich Mineralien und Metallen aus Konfliktgebieten und Kinderarbeit (VSoTr) (Ordinance), https://www.fedlex.admin.ch/eli/cc/2021/847/de
- Erläuternder Bericht zur Verordnung über Sorgfaltspflichten und Transparenz bezüglich Mineralien und Metallen aus Konfliktgebieten und Kinderarbeit (VSoTr) (Explanatory report), https://www.skmr.ch/cms/upload/pdf/2022/220803_Erlauternder_Bericht_VSoTr.pdf

reports and, hence, the implementation of the new regulation, at significantly reduced costs compared to manual assessment.

This report proceeds as follows. First, we outline the key challenges of analysing non-financial reporting. Based on this stocktaking, we discuss conceptual approaches to the analysis and explain the approach chosen for this report. Next, we present the technical approach to the analysis, the arrangement to test the outcome of the analysis, and the results. We conclude with recommendations for the way forward.

The challenges at hand

Analysing sustainability reports has three main challenges:

Challenge 1: Diversity

The first challenge comes from the diversity of topics and the inherent complexity of these topics that sustainability reporting, as required by the new regulation, is expected to cover. The five required topics include: Environmental issues including climate change, social issues, employee issues, human rights and corruption. Companies with a potential exposure to child labour and conflict minerals are subject to additional reporting requirements.

Companies covered by the Act have to report on different aspects of each of these five topics in a way that follows the generic “management cycle”. Namely companies must report on their business model, concepts and due diligence procedures relating to the five topics, measures taken and impact thereof, and a description of the risks emerging from those non-financial topics. This level of detail means that identifying whether a topic such as “climate change” is being reported on is not enough. The evaluation of a report also needs to assess whether these dimensions of the management cycle are covered for each topic. That said, the provisions in the law are too unspecific to serve as the sole source for the operationalisation of the analysis. A more detailed summary of the legal provisions is provided in Annex 1.

Challenge 2: Context-dependency

The second challenge relates to the fact that each company both impacts and is affected by each sustainability topic differently, depending on the business sector, business model, and the company’s geographical presence. Accordingly, the scope of reporting that could be reasonably expected from a company in order to meet the new regulatory requirements is different for each company. To clarify the scope of reporting, companies need to assess their exposure to the full range of sustainability

risks. Risks that affect a company in a significant way are called material risks and the analysis thereof is called ‘materiality analysis’. The consequence is that we have a generic understanding of what sustainability reporting covers, but we do not exactly know what a specific company can be reasonably expected to report about against the background of their specific operations. In addition, under certain circumstances, a company is exempt from reporting on some critical topics, such as child labour or conflict minerals.

Challenge 3: Lack of recognized baseline

As of now, sustainability reporting lacks precise definitions and requirements, in stark contrast to the strongly formalised and prescriptive nature of financial reporting. In sustainability reporting, there is no structured reporting template or systems that companies must follow when drafting their report. There are some international guidelines, with GRI[5] being the most prominent one, but these are voluntary templates and the Swiss legislation makes no formal prescription on the reporting format. In addition, language, including the wording of topics and processes, varies significantly across companies. Sustainability reports typically contain a highly varied mix of structured, tabular and quantitative data mixed with qualitative text and a varying number of pictures and charts. Finally, sustainability reports are often published, not as a single document, but as separate pieces of content, including documents hyperlinked in the main report, or, excel files containing results, raising the question of what exactly has to be included in the analysis and how this can be retrieved without missing key elements.

The formal requirements for the 2022 EU Corporate Sustainability Reporting Directive[6] are more rigid. Under the updated regime of the 2014 non-financial reporting directive, companies submit their report in XHTML format and ‘tag’ their reported sustainability information according to a digital categorisation system. Similarly, the German supply chain act[7] — the German counterpart to the new Swiss regulation — requires companies to submit a report in a structured format that is specified by the respective German administrative authority.

[5] https://www.globalreporting.org/
[6] https://www.consilium.europa.eu/en/press/press-releases/2022/06/21/new-rules-on-sustainability-disclosure-provisional-agreement-between-council-and-european-parliament/
[7] https://www.bafa.de/DE/Lieferketten/Berichtspflicht/berichtspflicht_node.html

Conceptual approach

As discussed in the previous section, the legal provisions remain vague regarding what to assess when analysing sustainability reports and how to evaluate what we find. We discuss three conceptual approaches to the analysis[8].

Option 1: Assessing compliance

One option is to develop a benchmarking system that results in a "yes" or "no" verdict, i.e., the company has met its reporting obligation or it has not.

Although this approach would, in theory, deliver an authoritative answer, the reporting requirements as formulated in the Act are far too generic and unspecific, making it impossible to evaluate reports against the provisions of the Act. It is unclear what information needs to be provided for a report to be considered sufficient. Is it sufficient to provide anecdotal evidence of some achievements? Or does adequate reporting on results require that information is provided on all aspects that were identified as relevant? Should analysts expect finding systematic data on the situation at the beginning and at the end of the reporting period, including a discussion of change that was measured, or why no change was observed? As mentioned in the section above, unlike in financial reporting, clear standards are missing. Even then, thresholds would have to be set for a verdict to be made.

We conclude that the conditions are not in place to assess compliance with the law because it is too ambiguous. Doing so would be marred by countless normative decisions by those who design the assessment rules.

Option 2: Categorising

An alternative approach is to map the scope and depth of a sustainability report. This means that we do not evaluate the compliance of a report with the law, but rather assess the scope and depth of a report. This approach allows for a more nuanced statement instead of directly answering whether a company meets the legal reporting requirements with a "yes or no" statement. For each of the five substantive sustainability topics spelled out in the Act, we can ask: Does the company report whether and how this sustainability topic is 'material' for the company? Does the company report a commitment to address the topic and formulate an ambition or a strategy to address it? Does the company detail measures that it has taken in the reporting year to mitigate the respective risks? Does the report explain the company's approach to measuring performance and the results of this monitoring?

[8] For the purpose here, we assume that we can identify all companies that are required to report under the Act and that all reports can be retrieved in full (i.e., including all relevant side documents that need to be considered) in an automated way.

However, similar to the first approach, it is still unclear what the conditions should be to definitively say that the company has addressed a topic. Does, for example, the reference to one single measure taken translate into “ticking the action box” or is more needed? How much is enough? Instead of answering this question with a yes/no assessment, one could count how many statements on a topic a company makes for each step of the management cycle. The results would show the relative distribution of the reporting effort of a company and can be visualised in a normalised way, as illustrated, for example, in Figure 1.

Such an approach has the advantage that readers can quickly get an idea of where the company places the emphasis in terms of topics as well as steps along the management cycle. Readers can easily see whether the company reports a commitment only or whether the company can report along the entire management cycle from the identification of the materiality all the way down to presenting indicators to measure progress and the respective results.

In this approach, NLP is used to map the relative scope and depth of the issues covered. It is important to emphasise that this approach does not produce a statement on legal compliance. Even if the NLP analysis finds that a report covers all topics and all aspects along the management cycle in considerable depth, this does not mean that the report complies with the legal requirements.

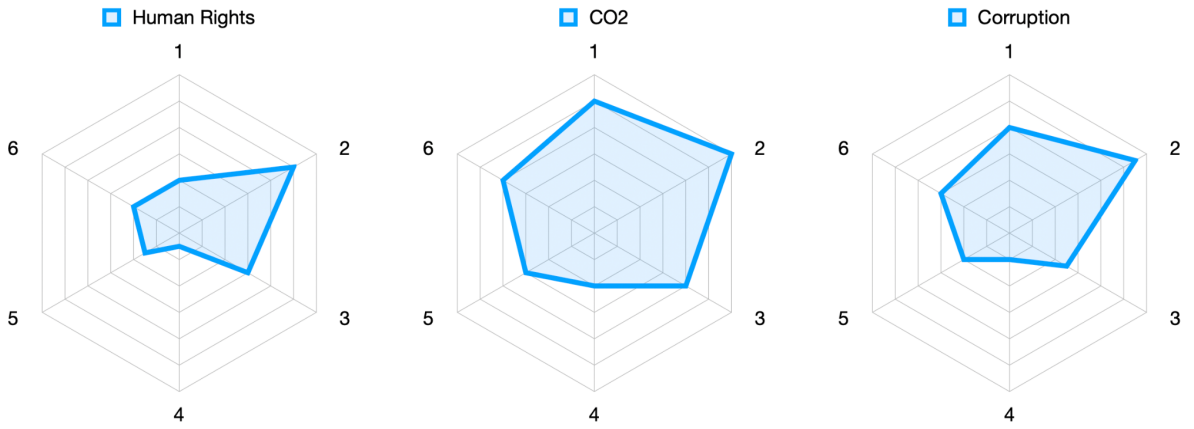

Figure 1: Illustrative visualisation of the scope and depth of reporting on the topics of human rights, CO_2 emissions and corruption found in a fictitious sustainability report.

Option 3: Highlighting

The third option is conceptualised as an information filtering system. It emphasises the importance of human — not machine — interpretation of what a company reports and uses NLP to simplify and speed up the human evaluation. In this option, instead of counting and summing occurrences of issues mentioned in a report, NLP is used to

highlight the segments of text where relevant content is found directly in the sustainability report. This has the advantage that a reader is guided directly to those sections within the report that are deemed relevant with regards to the mandatory reporting requirements. In addition, the reader also finds the statements on the topic embedded in the wider context of the report, which is important for proper understanding to enable an expert judgement.

On the downside, with this approach it is not possible to make aggregated statements on a report nor to make comparisons between reports, including overall findings and trends. Technically, this approach builds on a similar identification strategy as the second option; therefore, the two can be combined, providing additional support when analysing sustainability reports.

Decision on the conceptual approach

Based on this discussion, we first acknowledge that we consider it to be infeasible to answer the initial question in a fully automated way without human intervention after the fact, i.e., determining whether a non-financial report meets the legal reporting requirements. The legal provisions, including the explanations in the ordinance and the explanatory comment, are not clear enough to derive unambiguous evaluation criteria. Having said that, this lack of clarity affects both automated assessments and manual assessments by human assessors in a comparable way.

It will be necessary to rely on evaluation criteria that we derive from what is considered good practice in non-financial reporting, including those referred to above. With these realities in mind, we choose Option 2 as the preferred approach to analyse sustainability reports. It maps reporting patterns, making it possible to draw some conclusions regarding compliance with the new regulation, e.g., by assessing whether the reporting covers all aspects, from strategy to results, in a similar manner or whether it reports extensively about commitments and strategies, but not about results. This approach also enables comparison between companies form the same business sector or between those that have to report versus those that do not under this legislation. With Option 2 it is also possible to analyse the evolution of reporting patterns over time, e.g., before and after the introduction of the legislation, but also in the subsequent years following the enactment of the reporting obligation.

Technical approach

The transformer-based architecture

In recent years, there have been many advances in the field of data science and especially in the field of NLP. These advances have greatly reduced the effort required to accomplish many complex tasks, such as question answering, text prediction, generation and summarization, and sentiment analysis. The introduction of models built on a transformer-based architecture has been particularly impactful. In 2018, for instance, Google introduced a model called BERT[9], a powerful machine learning model for NLP applications that could outperform previous language models in different benchmark datasets. BERT and similar models are trained on huge amounts of text from an unprecedented number of online resources. This training allows the model to learn representations of words and patterns in everyday language. Hence, when it comes to complex textual modelling, pre-trained transformer-based models such as BERT are preferred because of their ease of use and improved performance.

Since the publication of the transformer-based architecture, many transformer models have been released by a wide range of actors. The following timeline[10] highlights a few of the most popular ones:

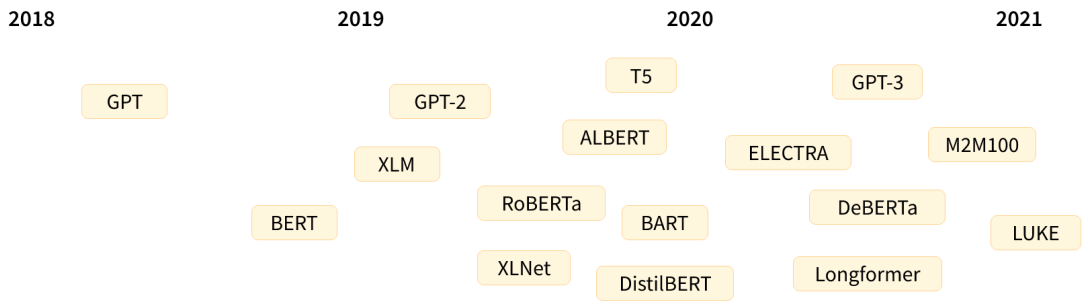

Figure 2: Some reference points in the (short) history of transformer models.

The big difference between transformer-based and earlier language models is that transformers understand the context in which a word is used. For instance, it is far from trivial for machines to determine the difference between “Bob is running a marathon” and “Bob is running a company”. Pre-transformer models would put both instances of “running” into the same semantic box of “walking fast”, while for transformer-based models, it is clear that “running” in the second sentence implies that Bob leads a company and that the sentence has nothing to do with the act of walking.

---

[9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova: “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, 2018; [http://arxiv.org/abs/1810.04805 arXiv:1810.04805]

[10] https://huggingface.co/course/chapter1/4

Identification of potential technical approaches

At the time of this project, a few well-documented application cases were already available for transformer-based models. These examples provided a broad basis for exploring whether a transformer-based model was a suitable technological solution to answer the question defined in the scope of this project. Several transformer-based techniques were considered to meet the challenge. Two approaches appeared to be the most promising:

Option 1: Text classification

Text classification is the process of categorising text into a group of words. Using NLP, it is thus possible to automatically analyse the text and then assign a set of predefined tags or categories to it according to its context. The most common form is binary classification, or assigning one of two categories to all documents in the corpus. For example, by analysing the content of an email, an incoming message could be classified as spam. Another example is ClimateBERT[11]. In this case, the BERT model has been specifically trained to improve its ability to process climate-related texts.

Text classification is a solution with great potential to tackle the challenges described in this project effectively. On the one hand, it promised very good performance, as it is one of the current state-of-the-art approaches in the NLP field. A relatively small dataset (approximately 2000 sentences) is required to train the model in a supervised manner, which would already be sufficient to obtain promising preliminary results. On the other hand, building a representative dataset is not a trivial task and its quality heavily depends on the labelling process.

Option 2: Semantic search

Semantic search is a task that involves finding sentences that are similar to a given sentence in meaning. Given a paragraph of several sentences, a semantic search model could return the sentence pairs that are the closest in meaning to each other. Sentence transformers such as SBERT, a modification of the standard BERT model, allow many sentences to be encoded and compared based on semantic similarity.

The potential of this technique is that there are working solutions available, which would allow moving from lexical search to semantic search logic. In concrete terms, in the case of lexical search, the search engine looks for literal matches of query words or variants of them, without understanding the overall meaning of the query. In the case of semantic search, the input could be full sentences, including their overall meaning. Training dataset requirements are in the same range as for option 1.

[11] Nicolas Webersinke, Mathias Kraus, Julia Anna Bingler, Markus Leippold: “ClimateBert: A Pretrained Language Model for Climate-Related Text”, 2021; [http://arxiv.org/abs/2110.12010 arXiv:2110.12010].

Decision on the technical approach

The deliberations within the project team led to the decision to follow *Option 2: build a text classifier based on sentence transformers*, i.e., make use of the ability to perform a semantic search on the analysed reports by inputting a set of query sentences. The idea is to pursue the conceptual approach presented above: assessing the scope and depth of reporting, and leverage the fact that sentence transformers make it possible to move away from a lexical search, i.e., counting keywords, to semantic search, which takes into account the whole input sentence.

One element that weighed heavily in this decision was the size of the project and therefore the limited capacity of both the thematic and technical teams. Indeed, initially *Option 1: text classification* was the preferred option. However, after a thorough analysis, it seemed unlikely to achieve a conclusive result since only a limited panel of domain experts was available, yet a large panel of experts is needed to correctly label the 2000 or so sentences to build the required training dataset.

To build a sentence transformer-based text classifier to assess sustainability reports, a set of query sentences, called topical sentences in the rest of this report, is a required central element. The topical sentences describe the search topic as exhaustively as possible. Based on these topical sentences, the text classifier performs a semantic search and returns the similar sentences recognised in the analysed text. The effort required from the domain experts to build the set of topical sentences seemed more reasonable than the effort required to obtain a valid dataset for the text classification approach. From a technical point of view, it also seems that less effort is required for implementation, making the sentence transformer-based text classifier achievable within the scope of this project.

The challenge of text extraction

The different transformer-based approaches described so far attempt to address two of the three challenges mentioned at the beginning of this report, namely the diversity of topics in sustainability reporting and their context dependency. However, there remains the third challenge: the format of sustainability reports. The lack of standardisation coupled with the fact that the reports generally contain a very diverse mix of structured, tabular and quantitative data, qualitative text, and varying numbers of images and graphics, represents a significant technical challenge.

The first step required to automate report analysis is text extraction, i.e., converting a PDF document into a format that can be recognised by a machine. For this purpose, there are many tools available. An interesting and especially relevant comparison for

this project was made by the Artificial Intelligence against Modern Slavery[12] (AIMS) project, in which the starting points are also reports available in PDF format, sometimes only in printed form.

However, these tools have clear limitations. First, the quality of the extracted text is often suboptimal. For instance, sometimes words are missing or sentences are split in an unintuitive manner. This is problematic as poor-quality data is fed to the NLP-based analysis method, whether it leverages a transformer-based technique or not. In some instances, this may drastically lower the output quality. Second, depending on the method used, all non-text information presented in a report, i.e., images, graphs, or tables, cannot be extracted in a straightforward way. This data is therefore lost at the first stage and does not even enter the subsequent analysis.

Although there are technologies that could remedy these problems, the efforts required to achieve satisfactory results seem very high. Therefore, it seems valid, from a technical point of view, to question the relevance of having a reporting obligation without defining a standard format that can be easily analysed — at least by machines. As already mentioned earlier in this report, the European Council as well as the German administration are taking concrete steps in the direction of standardised reporting.

Building a sentence transformer-based text classifier

Setting up a test framework

We decided to build a sentence transformer-based text classifier and to focus on the topic of human rights for practical reasons. We selected the topic of human for two reasons. First, we wanted to select a topic from the social realm of the sustainability universe, since environmental issues, in particular CO_2, get a lot more attention in the current discourse. Second, human rights is a broad topic as it includes not only fundamental human rights, but also labour rights, children's rights, rights of indigenous communities, political rights, etc. Consequently, several terms are used in reporting that address different aspects of human rights. We created a list of terms that are used in sustainability reporting based on the literature and expert knowledge (see Annex 2). This complexity is well suited to test the potential of NLP.

Our sentence transformer-based text classifier uses topical sentences as a starting point so that it can then perform a search function to identify all similar sentences in the text. Since there are many ways to report about the same topic (within a single

[12] https://github.com/the-future-society/Project-AIMS-AI-against-Modern-Slavery/tree/main/%F0%9F%97%84%EF%B8%8F%20Data%20and%20text%20extraction#analysis-and-learnings

report and even more diversity between reports), a single-sentence approach would not find all relevant information in a report. Therefore, we opted for a 'portfolio approach': We created a list (see Annex 2) of about 8-12 ideal-type topical sentences for each step along the management cycle based on domain expert knowledge and the analysis of sample reports. A portfolio of topical sentences to represent a topic has the advantage that we are more likely to capture the different reporting styles that characterise different sustainability reports. To make matters more complicated: The portfolio of topical sentences must be multiplied by the items on the list of key terms in order to represent the full range of information we are searching for.

An important downside to the portfolio approach is that a quantitative evaluation of the findings is more difficult since many of the sentences are alternatives. In other words, the equation "more topical sentences found = better result" does not hold. Also, assuming that we can reduce the number of topical sentences to those that get the most hits would be incorrect as there are topics on which fewer companies have risk exposure and hence, fewer companies that report, which translates into fewer — but not less important — reporting instances.

For testing and validation, we purposefully selected a range of sustainability reports considered comprehensive, average, or poor (based on expert judgement) from which we have distilled a set of validation sentences for each report.

Testing goals

The process of discussing different use cases for applying NLP on sustainability reporting and the exchange between the technical team and experts in the domain of sustainability reporting already yielded insights and qualitative assessments which will be presented later in the results section. Additionally, to assess the performance of the sentence transformer-based text classifier, a quantitative evaluation is required.

In the chapter about the conceptual approach, we decided to focus on assessing the scope and depth of reporting as this seems to be the most promising way to evaluate the reports. To be of assistance, our sentence transformer-based text classifier has to identify all relevant passages in the reports, as well as attribute them to the different steps in the management cycle. We therefore test the following two qualities in our quantitative evaluation:

1. Recall: The sentence transformer-based text classifier finds all text passages related to human rights in the sustainability report and assigns them to the respective step of the management cycle.
2. Precision: The text classifier finds the same passages that were also identified by the domain experts as relevant and — importantly — it does not label sentences that were not identified by the experts (false positives).

Testing strategy

To test for the qualities described in the preceding chapter we had to conduct the following steps:

1. Let domain experts devise many topical sentences that can be encoded using sentence transformers to identify relevant text passages in sustainability reports.
2. Let domain experts create a test dataset by identifying all relevant text passages in the reports.
3. Implement a baseline approach using a keyword search method (naïve baseline approach, or 'bag of words' approach) to qualify the results we produce with the sentence transformer-based text classifier.
4. Implement the sentence transformer-based text classifier.
5. Match the text passages identified by the experts to the parsed text data of the reports.
6. Assess recall and precision by comparing the text passages identified by our models with those identified by the domain experts.

The following sections describe the main design decisions made along the way.

Creating the topical sentences

The topical sentences are a central part of our classifier, as they form the link between the domain experts' knowledge and the machine. Based on these sentences, our classifier identifies relevant text passages in the sustainability reports. Close collaboration between the domain experts and the technical team was therefore required to create a set of topical sentences for each stage of the management cycle.

To achieve this, a tool to interact with an early version of our sentence transformer-based text classifier was built, so that the domain experts could get a better feeling of what output the machine could produce. Secondly, once the tool was in place, we went through several iterative rounds between the domain experts and the technical team to find a systematic approach to refine the topic sentences. We quickly noticed, however, that it was complicated to identify clear relationships between the input sentences and the machine output.

The first results produced by the text classifier looked promising. However, two central challenges emerged. On the one hand, it was not easy for domain experts to decide which topic sentence was representative and within which framework. Thus, getting a first draft of sentences took a lot longer than expected. On the other hand, without a first draft of sentences and clarity of what exactly was being sought, it was difficult to move forward on the technical side and establish a systematic way to improve the output of the machine, creating a bit of a "chicken and egg" problem. Nevertheless,

despite these challenges, we managed to create a list of topical sentences for the entire management cycle, available in Annex 2 or in the project repository[13].

The set of topical sentences is crucial for the quality of the search results. For the next iteration of the project, further enhancing the topical sentences would be one of our main priorities, as we expect it to lead to the most improvements to the results.

Creating a validation data set

The validation data is essential to be able to test our sentence transformer-based text classifier. First, the domain experts selected three sustainability reports considered comprehensive, average or poor, based on their judgment. They then independently extracted the passages they considered relevant for each stage of the management cycle, i.e., the extracts that must be found by the classifier for an analysis to be considered complete. Finally, the experts adopted a collaborative process to combine the identified extracts into a single validation document for each selected report. These documents form the final validation set used within the scope of this project.

The domain experts felt it was a challenge to maintain the relevance of information taken out of context. For example, a sentence loses considerable meaning when isolated from the surrounding paragraph. As sustainability reports are full of graphs and tables, the same applies to information presented in this form: some texts, mentioned in a coloured box for example, can lose all their meaning if presented separate from their surroundings in the report. To make things more complicated, from a technical point of view, isolated sentences were necessary, as these are what correspond to the text classifier output. It was therefore necessary to find a consensus and we finally agreed to work with paragraphs, rather than individual sentences. The final validation documents are found in Annex 3 or in the project repository[14].

Implementing the sentence transformer-based text classifier

We have implemented a sentence transformer-based text classifier to leverage the power of semantic search. Semantic search is a method of searching through body of text in which the search engine understands the intent and context of the query, rather than just matching keywords. It uses NLP techniques to understand the meaning of the query and returns results that are semantically related to the query, rather than just containing the keywords.

[13] https://github.com/planet-10/sentence-tranformer-based-text-classifier/blob/main/00_data/queries/query_sentences.json
[14] https://github.com/planet-10/sentence-tranformer-based-text-classifier/blob/main/00_data/validation_data/validation_paragraphs.json

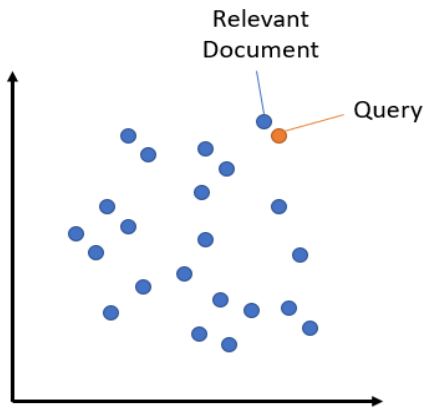

Figure 3: Vector representation of semantic search[15].

The concept behind semantic search is illustrated in Figure 3. The query and all documents are embedded into a high-dimensional vector space for the search engine to find relevant passages. These embeddings capture the meaning of words by placing semantically similar words close together in the vector space. Large pre-trained language models such as BERT, GPT-3 or MPNet can be used to create those embeddings.

Once the vector representations are created, the search engine can use similarity measures, such as cosine similarity[16] or dot product[17], to calculate the similarity between the query vector and the document vectors[18] and then return the most similar documents as the search results. While semantic search can be implemented with embeddings based on BERT or a similar language model alone, models exist which are specifically trained to embed whole sentences and search queries. Those more specific models generally perform better for tasks like semantic search.

In our case, we have chosen ***all-mpnet-base-v2***, which is based on the MPNet[19] model developed by Microsoft. This model was trained on "a large and diverse dataset of over 1 billion training pairs"[20]. We chose the model with the highest average performance according to an extensive evaluation of several models by the authors of

[15] https://www.sbert.net/examples/applications/semantic-search/README.html
[16] Cosine similarity is a measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them. By using cosine similarity, we can identify documents that are semantically similar or related, even if they contain different words or are structured differently. https://en.wikipedia.org/wiki/Cosine_similarity
[17] Dot product is an often method to compare document vectors or word embeddings. https://en.wikipedia.org/wiki/Dot_product
[18] A document vector is a mathematical representation of a document that captures its semantic meaning. It is a way to represent a piece of text as a numerical vector, where each element of the vector corresponds to a particular feature or characteristic of the document. One method for constructing document vectors is the "term frequency-inverse document frequency" (TF-IDF) weighting scheme, which assigns a weight to each term in a document based on how often it appears in the document compared to how often it appears in the entire collection of documents.
[19] Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, Tie-Yan Liu: "MPNet: Masked and Permuted Pre-training for Language Understanding", 2020; [http://arxiv.org/abs/2004.09297 arXiv:2004.09297]
[20] https://huggingface.co/sentence-transformers/all-mpnet-base-v2

the sentence-transformers library[21]. *all-mpnet-base-v2* outperforms other models like *distilroberta-v1* or *msmarco-bert-base-dot-v5* with regard to sentence embeddings and semantic search, while providing a decent encoding speed for development on consumer hardware. The results section discusses how the model selection could further improve the results.

We used the python framework *sentence-transformers*[22] to compute our sentence embeddings and *scipy*[23] for the cosine similarity between queries and sentences from the corpus. We then defined a threshold to only count the retrieved sentences which have a minimal distance to our input-query. Through experimentation we found that a threshold value of 0.65 tends to perform the best in terms of minimising the recall and maximising the precision.

Implementing a naïve baseline text classifier

A baseline system in the domain of machine learning is a simple approach that helps to qualify the performance of the model in question on a more global scale. Imagine a very complex model that predicts tornadoes in Switzerland for every day of the year. If this model has an accuracy of 99% it might seem impressive at first. However, since tornadoes are very rare in Switzerland, we could also design a very simple model that predicts no tornado for every day in Switzerland. If only five tornadoes occur in a given year in Switzerland, then this baseline system also an approximate accuracy of 99%, which would lead us to reassess the quality of the very complex model.

In our case, we decided to use a keyword search that is based on a term frequency-inverse document frequency (tf-idf) analysis of the experts' topical sentences. This approach is common in information retrieval and weights terms based on how often they occur in a specific document. If a term occurs more often than the weight applied is increased, as greater frequency is interpreted to mean the term has more significance in the document. If the term also occurs in many other documents, then the weight is decreased because the term seems to be more common in the whole corpora. For our classifier, we used all topical sentences generated by the experts as corpora and extracted bigrams (terms consisting of two words) with high tf-idf scores for each step in the management cycle. After that, all sentences in the reports that contained at least one of the search terms was marked as relevant for a topic.

We used common text transformation methods, such as lemmatization[24] of the words, to improve the performance of the classifier. However, we did not exhaust all possible

[21] https://www.sbert.net/docs/pretrained_models.html
[22] https://www.sbert.net/index.html
[23] https://scipy.org/
[24] Lemmatization is the process of reducing words to their base or root form, which is known as the lemma. The purpose of lemmatization is to normalize words so that they can be analysed and compared more easily. In natural language processing, lemmatization is often used as a pre-processing step before other text analysis techniques are applied.

improvement approaches. By only using the set of expert topical sentences, we used a suboptimal corpus to relativise the importance of the search terms, since the corpus is already very targeted towards the different topics we look for. We could improve this by using a more general corpus in combination with the topical sentences so that tf-idf detects terms important for the domain in general. In addition, project limitations did not allow to perform an extensive optimization process where we varied thresholds, such as the size of n-grams, to improve the results. But since the purpose behind a baseline classifier is to contrast the classifier in question with a very simple approach, we consider our approach to be sufficient.

Matching the parsed text data

As already described above in the chapter *The challenge of text extraction*, parsing the PDF files into truly machine-readable text data is not straightforward. Our experts used the reports in PDF format to identify the relevant passages because this is the best human readable format. This process resulted in two major challenges for the task of matching the parsed data with their selection.

First, when conducting NLP, the standard approach is to analyse text on a sentence level. However, our experts preferred to highlight whole paragraphs or even broader passages since sometimes the relevance of a sentence only becomes clear in the context of the surrounding sentences. Second, even when the highlighted passages corresponded to a paragraph or any other formatted form of text, the parsed data did not necessarily have to match the format in a suitable manner. For example, when a sentence is split up over two pages, then a human reader could easily comprehend that it still is one sentence. On a technical level, however, the text belongs to two different blocks. A block is the technical interpretation of a paragraph in a PDF file. Even though a block often corresponds to the human understanding of a paragraph, it can differ for many reasons, such as page breaks or special formatting. We could not achieve a one-to-one matching between the two different corpora, but we could achieve a reasonable convergence that is appropriate and proportional to the inaccuracies introduced by the circumstances of this matching.

We solved this problem by computing the Levenshtein distance between each sentence in the parsed data and the text highlighted by the experts. As soon as a sentence had a reasonably small distance, we flagged it as relevant. After that, we also flagged each block containing one of these relevant sentences.

We tried out two different libraries to parse the PDF files. The first library[25] analysed the source code of the files itself. This approach has the advantage that the tags and the technical layout of the source code itself can be used and all characters are passed without any errors. The disadvantage is that this only works for valid PDF files, not for

[25] https://github.com/pdfminer/pdfminer.six

e.g., pictures of PDFs. In addition, the technical layout does not have to correspond to the layout assumed by a human reader especially if a PDF is heavily stylized. The second library[26] used optical character recognition (OCR) for text recognition, where every PDF is treated like an image and the text extraction happens by optically interpreting the image. This approach has the advantage that even images of PDFs can be processed and the extracted blocks relate more to a human interpretation of the paragraphs. The disadvantage is that this approach requires a reinterpretation of the files' source code, which means that some characters or words might not be recognized correctly.

Which parsing approach works best depends heavily on the files themselves, the algorithms used, and many other factors. It can be argued that parsing the source code of the files yields good results because no interpretation errors occur. On the other hand, it can also be argued that a sprint with OCR produces more digestible and continuous text that has more value for NLP. In our case, we had to discover that the parsed text using OCR had by far bigger Levenshtein distances to the validation data than the text produced by interpreting the source code. The distances were so big that the matching produced a very sparse dataset. We therefore used the results from the source code interpretation for our evaluation of the models.

Results

The results section is split into three parts. First, we present the results of the model evaluation and assess the validity of the evaluation. Then we assess what parts of the process could be improved, what effort would be needed and how this would affect the quality of the results. In the third part, the domain experts interpret the findings and draw conclusions about the applicability of NLP for assessing sustainability reports. The code used to implement the classifiers and evaluate the results is available in the project repository[27].

Quantitative evaluation

Figure 4 shows the results of the evaluation. For each report there are two graphs, one for the performance of the baseline system and one for the performance of the sentence transformer-based text classifier. On the x-axis, each graph shows the different steps in the management process depicted as a topic. The y-axis shows the number of paragraphs (i.e., blocks) that were found for the specific topic. For every topic there are three bars, one for the number of validation paragraphs that belong to this topic (blue), one for the number of paragraphs found by the classification system (orange), and one bar showing the size of the overlap between the validation data and

[26] https://github.com/JaidedAI/EasyOCR
[27] https://github.com/planet-10/sentence-tranformer-based-text-classifier

the results of the classifier (green). The normalised performance metrics for recall, precision and F1 score[28] can be found in Annex 3.

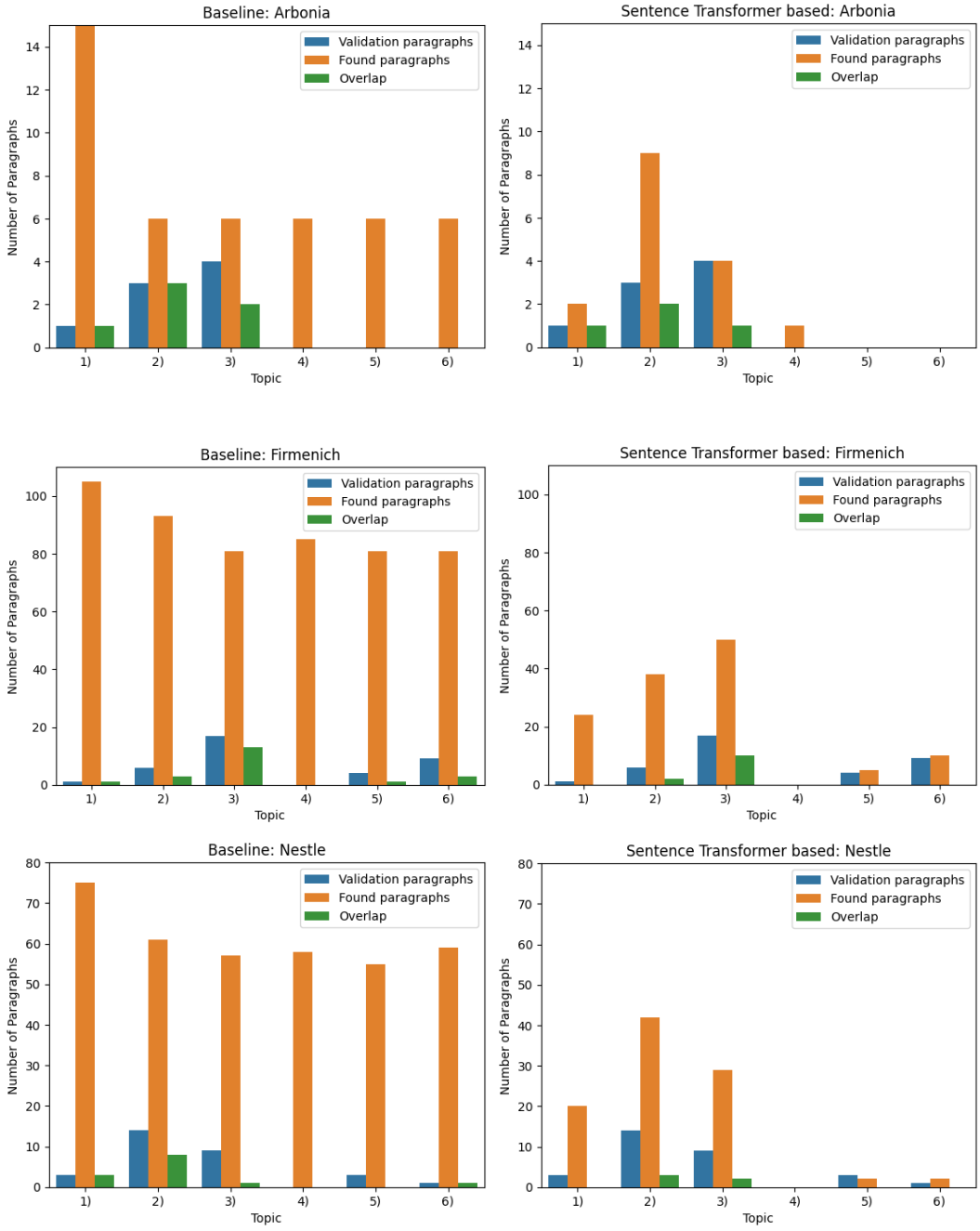

Figure 4: Results of the evaluation of the three reports (left column: Baseline results, right column: sentence transformer-based results)

[28] F1 score is a measure of the overall accuracy of a binary classification model that takes into account both precision and recall. It is the harmonic mean of precision and recall, with a value ranging from 0 to 1, where a higher value indicates better performance.

The different steps in the management process are represented by numbers for the sake of readability. The numbers stand for the following steps:

1. Materiality and risks
2. Policies or commitments regarding human rights
3. Actions to prevent or mitigate potential negative impacts on human rights
4. Actions to address or remedy actual negative impacts
5. Indicators and processes used to track or evaluate the effectiveness of the actions
6. Effectiveness of the actions, including progress toward the goals and targets

Recall

We first assess how many relevant passages the sentence transformers-based text classifier finds and assigns them to the respective step of the management cycle. In technical terms. The respective results are represented in the green bars in Figure 4. If the green bar has the same size as the blue bar, then the classifier found all passages that the domain experts deemed relevant. If the green bar is shorter, some important passages were missed.

When only looking at the recall, the naïve keyword search catches more relevant text passages for every step and report. The explanations for this result can be manyfold. One possible reason is that the queries devised by the domain experts are not suitable for usage with the sentence transformer-based system. This reasoning shows the difficulty of so-called “prompt design,” and the difficulty for human actors to estimate the effect of their input to a machine learning model on its output.

Another reason lies in the structure of the results, where the baseline approach generally (but not always) classified more passages as relevant, raising the probability of identifying a passage from the validation data. The following paragraphs elaborate further on this issue. Furthermore, the validation data contain very few sentences deemed relevant by the sustainability experts. Therefore, especially for topics five and six, the sentence transformer-based classifier persistently has a recall rate of zero percent because it missed the very few text passages (two and five, respectively) the evaluation data contained. Given our data, we cannot say whether the recall rate would improve or stay this low if we had more evaluation data.

Precision

The second quality assessed in this evaluation was described with the following sentence: *The text classifier finds the same passages that were also identified by the domain experts as relevant and — importantly — not different sentences*. This quality is well expressed through precision as a performance metric. The precision can be calculated by looking at the false negatives (passages missed by the classifier) and false positives (passages erroneously marked as relevant by the classifier). While the recall punishes false negatives, precision looks also at the false positives. In Figure 4, the false positives can be assessed by comparing the green bars to the orange bars.

The bigger the difference between the orange and the green bar, the more text passages were classified as relevant that could not be found in the validation data. When comparing the range of the y-axis between the baseline and transformer-based diagrams in Figure 4, it becomes clear that in all cases but one, the baseline approach marked distinctively more text passages as relevant compared to the sentence transformer-based classifier. Annex 4 shows that both classifiers never reach a performance score above 50 percent and that in eight of thirteen cases where there is validation data available for a topic, the baseline approach has higher precision.

The reasons for this behaviour can be manyfold. It could be that both classifiers might predict relevant text passages that were overlooked by the domain experts. In the case of the baseline system, the pitfall is quite likely that it is unaware of context. One of the search terms were the two words, human rights, which were used frequently in all three reports. However, in many cases the mere presence of the two words does not mean that meaningful statements relating to the management process are made. Similar to the analysis of recall, the biggest effect probably comes from the available validation data. While the low precision of the naïve keyword search is strongly influenced by false positives, the low precision of the sentence transformer-based classifier comes mainly from false negatives. In four cases the precision of the sentence transformer-based classifier is actually zero because the number of true positives is zero. However, in view of the sparsity of the validation data, this situation can happen if only a few of the relevant passages are missed. Hence, to increase the precision of the sentence transformer-based classifier, there need to be more true positives, meaning that the classifier must primarily detect more of the relevant text passages. This improvement could be achieved by tweaking thresholds or changing the topical sentences. However, with the available validation data it cannot be assessed ifincreasing the true positives would also lead to an increase in false positives (meaning the precision would stay low) or if the growth of false positives could be mitigated.

Robustness

One more interesting quality is the robustness of the classifier, which indicates how much the models' results vary depending on changes in the input. From this point of view, the sentence transformer-based text classifier comes again closer to the desired behaviour. While the number of found blocks was in all three cases somewhat relative to the number of identified blocks by the experts, the baseline classifier always produced a big number of results independent of the experts' indications.

This behaviour is not uncommon for keyword-based classifiers and could be mitigated by a more careful (maybe even human-supervised) selection of search terms. However, this result is also somewhat symbolic of the advantage of more complex text encoders, compared to the naïve baseline approach and other "bag of word" approaches.

Validity of the evaluation

This evaluation has multiple properties that limit the informative value and explanatory power of the results. Firstly, only three reports were assessed, which means that the number of data points is rather small. This small number of data points also made it superfluous to conduct further statistical analyses as the results would not have been of statistical significance. Secondly, the process of how the relevant passages were marked by the domain experts and other design decisions made it hard to implement a sound quantitative setup because technical workarounds had to be put in place to match the available data with a numeric framework. Thirdly, the performance of a classifier is not only the result of the statistical and technical methods underlying the model, but also a function of the effort and time put into preparing the data and tweaking the hyperparameters. The results of this evaluation should therefore not be used to rate the general capability of sentence transformer-based text classifiers, but rather provide a starting point to discuss how such classifiers could be used in the realm of sustainability reporting and what kind of resources would be needed to improve certain qualities.

Required effort for technical model improvements

Due to the exploratory nature of this project, the time spent tuning the sentence transformer-based text classifier was very limited. In that sense, we have only scratched the surface of applying sentence transformers to a corpus like the one evaluated in the scope of this project. While additional research is certainly needed, we already see several possibilities on how our *proof-of-concept* could be further enhanced.

To build a more sophisticated text classifier, our approach needs to be threefold: (1) expand the test and validation set, (2) define clearly in which direction to optimise, and (3) further develop and tune the underlying model.

(1) It is crucial for the tuning of any machine to have valid feedback loops. Hence, in order to improve the machine, a robust testing setup must be created, which in turn requires improvements to the whole testing process, including the test data. Further work is needed to improve the size and quality of the validation set to ensure that all reasonable sentences are recognised. These improvements will result in better performance analysis and more conclusive results.

(2) A central difficulty of this project is that it was exploratory at both the conceptual and technical levels in parallel. Thus, close collaboration with experts is continuously required, in order to identify more precisely the application cases for which a text classifier can add real value, and where the solution is not technical. Once the scope of the issue is framed more clearly, we can really tune the hyperparameters, which should allow us to greatly enhance the results.

(3) There are also several purely technical improvements to the classifier — starting with the model selection of the large-language model used for our sentence transformer. There are many other promising models that we could not test due to our limited time. We chose our model based on the average performance on sentence embeddings and semantic search for symmetric search. Depending on the prompts it could make sense to use a model optimised for asymmetric search, like one of the MS MARCO models suggested by the authors of the sentence-transformers library[29]. Analysing the effects of using different distance metrics instead of cosine-similarity.

    Since our corpus contains a lot of domain specific language, it would make sense to perform domain adaptive pre-training together with fine-tuning on labelled data to enhance the performance of our approach. Research on climate-change related texts have shown that such an approach can significantly enhance the capabilities of large language models[30].

We expect that with another iteration of a similar scale project, we could significantly improve the classifier and the validation system to draw more accurate and reliable conclusions about the potential of the technology itself. As mentioned, this new iteration must be done in close collaboration with the domain experts and would also allow for further refinement of the conceptual approach. As for the pre-training of a model for the specific domain of our corpus, estimates made during this project show that in order to reach the current state-of-the-art in the field, more resource-intensive projects are needed.

Assessment on applicability of NLP for sustainability reporting

Our efforts show that the applicability of NLP to evaluating sustainability reporting is influenced and likely limited by two important factors:

As mentioned in the chapter on the challenges at hand, there is a huge variety in any corpus consisting of a collection of current real-life sustainability reports, both in form (pictures, tabular data, charts, text) as well as topical content (environment vs. social issues, own operations vs. business partners, focus on positive stories vs. challenges). Given this diversity, correctly evaluating sustainability reporting manually is likely to be very labour intensive. We therefore believe that using NLP to assist human evaluators is a promising endeavour.

[29] https://www.sbert.net/examples/applications/semantic-search/README.html

[30] Nicolas Webersinke, Mathias Kraus, Julia Anna Bingler, Markus Leippold: “ClimateBert: A Pretrained Language Model for Climate-Related Text”, 2021; [http://arxiv.org/abs/2110.12010 arXiv:2110.12010].

Yet algorithms that perform automated assessments will have to deal with many of the same difficulties as human evaluators. Therefore, results of the automated assessments need to be analysed carefully, especially concerning the possible underrepresentation of sentences that are relevant, but not picked up by the algorithm due to the diversity of the corpus.

On the other hand, sustainability reporting is also marred by a lack of clarity on what is considered to be "good". And it is hard to come by universally accepted definitions that are valid across industries, sectors, sizes of companies, degrees of exposure to fundamental risks, etc. The fact that the law only very generally describes what is expected of companies in terms of sustainability reporting means that there is no authoritative benchmark against which an evaluation can be made. Hence, domain expert-defined criteria will inevitably vary and may involve some level of bias. Algorithms trained on that basis will carry the same limitations. This suggests that NLP can play a supportive role in identifying the critical text segments in sustainability reports, but human assessment based on transparent evaluation criteria needs to be involved.

Benefits of NLP

Our efforts showed that despite these limitations, there are tasks that NLP is potentially well-suited to assist humans with (see chapter on conceptual approach). While fully assessing a report's compliance with the law may not be an appropriate task for NLP at this stage, NLP can reduce the amount of human effort needed for this task by assisting with categorising reports according to their scope and depth, which may help humans to quickly focus on those cases where further human investigation is necessary. Similarly, highlighting areas of reports that contain content likely to be relevant for assessing the report can speed up the task.
Hence, combining "categorisation" and "highlighting" tasks seems to generate the most value for the specific task at hand for the time being.

Comparison of results from baseline and NLP analysis

Assessing the results from a qualitative perspective is particularly important to understand the relevance of the additional hits by the baseline approach and how this compares to the sentence transformer-based text classifier results.

Using bag-of-words creates significant amounts of false positives (25%, in extreme cases up to 65%) that have nothing to do with human rights, but instead with climate change or other topics. This problem might be driven by keywords like 'minerals' or 'resources', which are also used in contexts unrelated to human rights. Other false positives are related to human rights but not to the respective stage of the management cycle. And some of the latter categories are just keywords that appear in a long list of sustainability topics not related to human rights. The additional true positives that the baseline approach finds come with considerably more noise in the baseline results than the results from the text classifier.

The sentence transformer-based text classifier has no hits outside the human rights domain. It avoids those cases where human rights keywords are just mentioned in a longer list on unrelated other topics. For the false positives that are related to human rights, all that were found by the text classifier were also tagged by the baseline approach, but the baseline approach returned many more false positives.

A few false positive cases were found in both approaches that could also be counted as true positives, although the information content was in all cases very limited which is why they were not included in the validation set.

The results further show that there is a second problem, which is the allocation of human rights related hits to the respective step of the management cycle. The false positives do not show a pattern, i.e., they are not all related to a particular step, but seem to be randomly distributed across the management cycle.

Recommendations and outlook

As the role of non-financial reporting in sustainability governance grows, the number of companies reporting and the scope of reporting can also be expected to grow.

The findings of this limited study indicates that NLP has the potential to assist in — not replace — the evaluation of sustainability reports by human experts. Analysing scope and depth of reporting along the management cycle provides important information about a company's reporting effort and comprehensiveness.

The formulation of recommendations for next steps depends critically on how we assess the outcomes of this study, in particular the recall and precision performance of the baseline versus the sentence transformer-based classifier approach. In short: should a higher recall be preferred over precision or rather the other way round? The answer involves a judgement call. The higher recall comes with false positives, which can be significant in number, as the test results show. The number of false positives is lower with the text classifier approach, but it also returns a higher number of false negatives (i.e., not catching up all statements that expert analysis linked to the different steps of the management cycle).

Either way, both are not sufficiently accurate and need further development. The question then is: should we bet on improving basic search strategy, the bag-of-words-approach, or invest in advancing the transformer-based classifier algorithm? Which strategy will lead to better results that make the algorithm most useful in the long term?

Considering the search challenges inherent in evaluating the reports, we recommend focusing on the sentence transformer-based text classifier approach because of its semantic sensitivity. While both the transformer-based text classifier and the baseline approach could identify statements related to human rights with some degree of improvement, the greater challenge is determining where to allocate a statement along the management cycle. There are many more semantic nuances that make the difference between steps which the current text classifier approach seems to pick up slightly more accurately. Since the semantic sensitivity of the text classifier is what distinguishes it from 'bag of words', the chances are higher to advance to a point where we can further disentangle human rights related sentences according to the step of the management cycle they relate to.

While efforts are underway for NLP-assisted analysis of environmental reporting, especially on climate change, this is not yet the case for the social dimensions of sustainability. ClimateBERT has been trained specifically targeting the language used in environmental reporting and the results obtained are very promising. Therefore, we suggest that future efforts continue to focus on the social dimensions of sustainability, starting with human rights. Following the example of ClimateBERT, a next step could be to build ethicaLM, a model trained specifically to target the language used to report on social dimensions of sustainability.

At the technical level, the goal of the next step is to confirm the preliminary findings of this analysis and advance the algorithm's robustness. In essence, the proof-of-concept needs to be replicated at a larger scale, including the

- review of topical sentences by a larger pool of experts from academia and practice;
- iterations between the pool of experts and technical team around the topical sentences;
- further research and consolidation of the underlying technical approach;
- creation of a significantly larger pool of validation data;
- improvement of the evaluation process based on a refined application case.

At the political level the following recommendations emerge from this analysis:

First, efforts should be undertaken to request companies to adhere to more formal and structured analysis as is the case for the EU directive and the German Lieferkettengesetz. This reform will be an important way to increase the consistency of the input data and hence improve the quality of NLP-assisted evaluation.

Second, the body of sustainability reports to be analysed needs to be clarified. A registry containing the firms required to report each year could avoid ambiguity. In addition, sustainability reports should be completed using one template document to avoid ambiguity about what to include in an assessment.

Annexes

Annex 1: Sustainability reporting requirements as defined in the code of obligations.

Scope A

Under scope A of the regulation fall companies with more than 500 employees AND annually 40 million turnover in two subsequent years.

Duties for companies falling under scope A:

- Reporting on the environment, in particular CO2 emissions, social issues and worker issues, which are both not specified in detail; human rights and corruption;
- Reporting on strategy and due diligence procedures, measures taken and impact thereof, business risks emerging from those non-financial topics as described in OR Art. 964ter.

Scope B

Under scope B of the regulation fall companies that meet the following criteria:

- Companies that import and process at least one mineral from a defined list of minerals and which import above a certain threshold volume
- Exempt are companies that report adherence to either the OECD Due Diligence Guidance for Responsible Supply Chains of Minerals from Conflict-Affected and High-Risk Areas OR the EU ordinance (EU) 2017/821 on supply chain due diligence

Duties for companies falling under scope B:

- Document whether minerals are sourced from a conflict- or high-risk area. For companies sourcing form conflict- or high-risk areas, additional duties arise:
- Develop a supply chain policy to identify, assess, eliminate or mitigate the risk of potential negative externalities from sourcing minerals from conflict- or high-risk areas;
- Report on supply chain policy and the company's compliance with it;
- Establish a mechanism to report concerns related to the origin of the minerals and metals along the supply chain;

- Report on the outcome of the risk assessment, the assessment of the risks identified, measures taken to eliminate or mitigate those risks, and the results thereof
- Establish a traceability system along the supply chain

Scope C

Under scope C of the regulation fall companies that meet the following criteria:

- All companies except SMEs
- Exempt are companies that
    - have a low risk of child labour, defined as companies sourcing from countries classified by UNICEF Children's Rights in the Workplace Index as a "basic" due diligence response.
    - declare adherence to both ILO Agreement 138 and 182, and the ILO-IOE Child Labour Guidance Tool for Business AND either the OECD Due Diligence Guidance for Responsible Business Conduct or the UN Guiding Principles on Business and Human Rights

Duties for companies falling under scope C:

- Investigate and document whether there is a reasonable suspicion of child labour and document outcome of the investigation; For companies sourcing from a country with a high risk of child labor, additional duties arise:
- Develop supply chain policy to identify, assess, eliminate or mitigate the risk of child labor;
- Report on supply chain policy and the company's compliance with it;
- Establish a mechanism to report on child labor concerns along the supply chain;
- Report on the outcome of the risk assessment, the assessment of the risks identified, measures taken to eliminate or mitigate those risks, and the results thereof
- Establish a traceability system along the supply chain

Annex 2: List of key terms and topical sentences for human rights

A) <placeholder> values:

1. Human rights
2. Rights of Indigenous peoples
3. Rights of persons with disabilities,
4. Rights of migrant workers
5. Children's rights
6. Labour rights,
7. Workers' rights,
8. Right to organize,
9. Right to collective bargaining
10. Political rights
11. Freedom of assembly and association,
12. Right to protest

B) Queries for the six phases in the Management process

1) Materiality and risks

Example sentences

1. The sourcing of raw materials **may be linked to** adverse impacts on <placeholder>.;
2. Business activities in a conflict zone **may include violation** of <placeholder>.;
3. **We operate in a sector with known** <placeholder> **issues**.;
4. **Particular attention is required when we work in a country where** <placeholder> are not guaranteed.;
5. **We understand our exposure** to <placeholder> violations in our supply chain.;
6. The a**ctivities of our suppliers can have a significant impact on** local or indigenous communities, minorities, or vulnerable groups.;
7. **Heightened attention is required when a supplier operates in** a jurisdiction that experiences political instability, weak governance, or repression of minority groups.;
8. The **activities of our suppliers can have a significant impact on** worker's rights in jurisdictions with weak labour regulation and limited freedom of association.;
9. **We may potentially be exposed to** conflict minerals.;
10. **We may potentially be exposed to** child labour.;
11. **We may potentially be exposed to** forced labour;
12. **We may potentially be exposed to** modern slavery;
13. **We may potentially be exposed to** human trafficking;

2) Policies or commitments and ambitions regarding human rights

Example sentences

1. **We are committed to** <placeholder> and respect them as a key element of responsible business conduct.;
2. **We consider** <placeholder> issues in our supply chain;
3. **We recognize our responsibilities** as an employer in the area of <placeholder>.;
4. **We consider our responsibilities** in the area of <placeholder> as fundamental to how we do business.;
5. Upholding <placeholder> **is our key concern**.;
6. **We strive to assume our responsibilities** in accordance with the International Bill of human rights;
7. **We strive to assume responsibility** in accordance with the principles on human rights,, the UN Global Compact as well as conventions of the International Labour Organization.;
8. **We commit ourselves to** the United Nation's Protect, Respect and Remedy Framework.;
9. **We aim to** eliminate child labor and we aim to pay a living wage to all workers in our supply chain.;
10. **We will ensure that** labour rights and the right to unionise are respected.;
11. **We seek to play our part** in eliminating forced labour.;
12. **We seek to play our part** in eliminating child labour.;
13. **By 20# we will have** no child labour in our supply chain.;
14. **By 20# we will have** no forced labour in our supply chain.;
15. **By 20# we will pay** living wages to all workers in our supply chain.;

3) Actions to prevent or mitigate potential negative impacts on <placeholder>

Example sentences

1. **We have established** a structured due diligence assessments of our operations in place to identify where <placeholder> risks may exist.;
2. **We have established** a structured due diligence assessments of our suppliers in place to identify where <placeholder> risks may exist.;
3. **We conduct regular** <placeholder> **reviews** of our operations and suppliers;
4. **We conduct regular** <placeholder> impact assessments of specific operations.;
5. **We conduct regular training** of our employees on <placeholder>;
6. **We conduct regular training** of our suppliers on <placeholder>;
7. **Our actions include training** on <placeholder> .;
8. **We make** <placeholder> **part or the contract** with our suppliers.;
9. **We demand actions from suppliers** to reduce child labour.;
10. **We launched** an action plan on <placeholder>.;
11. **We have improved our system to detect** violation of <placeholder>;

12. **We make** living wages **part or the contract** with our suppliers.;

4) Actions to address or remedy actual negative impacts

Example sentences

1. We have an independent **grievance mechanism in place** that is available to employees, local communities and other stakeholders that are affected from our business operations.;
2. We have an independent **whistle-blower mechanism in place** for employees and other workers, local community and civil society members.;
3. Employees and affected communities and **stakeholders can file a complaint** with the office of the independent ombudsperson;
4. A committee of external experts **reviews all complaints** received;
5. Our independent board of experts **decides on corrective actions to remedy** <placeholder> complaints.;

5) Indicators and processes used to track or evaluate the effectiveness of the actions

Example sentences

1.
2. **Our monitoring system measures compliance with** <placeholder>;
3. **We regularly measure progress on** <placeholder>;
4. We have established a system of robust **indicators to monitor performance** in <placeholder> issues.;
5. **Our indicators to monitor** <placeholder> **are …;**
6. We carry out visits to **verify progress on** <placeholder>.;
7. All operations are subject to **regular performance review of** <placeholder> performance.;

6) Effectiveness of the actions, including progress toward the goals and targets

Example sentences

1. **We have trained** <#>percent of our suppliers on <placeholder> issues.;
2. The total number of <placeholder> incidents is <#> during the reporting period;
3. The **share of workers receiving living wages** is <#>.;
4. The total **number of children not engaging in** child labor is <#>.;
5. The **number of child labour cases** has been reduced by <#> during the reporting period;
6. The **number of forced labour cases** has been reduced by <#> during the reporting period;
7. <#> grievances related to <placeholder> **were filed during the reporting period.**;

8. <#> complaints related to <placeholder> **were filed during the reporting period**.;
9. **We have reviewed and addressed** <#> grievances and complaints related to <placeholder> during the reporting period.;
10. **We have resolved** <#> grievances / complaints related to <placeholder> during the reporting period.;
11. **We have resolved by mediation** <#> <placeholder> grievances during the reporting period.;

Annex 3: Validation data

The PDF of the reports are available in the project repository[31].

Nestlé 2021[32]

1) Materiality and risks

Human rights* play a key role in enabling *a just transition to regenerative food systems.
The path to regenerative agriculture is a long-term journey with challenges. This is why we will help farmers by offering investment, rewarding good practices and offering technical and scientific guidance. This, together with the respect and promotion of human rights, will contribute to a just transition to regenerative food systems.

By respecting and advancing human rights in our value chain, we are building a foundation that contributes to a resilient future for our planet and its people.

Human rights are inextricably linked to our shared future. By respecting and advancing them in our value chain, we are building a foundation that contributes to a resilient future for our planet and its people.

Our salient issues are those human rights at risk of the most severe negative impact on people through our activities or business relationships. By the end of 2022, we will develop
and publish a dedicated action plan for each of our salient issues. These will articulate our strategy for assessing, addressing and reporting on each salient issue, defining what we need to do across our value chain, as well as what collective action can be taken.

2) Commitments, policies and strategies regarding human rights

People and respect for human rights are at the core of Nestlé's culture and values. **We are committed to *raising awareness*, *promoting best practices and empowering people across our own operations and supply chains.***

We were early adopters of frameworks like the United Nations Guiding Principles on Business and Human Rights and the Organisation for Economic Co-operation and Development Guidelines for Multinational Enterprises. At the same time, we piloted many programs to assess and address risks on the ground.

Our commitment to respecting and promoting human rights is a key part of advancing regenerative food systems at scale, which is focused on transforming farming practices at the heart of the food systems while enabling a just and equitable transition. We aim to use our scale, experience and resources to contribute to this vision.

In December 2021, **we released our new Human Rights Framework and Roadmap**. Through implementing this framework, and with powerful collaborations, **we will enhance** due diligence and develop action plans to address our most salient human rights issues.

By 2022 year-end, **we will publicly launch action plans** for each of our 10 salient issues, and report our progress against them by 2025.

Forest Positive means moving beyond just managing deforestation risks in our supply chain to targeting a positive impact on our broader sourcing landscapes. **Our strategy aims to help** conserve and restore the

[31] https://github.com/planet-10/sentence-tranformer-based-text-classifier/tree/main/00_data/reports
[32] https://github.com/planet-10/sentence-tranformer-based-text-classifier/blob/main/00_data/reports/Nestle_creating-shared-value-sustainability-report-2021-en.pdf

world's forests and natural ecosystems while promoting sustainable livelihoods and respecting human rights, including empowering Indigenous Peoples and Local Communities to be stewards of critical natural ecosystems.

Understanding the drivers of deforestation and creating the right incentives for forest conservation and the preservation of natural ecosystems are key to our approach. This is why **we will go beyond our supply chain**. **Our actions will include** rewarding suppliers for practices that keep trees standing, regenerate the land and respect human rights.

We will ensure proactive action to help keep forests standing and restore degraded forests and natural ecosystems while respecting the rights of Indigenous Peoples and Local Communities.

The conservation and restoration of forests and other key natural ecosystems forms part of our Net Zero Roadmap. Sustainable livelihoods and respecting human rights are part of our Human Rights Framework and Roadmap.

Sustainable production, respect for human rights and investing in women and youth **are at the core of** Nestlé's activities to help boost rural development and livelihoods and strengthen communities. **We seek to play our part in tackling** child labor risks, improving animal welfare, increasing farmer incomes and investing in the next generation. From enabling access to education for children, farmers and communities, to investing in local infrastructure, working with partners to map supply chains and provide raw material certifications, **we use the many tools at our disposal *to support communities and help them thrive.***

LGBTQ+ community Nestlé has expressed support for the United Nations Standards of Conduct for Business on tackling discrimination against LGBTI people. In addition, **we are proud to be part of** the Partnership for Global LGBTI Equality, the only LGBTQ focused organization in the world where the private sector and civil society sit together as members, to accelerate equity, social and economic inclusion for the LGBTQ+ community. The Partnership is an initiative of Business for Social Responsibility, the Office of the United Nations High Commissioner for Human Rights and the World Economic Forum.

Our board-level Sustainability Committee **aims to ensure** that we carry out due diligence and report on our most severe risks to human rights, while our ESG and Sustainability Council manages salient issues (see right) in the upstream supply chain. It is supported by the work of the Human Rights Community, gathering more than 20 people from different functions with human rights responsibilities.

Our long-term Forest Positive strategy, announced in 2021, is helping us to find ways to integrate further protection for tenure-based rights for Indigenous People and Local Communities into our approach, while at the same time helping smallholder farmers to develop sustainable livelihoods.

3) Actions to prevent or mitigate potential negative impacts on human rights

In particular, ***the Sustainability Committee* reviews our plans and actions with regard to** climate change, plastics and packaging, water management and responsible sourcing, while ensuring that Nestlé carries out human rights due diligence and manages diversity, inclusion and employee health and well-being appropriately.

Stakeholder engagement and partnerships have long been an important part of our strategy. **We partner (and have partnered) with** a wide range of organizations on human rights issues, such as the Danish Institute of Human Rights, the Fair Labor Association and the International Cocoa Initiative, among many others. Our CARE program monitors internal human rights compliance at Nestlé facilities through external audits.

We worked with our suppliers and partners to develop time-bound action plans to address the gaps found and supported suppliers, mills, plantations and smallholders in our supply chain to address specific labor rights risks such as forced labor and child labor, through targeted interventions.

In 2021, we improved our grievance mechanism by integrating our former Integrity Reporting System (for employees) and our external platform (for all other stakeholders) into an independently operated system called 'Speak Up'.

In June 2021, our UK and Australian markets **worked together to *produce their first joint Modern Slavery and Human Trafficking report to address the requirements of their countries' modern slavery acts.*** This demonstrates the collaboration and consistency of our coordinated global approach.

Promoting human rights in agricultural supply chains **Our efforts to source sustainably have enabled us** to make important progress in promoting human rights in agricultural supply chains.
In 2021, **we launched a detailed labor rights action plan for palm oil.** We are working toward a palm oil supply chain where all workers, at all tiers of production, work and live in safe and healthy conditions, are provided contracts detailing their working conditions, are paid fairly, have the right to associate freely and collectively bargain and have access to grievance mechanisms.

Child labor risks and access to education
We were the first company in the cocoa sector to **introduce a Child Labor Monitoring and Remediation System (CLMRS),** and many companies have now adopted it as a leading tool that helps tackle child labor risks by working directly with communities on the ground. **Our CLMRS prioritizes** access to education, including building and renovating schools and securing birth certificates for registration, and tackling rural poverty through income diversification programs and support.
The CLMRS is a six-step process that starts with raising awareness. Community Liaison People visit farmers and cooperatives, and based on visits and surveys, identify children at risk. Families of children identified receive further visits where they are advised and supported by the Community Liason People. **Regular follow-up visits allow us** to measure how many children have been prevented from entering child labor or have stopped doing hazardous work. Each year, we identify some children in our supply chain who are at risk of engaging in child labor. **We carry out follow-up visits** with each of these children and record the number who report that they are no longer at risk during two consecutive visits. In 2021, the number of children who reported no longer being at risk at the two most recent visits was 6307 in Côte d'Ivoire and 738 in Ghana.

Collaborating to reimagine fairer food systems
In 2021, **Nestlé partnered with Tufts University to convene** a UN Food Systems Summit dialogue with stakeholders to discuss the nexus between regenerative food systems and the right to food.

In total, 57 participants from academia, non-governmental organizations, the private sector and UN organizations discussed the major barriers and corresponding levers to making healthy diets affordable, accessible and adequate for everyone, including the responsibilities of different stakeholders in ensuring access to safe and nutritious food for all, collectively moving toward the 2030 SDGs.

In January 2022, to expand our work to tackle poverty as a root cause of child labor risks, **we launched a novel approach that aims to** support farmers and their families in the transition to more sustainable cocoa farming. The Income Accelerator Program **will pay cash incentives** directly to farming families for activities such as school enrollment, sustainable agricultural practices, agroforestry and income diversification. **The incentives will encourage** behaviours and agricultural practices that are designed to steadily build social and economic resilience over time. These incentives are paid on top of the premium introduced by the governments of Côte d'Ivoire and Ghana that Nestlé pays and the premiums Nestlé offers for Rainforest Alliance certified cocoa. The payments are not linked to production volumes and reward cocoa-farming families for the benefits they provide to the environment and local communities.

4) Actions to address or remedy actual negative impacts

P57 Table with reports on non-compliance

Non-compliance concerns raised through Speak Up by category

Breakdown categories for Speak Up messages	Messages received	Messages substantiated
Abuse of power and/or mobbing/bullying	567	188
Unfair treatment	386	97
Labor practice	373	77
Safety and health 1	56	29
Fraud (misappropriation or misconduct on accounting/financial statement)	137	28
Harassment (excluding sexual harassment)	108	37
Third-party compliance	94	17
Gifts, families and relatives, conflicts of interest	91	8
Violation of laws/regulations	91	25
Violence and discrimination	78	19
Seeking compliance advice	63	10
Sexual harassment	59	22
Bribery and corruption	55	4
Confidential information, Privacy Policy (data privacy, trade secrets, intellectual property)	49	11
Human rights (child labor, forced labor and modern slavery risks)	49	1
Environmental impact	43	3
Antitrust and fair dealing	33	1
MANCOM members related	17	2
Non compliance with WHO Code	11	1
Trade sanctions	10	0
Executive Board member/senior managers in Switzerland	5	0

5) Targets, indicators, and processes used to track or evaluate the effectiveness of the actions

P40:
The **minimum criteria to define if a raw material is produced sustainably are**:
– Traceable back to the point of origin (farm or group of farms)
– Human rights and environmental due-diligence systems are in place to assess, address and report on the potential or actual impacts in the supply chain
– The tier-1 supplier is measurably progressing in addressing actual or potential human rights and environmental impacts identified in its supply chain, as well as animal welfare where applicable

Regular follow-up visits allow us to measure how many children have been prevented from entering child labor or have stopped doing hazardous work. Each year, we identify some children in our supply chain who are at risk of engaging in child labor. **We carry out follow-up visits with each of these children and record the number who report** that they are no longer at risk during two consecutive visits.

6) Effectiveness of the actions, including progress toward the goals and targets

57892 Employees **trained on** human rights in 2021

After launching mandatory human rights training for all employees, we identified in 2020 a handful of countries with gaps in terms of the number of employees trained. These were mainly low-risk countries with a substantial number of factory workers with no computer access and where in-person training was made difficult because of COVID-19 restrictions. By the end of 2021, **we closed this gap** in most countries. In addition, it is part of the mandatory training for all new employees, which will ensure that all future employees are trained.

Côte d'Ivoire
156974
Cumulative total number of children who have received support (127550 in 2020)
6307
Number of children identified who reported no longer engaging in child labor at the two most recent follow-up visits (4838 in 2020)

Ghana
2809
Cumulative total number of children who have received support (2399 in 2020)
738
Number of children identified who reported no longer engaging in child labor at the two most recent follow-up visits (693 in 2020)

85
Company and supplier representatives in Turkey's hazelnut supply chain **received training in** labor rights issues in 2021

Firmenich 2021[33]

1) Materiality and risks

Due to the nature of our business, Firmenich's exposure to conflict minerals is indirect and very limited. **We may potentially be exposed to** "conflict minerals" through the use of catalysts in the manufacturing process of our products. We conduct due diligence checks to find out the origin of the relevant materials and ensure

[33] https://github.com/planet-10/sentence-tranformer-based-text-classifier/blob/main/00_data/reports/Firmenich_KOZQjRz6ornoYorSa601gLSGvA830xU4JD2UzZghMew.pdf

traceability through the following internal procedures: supplier qualification and raw materials introduction process.

2) Commitments, policies and strategies regarding human rights

As members of the UN Global Compact (UNGC) since 2008, and a UN Global Compact LEAD company since 2019, **we continue to be guided also by** the UNGC's Ten Principles in the areas of human rights, labor, the environment and anti-corruption.

We published **our ESG Ambitions** 2030 in FY21, working across the ESG spectrum to set our strategy for the decade ahead. **Our ambition is** to be #1 in renewable ingredients, in conscious perfumery, and in dietary transformation. **We have mapped out this transformational journey with long-term goals and mid-term targets** in three streams: Acting on Climate; Embracing Nature; and Caring about People. They include the boldest carbon emissions commitment in our industry: carbon neutrality by 2025 and a carbon positive impact beyond that date. By 2030, we will have achieved absolute carbon emission reduction in line with the 1.5°C Science-Based Targets; as well as pace-setting goals, ranging from water use and regenerative agriculture to human rights and equal pay.

Human rights **are one of our key concerns,** from ensuring health and safety during a pandemic to standing up for social justice. **Firmenich strives to** protect individuals and reduce inequalities. **We demand the highest human rights standards** in our business and our supply chain.

Our Human Rights Policy **outlines our commitments and expectations from** colleagues and suppliers while encouraging our business partners to follow similar principles. It complements our Code of Ethics and our Responsible Sourcing Policy, which states what we expect from all business partners.

Firmenich is adamant that materials and services be procured from reputable suppliers who are aligned with the Firmenich Code of Ethics (CoE) and Responsible Sourcing Policy (RSP), stemming from our commitment to operate in the most ethical, traceable and
responsible supply chain.

As a responsible company, we are committed to operate within a responsible supply chain, respect and support human rights as evidenced by our Human Rights Policy Statement, our Code of Ethics and Business as well as our Responsible sourcing policy. **We are also committed to support our customers to comply** with their reporting requirements related to their value chain exposures.

3) Actions to prevent or mitigate potential negative impacts on human rights

It is important to note that our Responsible Sourcing due diligence at material level does not only focus on biodiversity risk but also includes human rights topics.

Train 100 major suppliers on human rights and responsible sourcing
10 new initiatives at source including focus on women empowerment, education, human rights practices and living wages

Conducting regular human rights due diligence is central to our approach.

Our human rights impacts are independently assessed through SMETA audits by Sedex, EcoVadis questionnaires, and the Union for Ethical BioTrade field audits..

Our human rights-based approach is managed by a transversal Executive Human Rights Committee, whose membership includes our Chief Human Resources Officer; Senior Vice President Quality, Health, Safety and Environment; Senior Vice President Human Resources for Ingredients and Research; as well as the General Manager of Firmenich Geneva; Senior Vice President Legal Counsel; Senior Vice President Supply Chain; Vice President Business

Ethics; and led by the Chief Sustainability Officer. The committee meets monthly to review progress on our Human Rights action plan, review any new policy requirements in Switzerland, in the EU and beyond and take necessary decisions to meet the Company's ESG ambitions.

Right after the launch of our ESG goals, we began our Human Rights Training program, following its review and approval by the Firmenich Executive Committee on Human Rights.

Our **training and learning approach** on human rights covers five key areas:
1. Training of human rights coaches
These coaches will become reference colleagues for any colleague who has a human rights-related question, issue or project. The coaches have been trained by the UN Global Compact Network Switzerland through a 15-hour online program.
2. **Human rights training f**or procurement
Through the Firmenich Sustainability Academy, we produced a 30-minute training video covering all procurement functions. Adapted from a training module designed by the UN Global Compact Academy, the video explains the role a procurement team can play in improving working conditions in global supply chains.

3. Human rights for managers
This course is planned to launch in FY22

4. **Human rights executive coaching**
In the past three years, Firmenich has been working with a leading human rights expert who participated in the drafting of the UN Guiding Principles on Human Rights. This expert participates in all our committee meetings and offers **coaching and insight to the Executive Committee on Human Rights on a monthly basis**.

5. **Human rights as part of other trainings**
Besides targeted human rights trainings, other available **trainings addressing human rights includes:** training on the United Nations Development Goals; safety trainings; procurement trainings; training on policies; training on biases; and leadership development programs promoting inclusive behavior.

11000 **hours of human rights training**

100% of our production sites are regularly SMETA audited, and their reports are shared on the platform with over 100 customers. We decided to leverage Sedex platform in our human rights due diligence approach because it includes material human rights dimensions,

To address human rights with our suppliers, **our strategy is** first to raise awareness and train the entire procurement community on human rights, including on the UN Human Rights Guiding Principles for Business, as well as on emerging human rights laws and management practices. These efforts include a strong focus on the roles of EcoVadis and the Union For Ethical Biotrade (UEBT) and their evaluation of our supply chains. Both organizations' stan-
dards pay great attention to human rights, with one entire EcoVadis pillar devoted to assessing a company's record of accomplishment on human rights.

› **Our new Supplier Expectations Manual**, published in April 2021, is fully aligned with our Responsible Sourcing Policy and includes a series of requirements and expectations based on ethics, human rights and labor rights. Our Chief Procurement Officer has mandated that the entire procurement teams attend a new e-training module developed by the United Nations Global Compact: "How procurement decisions can advance decent work in supply chains." This module provides guidance on engaging with suppliers on executive Human Rights Committee

Where a catalyst uses a material defined as "conflict minerals", our Global Regulatory Services review the supplier's certificate(s) to guarantee compliance with the applicable regulations. As it is the case for any other material sourced by Firmenich, if a supplier does not satisfy the requirements, we reserve the right to look for alternative sources and/or substitute the raw material with a different one.

4) Actions to address or remedy actual negative impacts

5) Targets, indicators, and processes used to track or evaluate the effectiveness of the actions

Firmenich ESG **ambitions** …
Zero human rights non compliance in our operations
No gender Pay Gap - No Ethnic Pay Gap
+50% of Senior Leaders are diverse
100% Living Wage in our operations

By 2030, we aim to raise the global score of our suppliers on the Labor & Human Rights pillar to 60 out of 100 points total.

NO DISCRIMINATION IS PRACTICED

No discrimination in hiring, compensation, access to training, promotion, termination or retirement based on race, caste, national origin, religion, age, disability, gender, marital status, sexual orientation, union membership or political affiliation

There shall be no new recruitment of child labor

CHILD LABOR SHALL NOT BE USED
Companies shall develop or participate in and contribute to policies and programs which provide for the transition of any child found to be performing child labor to enable her or him to attend and remain in quality education until no longer a child

Children and young persons under 18 **shall not be employed** at night or in hazardous conditions

NO HARSH OR INHUMANE TREATMENT IS ALLOWED

Physical abuse or discipline, the threat of physical abuse, sexual or other harassment and verbal abuse or other forms of intimidation **shall be prohibited**

Companies **should provide access to a confidential grievance mechanism** for all workers

6) Effectiveness of the actions, including progress toward the goals and targets

ESG **performance highlights** …… Zero human rights non-compliance in our operations (p10)

P70-71: Relevant visual representation of whether the company is on track to meet its 2025 goals – how to capture this?

In FY21, SMETA **audits conducted in Firmenich production sites have identified** zero non-compliance on the three human rights dimensions:
> No discrimination
> No child labor
> No harsh or inhumane treatment

In our first year of tracking this data, we have now 73% of our suppliers reporting to EcoVadis and 38% of the assessed supplier spend received a silver rating or higher. In addition, our suppliers' current average score on the Labor & Human Rights pillar is 52 out of 100 points total.

P138 and 142:
Collective bargaining figures # union # non union % of union employees

LGBTI COLLEAGUES

In September 2020, Firmenich SA **received the Swiss LGBTI-Label for** our inclusive organizational culture for LGBTI people in Switzerland. Following a comprehensive assessment of the effectiveness of our policies, actions and communication, Firmenich SA became one of the first four companies and institutions in French-speaking western Switzerland to be **awarded the certification.**
With the Swiss LGBTI-Label, **we were commended for** being a role model through our open public commitments, including by signing the United Nations LGBTI Standards of Conduct for Business in 2019, and for our participation in the UN General Assembly panel discussion held in New York on advancing LGBTI rights.

Arbonia 2021[34]

1) Materiality and risks

The topic of procurement and supply chain at Arbonia comprises on the one hand the procurement management for the most commonly used materials and semi-finished goods – in other words, wood, steel, glass and aluminium. On the other hand, **the assessment of suppliers according to ecological and social criteria is also a part of it. In this context, respect for human rights in the supply chain is of vital importance.** Arbonia also pays attention to maximum sustainability with its suppliers. **Since more than 96% of used materials are sourced from suppliers in Europe, a high standard is already enshrined in law.**

2) Commitments, policies and strategies regarding human rights

Arbonia **is aware of its economic, ecological, and social responsibility** and **has committed itself in its Code of Conduct** 1.) to respecting human rights, especially taking into account the prohibition of child labour, 2.) to ensuring the occupational health and safety of ist employees, 3.) to cooperating with suppliers who meet their obligation to sustainability and social responsibility, 4.) to observing environmental protection standards, and 5.) to using resources carefully.

As an internationally active company, **Arbonia is aware of its responsibility for respecting human rights and avoiding child labour.** In all countries in which it is active, it complies with the United Nations' General Declaration on Human Rights, the UN Convention on overcoming discrimination against women, the UN Convention on the Rights of the Child and additional international humans rights protection standards. In the reporting year, Arbonia also initiated the accession to the UN Global Compact and has **committed to supporting the implementation of the ten principles** in the sub-areas of human rights, labour standards, environmental protection and fighting corruption. The accession was initiated in 2021 and confirmed in January 2022.
The Code of Conduct is supplemented by further directives, such as the anticorruption directive, the directive concerning insider trading, the directive for protection against sexual harassment, bullying, and discrimination at the workplace, and many more.

The diversity of employees, their equal opportunity, non-discrimination, as well as equal treatment in the company have great importance at Arbonia – regardless of sex, nationality, ethnic origin, skin colour, religion, or impairment.

3) Actions to prevent or mitigate potential negative impacts on human rights

The group **is raising awareness of ecological and social factors** in purchasing and is **working on a company-wide standard for supplier assessment according to ESG criteria** (environmental, social, governance). Starting in 2022, this assessment is to be mapped on the new e-procurement platform of the Group. **For evaluation of the suppliers,** external key figures (e. g. Creditworthiness, risk indicators, ESG ratings) are also to be collected and combined with the internally generated data for a holistic assessment in the future. For this purpose, Arbonia **decided in the reporting year to procure external ESG ratings via EcoVadis**

[34] https://github.com/planet-10/sentence-tranformer-based-text-classifier/blob/main/00_data/reports/Arbonia_Sustainability_Report_2021.pdf

starting on 1 January 2022. The aim is to check how many suppliers and what portion of the purchasing volume is covered by an ESG assessment. **Suppliers already have to guarantee that human rights are respected and, in particular, that child labour and forced labour are prevented.**

For example, 80% of the direct expenditures are covered by **supply agreements** with the most important suppliers. **These include ecological aspects (e. g. environmental protection, prohibited substances) and social issues** (e. g. respect for human rights, prohibition of child, forced and compulsory labour). **The division checks the suppliers' positions on the regulations concerning prohibited materials annually**. In addition, Sabiana has initiated a data survey to **check the positions of the suppliers on the topic of conflict minerals (3TG)**.

To ensure that the criteria regarding procurement and supply chain are observed, all companies of the division **carry out internal as well as external audits** in the areas of quality, social issues and energy efficiency. For this reason, a strategic category management was further expanded in the division during the reporting year. This **continuously collects and evaluates market information to react to potential risks in the supply chain** early on.

4) Actions to address or remedy actual negative impacts

5) Targets, indicators, and processes used to track or evaluate the effectiveness of the actions

6) Effectiveness of the actions, including progress toward the goals and targets

Annex 4: Normalised performance metrics