ASSESSING THE DATA CHALLENGES OF CLIMATE-RELATED DISCLOSURES IN EUROPEAN BANKS. A TEXT MINING STUDY

2023

BANCO DE ESPAÑA
Eurosistema

Documentos de Trabajo
N.º 2326

Angel-Ivan Moreno and Teresa Caminero

ASSESSING THE DATA CHALLENGES OF CLIMATE-RELATED DISCLOSURES
IN EUROPEAN BANKS. A TEXT MINING STUDY

ASSESSING THE DATA CHALLENGES OF CLIMATE-RELATED DISCLOSURES IN EUROPEAN BANKS. A TEXT MINING STUDY

Angel-Ivan Moreno

BANCO DE ESPAÑA

Teresa Caminero

BANCO DE ESPAÑA

Documentos de Trabajo. N.º 2326

September 2023

https://doi.org/10.53479/33752

ISSN: 1579-8666 (on line)

Abstract

The Intergovernmental Panel on Climate Change (IPCC) – a United Nations body responsible for assessing the climate change science and producing comprehensive assessment reports on climate change, its impacts, future risks, and strategies to mitigate its progress – estimates that global net zero should be achieved by 2050. To this end, many private firms are pledging to reach net-zero emissions by 2050. The Climate Data Steering Committee (CDSC) is working on an initiative to create a global central digital repository of climate disclosures, which tries to address the current data challenges.

This paper assesses the progress within European financial institutions in addressing the data challenges outlined by the Climate Data Steering Committee (CDSC), which was established in 2022 with the aim of supporting the climate objectives of the United Nations regarding high-quality climate data. Using a text-mining approach, coupled with the application of commercial Large Language Models (LLMs) for context verification, we calculate a Greenhouse Gas Disclosure Index (GHGDI), by analysing 23 very granular disclosures in the ESG reports between 2019 and 2021 of most of the significant banks under the ECB's direct supervision. This index is then compared to the score given by CDP. The results indicate a moderate correlation between institutions not reporting to CDP upon request and a low GHGDI. Institutions with a high CDP score do not necessarily correlate with a high GHGDI.

Keywords: ESG, sustainability, environment, climate change, carbon emissions, natural language processing, climate data challenges, OpenAI’s ChatGPT, Google’s text-bison.

JEL classification: C88, G32, Q56.

Resumen

El Grupo Intergubernamental de Expertos sobre el Cambio Climático (IPCC) —un organismo de las Naciones Unidas encargado del estudio de la ciencia del cambio climático, así como de generar informes de evaluación exhaustivos sobre dicho fenómeno, sus impactos, sus amenazas futuras y las estrategias para mitigar su avance— estima que se debe alcanzar el objetivo de cero emisiones netas a escala global para el año 2050. A fin de conseguirlo, muchas empresas privadas se han comprometido a lograr las cero emisiones netas para esa fecha. El Comité Directivo de Datos Climáticos (CDSC), entidad creada en 2022 con el propósito de respaldar los objetivos climáticos de las Naciones Unidas en lo que se refiere a la obtención de datos climáticos de alta calidad, está trabajando en una iniciativa para crear un repositorio digital central global de divulgaciones climáticas que intente abordar los desafíos de datos actuales.

Este documento evalúa el progreso de las instituciones financieras europeas a la hora de abordar los desafíos de datos descritos por el CDSC. Utilizando un enfoque de minería de textos, junto con la aplicación de modelos de lenguaje de gran tamaño (LLM) para la verificación de contexto, calculamos un Índice de Divulgación de Gases de Efecto Invernadero (GHGDI), analizando 23 divulgaciones muy específicas presentes en los informes ESG comprendidos entre 2019 y 2021 y pertenecientes a la mayoría de los bancos significativos bajo supervisión directa del BCE. Posteriormente se compara este índice con la puntuación otorgada por el Carbon Disclosure Project (CDP). Los resultados indican una correlación moderada entre aquellas instituciones que no informan al CDP cuando se les solicita y un bajo GHGDI. Las instituciones con una puntuación alta, por su parte, no guardan necesariamente correlación con un GHGDI alto.

Palabras clave: ESG, sostenibilidad, medio ambiente, cambio climático, emisiones de carbono, procesamiento de lenguaje natural, desafíos de datos climáticos, ChatGPT, text-bison.

Códigos JEL: C88, G32, Q56.

1 Introduction

In 1992, 154 states signed a treaty known as the United Nations Framework Convention on Climate Change (UNFCCC), to combat "dangerous anthropogenic interference with the climate system" (UNFCCC, 1992). Since then, several Conferences of the Parties (COPs) have taken place to discuss how to achieve that goal. In 2015, during the COP21, the Paris Agreement was signed, with the objective to hold "the increase in the global average temperature to well below 2°C above pre-industrial levels" and pursue efforts to "limit the increase of global temperature to 1.5º above pre-industrial levels" (UNFCCC, 2015). To that end, it also set the aim "to achieve a balance between anthropogenic emissions by sources and removals by sinks of greenhouse gases in the second half of this century". This balance objective is commonly referred as net-zero.

The Paris Agreement also included the need to provide a "national inventory report of anthropogenic emissions by sources and removals by sinks of Greenhouse Gases" (GHG). In 2018, the UNFCCC adopted the 2006 Intergovernmental Panel on Climate Change (IPCC) [1] Guidelines as the methodology for this inventory.

Every 5 to 7 years, IPCC publishes comprehensive Assessment Reports (AR) to provide scientific assessments and policy recommendations on climate change. Throughout the IPCC assessments, the year 2050 has become a reference for emission reduction targets and net zero emissions balance, prompting to not only governments but also private corporations to commit to achieving net zero CO_2 emissions by 2050.

The growth in pledges to reach net-zero emissions has led to a proliferation of criteria and benchmarks, making it difficult to measure and hold accountable entities making those commitments. This has led to a perceived lack of clear and generally accepted standards, with the risk of undermining serious stakeholders and enable greenwashing. To ensure credibility and accountability of net-zero pledges, the UN appointed a High-Level Expert Group on the net zero Emissions Commitments of Non-State Entities. They presented their recommendations in the COP27 (2022). As part of their recommendations they included the creation of an open-source repository of climate disclosures overseen by the UNFCCC and available at their Global Climate Action Portal. (UN HLEG on the net zero emissions of non-state entitities, 2022)

In line with this recommendation, the Climate Data Steering Committee (CDSC) was created in 2022 with aim to serve the climate objectives of the United Nations regarding high-quality climate data, presenting the net zero Data Public Utility (NZDPU) as "an open, free, and centralized data repository that would allow all stakeholders to easily access key climate

[1] The IPCC was established by the World Meteorological Organization (WMO) and the United Nations Environment Programme (UNEP) in 1988 to provide scientific assessments and policy recommendations on climate change.

transition-related data, commitments, and progress of businesses and financial institutions toward those commitments" (CDSC, 2023), with the declared intention to be hosted in the Climate Action Portal once a pilot product is built, which is expected to be ready by the end of 2023.

As of today, the most similar initiative to the NZDPU already in place is the voluntary questionnaire created and overseen by CDP. CDP is a not-for-profit charity (formerly known as Carbon Disclosure Project) that runs a global disclosure system. It houses and evaluates self-reported data from over 18,700 companies and 1,100 cities, states, and regions in three dimensions: climate, forests and water. CDP provides open access to several datasets via its portal. Each year, a scoring team within CDP evaluate the responses received and assigns a score to each of the dimensions, following a methodology that assesses the level of detail and comprehensiveness of each response.

The climate dimension of CDP is aligned with the recommendations of the Task Force on Climate-related Financial Disclosures (TCFD). These recommendations are widely adopted and applicable to organizations across sectors and jurisdictions and are structured around four thematic areas that represent core elements of how organizations operate: Governance, Strategy, Risk Management, and Metrics and Targets. However the design of the NZDPU closely follows the GHG Protocol methodology[2], which has a much narrower scope than the TCFD, as it focuses specifically on emissions disclosures. The NZDPU will also focus on trying to overcome the existing data challenges experienced by users of climate transition related data and identified by the CDSC[3].,

The present paper assesses the progress towards overcoming those data challenges in the GHG emissions disclosures within European financial institutions using text mining techniques.

The application of these techniques to climate-related disclosures is not new. One of the earliest studies was performed in 2008 (Doran & Quinn, 2008) where they basically looked for the terms "climate change", "global warming" and "greenhouse gas" in the SEC 10K filings, although in relation to corporate disclosures, the focus seems to have been more on techniques and applications related to sentiment analysis, topic modelling and complexity analysis (Bholat, Hansen, Santos, & Schonhardt-Bailey, 2015). With the publication of the TCFD recommendations, several papers have tried to evaluate the level of compliance with the recommendations using different Natural Language Processing (NLP) techniques. In their 2018 and 2019 (TCFD 2018 and 2019) Status Reports, the TCFD made use of supervised machine learning techniques to identify areas of the corporate reports potentially containing information related to each one of 11 recommended disclosures. Luccioni, Baylor and Duchene (2020) trained a model (ClimateQA) to identify climate-relevant sections on financial

[2] While the IPCC guidelines are primarily focused on national-level reporting of emissions by countries, the GHG Protocol is the most internationally known and used standard for calculating GHG emissions by corporations. The standard ISO 14064, developed afterwards, is very similar to the GHG Protocol

[3] See "Recommendations for the Development of the Net-Zero Data Public Utility" (CDSC, 2022)

reports. Later, Bingler et al (2022) also trained a Large Language Model (LLM), ClimateBERT, specifically on climate disclosures related to TCFD. With the popularization of general purpose LLMs such as ChatGPT, numerous papers have used it as a replacement or enhancement of traditional NLP techniques. Central Banks have also showed interest in the benefits of general purpose commercial LLMs and have found for example that ChatGPT seems to better capture the sentiment compared to dictionary approaches or even smaller models such as BERT (Alonso-Robisco & Carbó, 2023).

In this paper we use an approach similar to previous works of the Bank of Spain (Moreno & Caminero, 2020; Moreno & Caminero, 2022, Fernández-Rosillo, Koblents & Morales, 2023), coupled with the application of commercial Large Language Models (LLM) for context verification. The application of LLMs as a filtering step is a novelty to previous approaches adding precision to the results, allows having less iterations in the creation of the lexicons and proves to be more efficient than a manual verification process. We calculate a GHG Disclosure Index, by analysing 23 granular disclosures (see table 2) in the ESG reports between 2019 and 2021 of most of the significant banks under the ECB direct supervision, starting from the list of Significant Institutions (SI) of the Single Supervisory Mechanism (SSM) as of November 2022[4]. This index is then compared to the climate score given by the non-for profit organization CDP. These disclosures are based on the data challenges and recommendation outlined by the CDSC and can be considered challenging disclosures because of the difficulty for the Companies to actually obtain that information. In this sense, this paper presents a text-mining approach to assess the progress towards these requirements, exploring the power of new LLMs to perform context verification, which is one of the most important weaknesses of keyword and rule-based approaches.

The results indicate a moderate correlation between institutions not reporting to CDP upon request and a low GHG Disclosure Index (GHGDI). Institutions with high CDP score do not necessarily correlate with a high GHGDI which might be due to multiple factors, such as the specificity of the Index and potential differences between the public corporate disclosures and the answers to the CDP questionnaire.

The coupling of traditional NLP methods together with LLM for filtering provide a better individual precision in the excerpt retrieval of the targeted context. The resulting GHDI index allows measuring the progress in overcoming current challenges of climate-related disclosures and, despite its specificity, the moderate correlation with the CDP Score provides an indication of its robustness.

The report is organized as follows: first we describe the methodology, tools used and sample of study; then, we analyse the results and compare them with the CDP scores and finally, we summarize the conclusions.

[4] https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.listofsupervisedentities202212.en.pdf.

2 Methodology

We gathered the different ESG reports in English from the corporate websites of each institution. If no English version was found, the institution was not considered. If no ESG report was found, the Annual or Integrated report was used instead. In the event that a subsidiary was also a SI, we considered the highest level of consolidation, but also included the reports of the individual subsidiary, if available. This meant that since the disclosures of certain banks were considered at international consolidation level, sometimes they were outside the SSM scope (e.g. Bank of Canada, which has a subsidiary that is a SI in Luxembourg). The only exception was Sberbank Europe AG, which was considered at European consolidation level, since it was the only level available. In any case, disclosures by institutions with several subsidiaries in different SSM countries were only counted once. Those institutions for which reports were not available in English for the three years were discarded. As a result, the sample consisted in 99 consolidated SIs distributed as shown in Table 1

Country	**AT**	**BE**	**BG**	**CY**	**DE**	**EE**	**ES**	**FI**	**FR**	**GR**	**IE**	**IT**	**LU**	**LV**	**MT**	**NL**	**PT**	**SI**	**Total**
Significant Institutions	**7**	**4**	**1**	**1**	**19**	**2**	**10**	**4**	**10**	**4**	**7**	**12**	**4**	**2**	**2**	**7**	**2**	**1**	**99**
>€500 billion assets	1	1			4	1	2	1	6		3	2	1	1		2			25
€100 billion - €500 billion assets	2	2			6	1	2	2	1		1	3				2			22
€30 billion - €100 billion assets	2	1			9		6	1	3	4	2	7	2			3	2		42
<€30 billion assets	2		1	1							1		1	1	2			1	10

Table 1: Overview of the distribution of the sample according to institution size and country. (*Source: own elaboration)*

Note that we did not gather CDP reports. This was intentional and not only because most of the institutions do not publish the CDP reports on their website, but also because we wanted to analyse their own corporate reporting strategy.

The CDSC identified 6 challenges related to Emissions Accounting and Transition plan metrics and targets. (CDSC, 2022). Following these data challenges, we identify 23 specific disclosures and we try to identify whether an institution provides some information related to them, using a text mining approach. Table 2 shows the list of Data challenges outlined by the CDSC on those two areas and the specific disclosures used in the paper as a proxy for the index.

Data Challenge	Short description of the challenge	Specific disclosures.
Emissions accounting		
Emissions reporting	Low levels of disclosure	1. Market and location based calculations for scope 2 emissions 2. Emissions disclosed using the standard unit CO2 equivalent. 3. GHG Protocol being used as a reference 4. Accounting of emissions for all 7 gases outlined in the GHG Protocol 5. Control approach used for the organizational boundary 6. Limited or reasonable auditing assurance explicitly mentioned 7. Limited or reasonable auditing assurance explicitly mentioned in the context of ESG
Financed emissions	Lack of data on financed emissions.	8. Mention to the NetZero banking alliance 9. Financed emissions 10. Intensity of financed emissions 11. Sector of the financed emissions 12. PCAF quality score
Emissions estimates	High use of estimates rather than reported data; lack of clarity with regard to estimated versus reported data.	13. Distinction of actual vs estimated emissions
Carbon credits disclosure	Lack of granularity in carbon credit disclosures	14. Carbon credits 15. Carbon credit certificates
Transition plan metrics and targets		
Emissions Reduction Targets	Inconsistent target-setting and tracking methodologies	16. Science Based Targets 17. Emission targets 18. Emission targets per scope 19. Emission reduction targets per scope
Financed Emissions Reduction Targets	Inconsistent target-setting and tracking methodologies	20. Emission reduction targets for financed emissions 21. Sectors for financed emission targets 22. Exposure to Green bonds 23. Renewable energy financing

Table 2: Data Challenges identified by the CDSC and specific disclosures used as a proxy. (*Source: own elaboration)*

(Moreno & Caminero, Application of text mining to the analysis of climate-related disclosures, 2020). Figure 1 shows the pipeline of the process.

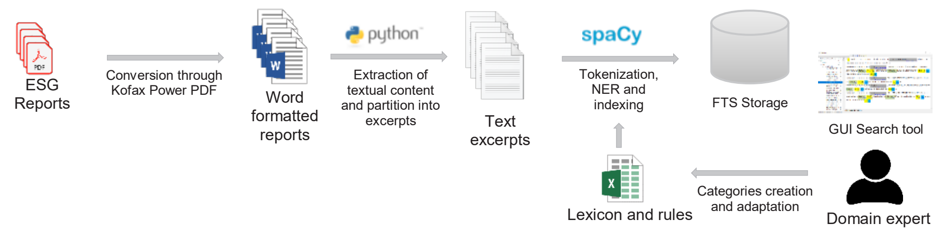

Figure 1 Representation of the workflow of the first part of the process. Source: own elaboration

Each disclosure was addressed by a query. Institutions were awarded one point for each disclosure that had at least one matching excerpt, and zero points for each disclosure that had no matching excerpts. As explained in our previous paper, keyword search has the risk to returning many false positives, that is, results that contain the searched concepts within a non-applicable context. For example, one of the keywords used was "market-based". Its purpose was to identify excerpts related to specifying whether market-based or location-based calculation was used in relation to scope 2 emissions. One excerpt matching that keyword contained the following text: "Only applying the regulatory policy makes little sense economically. It must rather serve as a framework, within which market-based instruments will create an efficient solution. Market-based state instruments include taxes/subsidies and the certificate trading.". The context of this excerpt is not related to scope 2 emissions and, thus, would be considered a false positive at the excerpt level. If this was the only excerpt applicable to a given bank, the bank would have been awarded one point for having at least a matching excerpt, becoming a false positive at the bank level. Note that if there were other applicable matching excerpts (true positives), the bank would have been righteously awarded the point, despite the one false positive excerpt. So a false positive at the excerpt level does not always translate into a false positive at the bank level.

In our previous paper we performed a manual review of all the matching excerpts and tweaked the queries in order to avoid false positives. For this paper we improved this approach by taking advantage of the availability of general-purpose commercial LLMs, which proved to have additional benefits in precision and efficiency. For each excerpt returned by each query, we performed a second filtering using two commercial LLMs: Google's text-bison[5] and OpenAI's chatGPT[6]. We used the LLMs with a specific prompt for each disclosure to confirm that the keywords matched by the search query applied to the appropriate context. An example of prompt for the first disclosure was:

[5] At the time of this writing, Google's text-bison model, based on Palm2, was in preview stage.

[6] At the time of this writing, the OpenAI's chatGPT version was gpt-3.5-turbo-0301

"*Classify the following text with an integer with a value of 0 or 1. 0 means that the text does not provide market or location-based scope 2 emissions or does not mention market or location-based calculation as a methodology for calculating electricity Scope 2 emissions. 1 means that the text refers to location or market based calculation as a methodology for calculating electricity Scope2 emissions or provides a value for market or location-based emissions. Please provide your answer as a list in the format [i, v], where i is the index of the original text and the v is the value 0 or 1 as per the previous instruction. The text is enclosed between {''}. There is just one text in total.*
Text to analyse: {plain_text}
Answer in the form of [1,v] being v a value of 0 or 1:"

Although the prompt was slightly different for the two models, the differences were mainly directed to force a specific output format in a reliable manner, without conveying a different interpretation of the main instruction.

An institution was considered compliant with a disclosure for a given year if there was at least one matching excerpt for that year for the corresponding query. We did the calculations before and after the LLM filtering step using both models. Finally, the discrepancies between the two models were manually reviewed in order to obtain the final index. This process is much more efficient that performing a manual review of all the results, as the manual review only focuses on the discrepancies. It also allows to reduce the iterations to refine the rules approach, as the false positives are more likely to be identified by the LLM filtering step. So, the fast retrieval of the keyword search is complemented with the precision of the context identification of LLMs, making it an efficient approach.

The GHG Disclosure Index (GHGDI) was calculated by awarding one point for each of the 23 disclosures identified and rescaling the final value between 1 and 10. Figure 2 shows the pipeline of this second part of the process for calculating the individual GHGDI for each of the 99 institutions of the sample, once the categories and keywords of the queries have been defined.

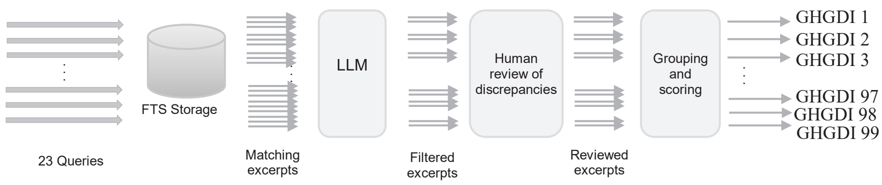

Figure 2 Representation of the workflow of the second part of the process. Source: own elaboration

CDP scores the responses with a letter between D- and A (8 levels), representing the "level of action reported by the company to assess and manage its environmental impacts during the reporting year" (CDP, 2023). The CDP may request responses to certain corporations that the signatories consider relevant. Those corporations for which a response has been requested, but no answer has been received, are scored with an F. To be able to compare the GHGDI with the CDP Score, we mapped the scores D- to A, to values 3 to 10 and assigned 0 to the F score. We then calculated the correlation of the individual CDP Scores and our GHGDI, both including corporations scored with an F and without including them. Since only a subset of 71 institutions with GHGDI were present in the CDP database, the calculation was performed only on that subset, although not all institutions reported in all years of the range of study. Additionally, some institutions can decide not to make their scores publicly available for any given year, and some institutions were not scored in a specific year for undisclosed reasons (although it is possible that they submitted their responses past the CDP deadline). These cases were filtered out from the subset. Table 3 shows the number of institutions considered per year and their average CDP score, according to the mapping described above.

	2019		2020		2021	
Size	**Number of SI**	**Average Score**	**Number of SI**	**Average Score**	**Number of SI**	**Average Score**
<30			1	0	1	0
30-100	24	5,3	28	5,0	28	5,3
100-500	13	7,2	15	6,3	15	6,9
>500	24	7,9	24	7,7	18	7,7
Total	61	6,7	68	6,2	62	6,3

Table 3 Number of SI reporting to CDP between 2019 and 2021 and average score per size in € billion

3 Results

Table 4 shows the evolution of the GHGDI of the subset of institutions reporting to CDP, including the LLM filtering, while Table 5 shows the GHGDI of all institutions of the sample.

	unfiltered			text-bison-filtered			gpt3.5-filtered		
Size	**2019**	**2020**	**2021**	**2019**	**2020**	**2021**	**2019**	**2020**	**2021**
<30		2,2	4,4		2,2	4,4		2,2	4,4
30-100	2,3	2,8	3,8	2,1	2,7	3,6	2,2	2,7	3,7
100-500	2,8	3,8	5,3	2,8	3,7	5,0	2,8	3,7	5,2
>500	3,4	4,1	5,6	3,2	3,9	5,4	3,3	4,0	5,5
Average	**2,8**	**3,2**	**4,8**	**2,7**	**3,1**	**4,6**	**2,8**	**3,1**	**4,7**

Table 4 Calculated GHGDI for the SI that report to CDP, per year and size in € billion

	unfiltered			text-bison-filtered			gpt3.5-filtered		
Size	**2019**	**2020**	**2021**	**2019**	**2020**	**2021**	**2019**	**2020**	**2021**
<30	0.5	0.7	2.3	0.4	0.6	2.1	0.4	0.6	2.1
30-100	2.0	2.4	3.2	1.7	2.3	3.0	1.8	2.3	3.1
100-500	2.5	3.2	4.7	2.4	3.0	4.4	2.4	3.2	4.6
>500	3.4	3.8	5.2	3.2	3.8	5.0	3.3	3.9	5.0
Average	**2.1**	**2.6**	**3.8**	**1.9**	**2.4**	**3.6**	**2.0**	**2.5**	**3.7**

Table 5 Calculated GHGDI for all SI in the sample, per year and size in € billion

Note that the average GHGDI is in general very low, since we are looking at very granular and challenging disclosures. Besides, larger banks tend to have better GHGDI values, and they improve over time. Regarding the LLM filtering step, the text-bison model rejects more excerpts than the gpt3.5 model. There are 6,831 combination of year-bank-disclosure[7]. Out of these, only 2,059 combinations actually had at least one matching excerpt, meaning that a given bank for a given year and disclosure was awarded 1 point. After processing the matching excerpts through the two LLMs, there were only 66 year-bank-disclosure combinations for which both models considered they were out-of-context and thus, false positives. In addition, there were 76 for which the models had different results. 10 of those were evaluated as in-context by the gpt3.5 model while being evaluated as out-of-context by the text-bison model. A manual review revealed that all of them were actually true-positives, so the keyword search was originally correct awarding them 1 point. On the other hand, there were 66 combinations that

[7] 3 years x 99 banks x 23 disclosures = 6,831

gpt3.5 evaluated as in-context while text-bison evaluated them as out-of-context. A manual review indicated that 28 were true positives, while 38 where out-of context. As a result, out of the 2,059 original matching combinations, after the filtering, they were reduced to 1,955 matching combinations. In any case, the final average scores obtained after filtering are not much different from the unfiltered versions. Considering that processing excerpts through LLM models is much slower than a keyword search, the keyword approach still has its advantage when processing large amounts of text excerpts and evaluating the results at an aggregated level. One of the main weaknesses of a keyword approach is the lack of context-awareness. This becomes more apparent when a word can have multiple meanings depending on the context. For example, in relation to the first disclosure ("Market and location based calculations for scope 2 emissions") when looking for the keyword "market-based", most of the results are related to the concept being searched, but there are some instances where the expression is present as part of "market-based instruments" which should be considered out-of context or false positive. Note also that the high level of precision of the keyword search in this case, is due to the specificity of the concepts being searched and the unlikeliness that those concepts appear in a different context than the one intended. In fact, most false positives were found within a relatively small subset of disclosures, for which the keywords were not specific enough. In these situations, LLMs might show an advantage at better identifying the context and, at individual institution level, the additional precision provided by LLMs would also mean a better insight on the actual disclosure level of its reporting. In the given example of "market-based instruments", both LLMs were able to identify them as false positives. Although in this specific study, using LLMs produced comparable results to those obtained without them, their use acts as a sort of robustness test providing more confidence in the results.

After the manual review of the discrepancies between the two models (the cases where both models agreed did not go through a manual review), we obtained a final GHGDI. Table 6 shows the evolution of this final GHGDI of the subset of institutions reporting to CDP while Table 7 shows its evolution for all institutions in the sample. Figure 3 shows a chart of the final GHGDI for all institutions in the sample differentiating between the disclosures related to "Emissions accounting" and the disclosures related to "Transition plan metrics and targets". Note that "Emissions accounting" accounts for 15 specific disclosures while "Transition plan metrics and targets" accounts for 8 specific disclosures (see Table 2), which translates in a larger area in the chart for "Emissions accounting".

Size	**2019**	**2020**	**2021**
<30		2,2	4,4
30-100	2,1	2,7	3,7
100-500	2,8	3,7	5,1
>500	3,2	4,0	5,4
Average	**2,7**	**3,1**	**4,6**

Table 6 Calculated final GHGDI for the SI that report to CDP, per year and size in € billion

Size	2019	2020	2021
<30	0.4	0.6	2.2
30-100	1.8	2.3	3.0
100-500	2.4	3.1	4.5
>500	3.2	3.8	5.0
Average	**2.0**	**2.4**	**3.7**

Table 7 Calculated final GHGDI for all SI in the sample, per year and size in € billion

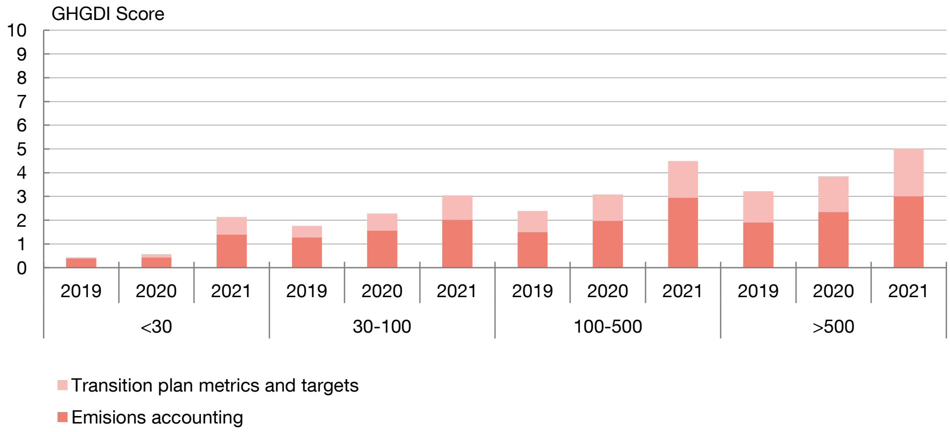

Figure 3. Calculated final GHGDI for all SI in the sample, per year and size in € billion

From the results, it is clear that, after incorporating the institutions that do not report to CDP, the GHGDI is lower, suggesting that not reporting to CDP is an indication of low level of challenging disclosures in ESG reports. In fact, if we create a binary variable indicating whether a corporation reports to CDP or not (its status for a given year is "Submitted"), regardless the CDP score given, we get a moderate statistical significant correlation of 0.51*** with our final GHGDI, reinforcing the relationship between reporting to CDP and providing better granular disclosures on GHG emissions.

The correlation between the final GHGDI and the CDP score also results in a moderate correlation (0.35***). It is worth mentioning, though, that when excluding the institutions not reporting to CDP upon request (the ones with a score of zero), the correlation is not significant, which might be an indication that institutions with high CDP score do not necessarily correlate with a high GHG Disclosure Index. This might be due to multiple factors, such as the specificity of the Index, which focuses on GHG disclosures instead of general TCFD disclosures, and potential differences between the public corporate disclosures and the answers to the CDP questionnaire. Figure 3 shows the relationship between CDP and GHGI where it can be seen that institutions with a CDP score of 0 (the ones that do not report upon request) tend to have also a low GHGDI

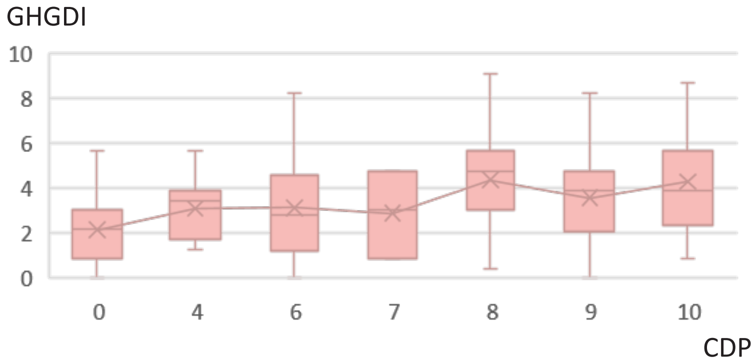

Figure 4. Relationship between GHGDI and CDP scores.

4 Conclusions

This paper assesses the progress towards overcoming the data challenges outlined by the CDSC within European financial institutions. We use a text-mining approach complemented with a context-filtering using two commercial LLMs. We then calculate a Greenhouse Gas (GHG) Disclosure Index (GHGDI), by analysing 23 challenging disclosures in the ESG reports between 2019 and 2021 of most of the significant banks under the ECB direct supervision. The use of both techniques, keyword search and LLMs, prove to be an efficient approach, complementing each other, reducing the potential errors of out-of-context matches. In the specific case of this study, where the keywords are not prone to be found in multiple contexts, the results do not present major differences, but the usage of the LLMs acts as a sort of robustness test providing more confidence in the results.

The GHGDI is in general very low, which is otherwise expected, since it focuses on very granular and challenging disclosures. Bigger banks tend to have better GHGDI. Nevertheless, there is a progressive increase in its value for all size ranges.

We compare this index with the score given by CDP. The results indicate a moderate correlation between institutions not reporting to CDP upon request and a low GHG Disclosure Index, but institutions with high CDP score do not necessarily correlate with a high GHGDI which might be due to multiple factors, such as the specificity of the Index and potential differences between the public corporate disclosures and the answers to the CDP questionnaire.

We also noticed that very few institutions publish their CDP report on their website, which might hinder transparency and data availability.

As a future area of study, we consider investigating further the scoring differences with the CDP, in order to confirm whether they are caused by the fact that the two approaches evaluate different disclosures or whether there is a tendency to disclose more to CDP than what is present in the sustainability reports, potentially prioritizing higher scores over higher transparency.

References

Alonso Robisco, Andrés, and José Manuel Carbó Martinez. (2023). "Analysis of CBDC narrative of central banks using large language models". Documentos de Trabajo, 2321, Banco de España. https://doi.org/10.53479/33412

Bholat, David, Stephen Hans, Pedro Santos and Cheryl Schonhardt-Bailey. (2015). "Text mining for central banks". Centre for Central Banking Studies, Bank of England, Handbook No. 33. https://doi.org/10.2139/ssrn.2624811

Bingler, Julia Anna, Mathias Kraus, Markus Leippold and Nicolas Webersinke. (2022). "Cheap talk and cherry-picking: What ClimateBERT has to say on corporate climate risk disclosures". *Finance Research Letters,* 47, Part B, 102776. https://doi.org/10.1016/j.frl.2022.102776

CDP. (2023a). "CDP Scores Explained". https://www.cdp.net/en/scores/cdp-scores-explained

CDP. (2023b). "The Net-Zero Data Public Utility". https://www.nzdpu.com/

CDSC. (2022). *Recommendations for the Development of the Net-Zero Data Public Utility.* https://assets.bbhub.io/company/sites/71/2022/11/development-of-the-net-zero-data-public-utility-november-2022.pdf

Doran, Kevin L., and Elias Leake Quinn. (2009). "Climate change risk disclosure: A sector by sector analysis of sec 10-k filings from 1995-2008". *North Carolina Journal of International Law and Commercial Regulation,* 34, pp. 721-766. https://scholarship.law.unc.edu/ncilj/vol34/iss3/2

Fernández-Rosillo San Isidro, Borja, Eugenia Koblents Lapteva and Alejandro Morales Fernández. (2023). "Micro-database for sustainability (ESG) indicators developed at the Banco de España (2022)". *Notas Estadísticas - Banco de España,* 17. https://repositorio.bde.es/handle/123456789/29496

IPCC. (2022). *Climate Change 2022 Mitigation of Climate Change. Summary for Policymakers.* https://www.ipcc.ch/report/ar6/wg3/downloads/report/IPCC_AR6_WGIII_SummaryForPolicymakers.pdf

Luccioni, Sasha, Emi Baylor and Nicolas Duchene. (2020). "Analyzing Sustainability Reports Using Natural Language Processing". NeurIPS 2020 Workshop on Tackling Climate Change with Machine Learning, 31. https://www.climatechange.ai/papers/neurips2020/31

Moreno Bernal, Ángel Iván, and María Teresa Caminero García. (2020). "Application of text mining to the analysis of climate-related disclosures". Documentos de Trabajo, 2035, Banco de España. https://repositorio.bde.es/handle/123456789/14185

Moreno Bernal, Ángel Iván, and María Teresa Caminero García. (2022). "Analysis of ESG disclosures in Pillar 3 reports: a text mining approach". Documentos Ocasionales, 2204, Banco de España. https://repositorio.bde.es/handle/123456789/20562

TCFD. (2018). *2018 Status Report.* https://assets.bbhub.io/company/sites/60/2020/10/FINAL-2018-TCFD-Status-Report-092518.pdf

TCFD. (2019). *2019 Status Report.* https://assets.bbhub.io/company/sites/60/2020/10/2019-TCFD-Status-Report-FINAL-0531191.pdf

UN HLEG on the Net Zero Emissions Commitments of Non-State Entities. (2022). *Integrity Matters: Net Zero Commitments by Businesses, Financial Institutions, Cities and Regions.* https://www.un.org/sites/un2.un.org/files/high-level_expert_group_n7b.pdf

UNFCCC. (1992). *United Nations Framework Convention on Climate Change.* https://unfccc.int/files/essential_background/background_publications_htmlpdf/application/pdf/conveng.pdf

UNFCCC. (2015). *Paris Agreement.* https://unfccc.int/files/essential_background/convention/application/pdf/english_paris_agreement.pdf

BANCO DE ESPAÑA PUBLICATIONS

WORKING PAPERS

2215 JOSÉ MANUEL CARBÓ and SERGIO GORJÓN: Application of machine learning models and interpretability techniques to identify the determinants of the price of bitcoin.

2216 LUIS GUIROLA and MARÍA SÁNCHEZ-DOMÍNGUEZ: Childcare constraints on immigrant integration.

2217 ADRIÁN CARRO, MARC HINTERSCHWEIGER, ARZU ULUC and J. DOYNE FARMER: Heterogeneous effects and spillovers of macroprudential policy in an agent-based model of the UK housing market.

2218 STÉPHANE DUPRAZ, HERVÉ LE BIHAN and JULIEN MATHERON: Make-up strategies with finite planning horizons but forward-looking asset prices.

2219 LAURA ÁLVAREZ, MIGUEL GARCÍA-POSADA and SERGIO MAYORDOMO: Distressed firms, zombie firms and zombie lending: a taxonomy.

2220 BLANCA JIMÉNEZ-GARCÍA and JULIO RODRÍGUEZ: A quantification of the evolution of bilateral trade flows once bilateral RTAs are implemented.

2221 SALOMÓN GARCÍA: Mortgage securitization and information frictions in general equilibrium.

2222 ANDRÉS ALONSO and JOSÉ MANUEL CARBÓ: Accuracy of explanations of machine learning models for credit decisions.

2223 JAMES COSTAIN, GALO NUÑO and CARLOS THOMAS: The term structure of interest rates in a heterogeneous monetary union.

2224 ANTOINE BERTHEAU, EDOARDO MARIA ACABBI, CRISTINA BARCELÓ, ANDREAS GULYAS, STEFANO LOMBARDI and RAFFAELE SAGGIO: The Unequal Consequences of Job Loss across Countries.

2225 ERWAN GAUTIER, CRISTINA CONFLITTI, RIEMER P. FABER, BRIAN FABO, LUDMILA FADEJEVA, VALENTIN JOUVANCEAU, JAN-OLIVER MENZ, TERESA MESSNER, PAVLOS PETROULAS, PAU ROLDAN-BLANCO, FABIO RUMLER, SERGIO SANTORO, ELISABETH WIELAND and HÉLÈNE ZIMMER. New facts on consumer price rigidity in the euro area.

2226 MARIO BAJO and EMILIO RODRÍGUEZ: Integrating the carbon footprint into the construction of corporate bond portfolios.

2227 FEDERICO CARRIL-CACCIA, JORDI PANIAGUA and MARTA SUÁREZ-VARELA: Forced migration and food crises.

2228 CARLOS MORENO PÉREZ and MARCO MINOZZO: Natural Language Processing and Financial Markets: Semi-supervised Modelling of Coronavirus and Economic News.

2229 CARLOS MORENO PÉREZ and MARCO MINOZZO: Monetary Policy Uncertainty in Mexico: An Unsupervised Approach.

2230 ADRIAN CARRO: Could Spain be less different? Exploring the effects of macroprudential policy on the house price cycle.

2231 DANIEL SANTABÁRBARA and MARTA SUÁREZ-VARELA: Carbon pricing and inflation volatility.

2232 MARINA DIAKONOVA, LUIS MOLINA, HANNES MUELLER, JAVIER J. PÉREZ and CRISTOPHER RAUH: The information content of conflict, social unrest and policy uncertainty measures for macroeconomic forecasting.

2233 JULIAN DI GIOVANNI, MANUEL GARCÍA-SANTANA, PRIIT JEENAS, ENRIQUE MORAL-BENITO and JOSEP PIJOAN-MAS: Government Procurement and Access to Credit: Firm Dynamics and Aggregate Implications.

2234 PETER PAZ: Bank capitalization heterogeneity and monetary policy.

2235 ERIK ANDRES-ESCAYOLA, CORINNA GHIRELLI, LUIS MOLINA, JAVIER J. PÉREZ and ELENA VIDAL: Using newspapers for textual indicators: which and how many?

2236 MARÍA ALEJANDRA AMADO: Macroprudential FX regulations: sacrificing small firms for stability?

2237 LUIS GUIROLA and GONZALO RIVERO: Polarization contaminates the link with partisan and independent institutions: evidence from 138 cabinet shifts.

2238 MIGUEL DURO, GERMÁN LÓPEZ-ESPINOSA, SERGIO MAYORDOMO, GAIZKA ORMAZABAL and MARÍA RODRÍGUEZ-MORENO: Enforcing mandatory reporting on private firms: the role of banks.

2239 LUIS J. ÁLVAREZ and FLORENS ODENDAHL: Data outliers and Bayesian VARs in the Euro Area.

2240 CARLOS MORENO PÉREZ and MARCO MINOZZO: "Making text talk": The minutes of the Central Bank of Brazil and the real economy.

2241 JULIO GÁLVEZ and GONZALO PAZ-PARDO: Richer earnings dynamics, consumption and portfolio choice over the life cycle.

2242 MARINA DIAKONOVA, CORINNA GHIRELLI, LUIS MOLINA and JAVIER J. PÉREZ: The economic impact of conflict-related and policy uncertainty shocks: the case of Russia.

2243 CARMEN BROTO, LUIS FERNÁNDEZ LAFUERZA and MARIYA MELNYCHUK: Do buffer requirements for European systemically important banks make them less systemic?

2244 GERGELY GANICS and MARÍA RODRÍGUEZ-MORENO: A house price-at-risk model to monitor the downside risk for the Spanish housing market.

2245 JOSÉ E. GUTIÉRREZ and LUIS FERNÁNDEZ LAFUERZA: Credit line runs and bank risk management: evidence from the disclosure of stress test results.

2301 MARÍA BRU MUÑOZ: The forgotten lender: the role of multilateral lenders in sovereign debt and default.

2302 SILVIA ALBRIZIO, BEATRIZ GONZÁLEZ and DMITRY KHAMETSHIN: A tale of two margins: monetary policy and capital misallocation.

2303 JUAN EQUIZA, RICARDO GIMENO, ANTONIO MORENO and CARLOS THOMAS: Evaluating central bank asset purchases in a term structure model with a forward-looking supply factor.

2304 PABLO BURRIEL, IVÁN KATARYNIUK, CARLOS MORENO PÉREZ and FRANCESCA VIANI: New supply bottlenecks index based on newspaper data.

2305 ALEJANDRO FERNÁNDEZ-CEREZO, ENRIQUE MORAL-BENITO and JAVIER QUINTANA: A production network model for the Spanish economy with an application to the impact of NGEU funds.

2306 MONICA MARTINEZ-BRAVO and CARLOS SANZ: Trust and accountability in times of pandemic.

2307 NATALIA FABRA, EDUARDO GUTIÉRREZ, AITOR LACUESTA and ROBERTO RAMOS: Do Renewables Create Local Jobs?

2308 ISABEL ARGIMÓN and IRENE ROIBÁS: Debt overhang, credit demand and financial conditions.

2309 JOSÉ-ELÍAS GALLEGOS: Inflation persistence, noisy information and the Phillips curve.

2310 ANDRÉS ALONSO-ROBISCO, JOSÉ MANUEL CARBÓ and JOSÉ MANUEL MARQUÉS: Machine Learning methods in climate finance: a systematic review.

2311 ALESSANDRO PERI, OMAR RACHEDI and IACOPO VAROTTO: The public investment multiplier in a production network.

2312 JUAN S. MORA-SANGUINETTI, JAVIER QUINTANA, ISABEL SOLER and ROK SPRUK: Sector-level economic effects of regulatory complexity: evidence from Spain.

2313 CORINNA GHIRELLI, ENKELEJDA HAVARI, ELENA MERONI and STEFANO VERZILLO: The long-term causal effects of winning an ERC grant.

2314 ALFREDO GARCÍA-HIERNAUX, MARÍA T. GONZÁLEZ-PÉREZ and DAVID E. GUERRERO: How to measure inflation volatility. A note.

2315 NICOLÁS ABBATE, INÉS BERNIELL, JOAQUÍN COLEFF, LUIS LAGUINGE, MARGARITA MACHELETT, MARIANA MARCHIONNI, JULIÁN PEDRAZZI and MARÍA FLORENCIA PINTO: Discrimination against gay and transgender people in Latin America: a correspondence study in the rental housing market.

2316 SALOMÓN GARCÍA: The amplification effects of adverse selection in mortgage credit suply.

2317 METTE EJRNÆS, ESTEBAN GARCÍA-MIRALLES, METTE GØRTZ and PETTER LUNDBORG: When death was postponed: the effect of HIV medication on work, savings and marriage.

2318 GABRIEL JIMÉNEZ, LUC LAEVEN, DAVID MARTÍNEZ-MIERA and JOSÉ-LUIS PEYDRÓ: Public guarantees and private banks' incentives: evidence from the COVID-19 crisis.

2319 HERVÉ LE BIHAN, DANILO LEIVA-LEÓN and MATÍAS PACCE: Underlying inflation and asymmetric risks.

2320 JUAN S. MORA-SANGUINETTI, LAURA HOSPIDO and ANDRÉS ATIENZA-MAESO: Los números de la regulación sobre igualdad. Cuantificación de la actividad normativa sobre no discriminación en España y su relación con las brechas de género en el mercado de trabajo.

2321 ANDRES ALONSO-ROBISCO and JOSÉ MANUEL CARBÓ: Analysis of CBDC Narrative of Central Banks using Large Language Models.

2322 STEFANIA ALBANESI, ANTÓNIO DIAS DA SILVA, JUAN F. JIMENO, ANA LAMO and ALENA WABITSCH: New technologies and jobs in Europe.

2323 JOSÉ E. GUTIÉRREZ: Optimal regulation of credit lines.

2324 MERCEDES DE LUIS, EMILIO RODRÍGUEZ and DIEGO TORRES: Machine learning applied to active fixed-income portfolio management: a Lasso logit approach.

2325 SELVA BAHAR BAZIKI, MARÍA J. NIETO and RIMA TURK-ARISS: Sovereign portfolio composition and bank risk: the case of European banks.

2326 ANGEL-IVAN MORENO and TERESA CAMINERO: Assessing the data challenges of climate-related disclosures in european banks. A text mining study.