Environmental Research Communications

LETTER

Using AI to assess corporate climate transition disclosures

Chiara Colesanti Senni[1,*], **Tobias Schimanski**[1,*], **Julia Bingler**[2,3], **Jingwei Ni**[1,4] and **Markus Leippold**[1,5]

[1] Department of Finance, University of Zürich, Zürich, Switzerland
[2] Smith School of Enterprise and the Environment, University of Oxford, Oxford, United Kingdom
[3] Council on Economic Policies, Zürich, Switzerland
[4] Center of Economic Research, ETH Zürich, Zürich, Switzerland
[5] Swiss Finance Institute (SFI), Zürich, Switzerland
[*] Authors to whom any correspondence should be addressed.

E-mail: chiara.colesantisenni@df.uzh.ch **and** tobias.schimanski@df.uzh.ch

Keywords: climate disclosures, large language models, RAG system, human evaluation, CA100+, transition plans

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
28 August 2024

REVISED
6 December 2024

ACCEPTED FOR PUBLICATION
12 December 2024

PUBLISHED
21 February 2025

Abstract

Company transition plans toward a low-carbon economy are key for effective capital allocation and risk management. This paper proposes a set of 64 indicators to comprehensively assess transition plans and develops a Large Language Model-based tool to automate the assessment of company disclosures. We evaluate our tool with experts from 26 institutions, including financial regulators, investors, and non-governmental organizations. We apply the tool to the sustainability reports from carbon-intensive Climate Action 100+ companies. Our results show that companies tend to disclose more information related to target setting (talk), but less information related to the concrete implementation of strategies (walk). In addition, companies that disclose more information tend to have lower emissions. Our results highlight the need for increased scrutiny of companies' efforts and potential greenwashing risks. The complexity of transition activities presents a major challenge for comprehensive large-scale assessments. As shown in this paper, novel and flexible approaches using Large Language Models can serve as a remedy.

1. Introduction

As the impacts of climate change become increasingly severe, the urgency for the global community to transition to a net-zero economy has never been more critical. This transition is essential to mitigate the negative effects of climate change and to ensure long-term economic stability and sustainability. Corporations are central to this transformation as significant contributors to greenhouse gas emissions. They are tasked with adjusting their operations and strategies to align with climate goals. This requires comprehensive planning, implementation, and transparency. Furthermore, this transformation requires a large amount of financial resources. Financial institutions are pivotal in directing capital towards sustainable activities and enabling the transition of corporate business models and technologies. However, they need information on comprehensive corporate transition strategies to assess the future risks and opportunities associated with their investments. Similarly, regulators rely on these plans to ensure that companies make genuine efforts to reduce their carbon footprints, which is critical to meeting national and international climate targets. Hence, the evaluation of transition strategies is key, as shown by recent contributions in this direction [1–5].

Despite the critical importance of corporate transition strategies, several challenges can undermine their effectiveness in supporting the achievement of the net-zero target. Reporting and credibility issues whereby companies misleadingly portray their climate and environmental efforts (greenwashing) are at the forefront. This can undermine trust and lead to a misallocation of resources in the economy, putting climate and environmental targets at risk and coming with negative implications for micro- and macro-financial stability [6–11].

The lack of standardization in company disclosures reduces the ability of stakeholders to compare and assess the ambition, credibility, and feasibility of the transition strategies of different companies. This poses a substantial barrier to effectively evaluating corporate contributions to climate goals and strategies to steer their business to the future. At the same time, the ongoing pressure to show progress in reducing climate impact leads companies and public institutions to release information that often results in vast amounts of unstructured data about transitioning toward net zero [12].

This paper addresses these challenges by providing a practical framework that assesses corporate climate transition disclosures. Our approach uses a Large Language Model (LLM)–based tool to automate and enhance the analysis of sustainability disclosures and identify potential risks of greenwashing. We build on previous work that uses language models in the field of sustainable finance and corporate climate risk analysis [9, 13–16, 28] and extend the literature by providing a detailed, expert-driven and scalable assessment process. Thus, we respond to the call for more fine-granular AI assessments in sustainability [17].

Our contribution is threefold. First, we define meaningful indicators to assess transition strategies. To this end, we review 28 transition strategy disclosure frameworks to identify common criteria for assessing corporate transition strategies and elicit experts' opinions to develop a unified framework encompassing 64 indicators.

Second, we build and validate an LLM-based tool for the automated analysis of transition plans based on the 64 common ground indicators. We assess the performance of the LLM-based tool in a pilot study involving users from 26 different institutions including financial regulators, investors, and NGOs to gain insights into practitioners' perception of the tool, its trustworthiness, and practical usage. We find that users are generally satisfied with the tool's performance, but there is user-specific disagreement regarding the priorities for improvement.

Third, we assess the quality of disclosures of the Climate Action 100+ companies to identify potential inconsistencies in their transition strategies.[6] Our findings show that companies tend to disclose more indicators related to target setting (talk) but fewer indicators associated with the concrete implementation of strategies (walk). Moreover, we find that the correlation between the quality of disclosures and carbon emissions tends to be negative in our sample.

This research demonstrates a flexible, comprehensive, and scalable approach to analyzing companies' transition efforts. This has multiple significant implications for stakeholders from fostering a better understanding of companies to questioning their actionable agendas and tackling information asymmetries concerning climate change.

2. Methods

The target of this work is to develop a sound and comprehensive evaluation method for companies' transition disclosure that is automated through an LLM-based tool. Our method comprises three steps. First, we define a set of indicators that enable a holistic assessment of transition planning. Second, we develop an LLM-based tool based on the latest technology advances. Third, we evaluate the tool with domain experts.

2.1. Assessment indicators

We initially define a set of indicators along which company disclosure related to transition strategies can be assessed. To identify these assessment indicators, we proceed in three steps.

In the first step, we review 28 existing transition plan frameworks published over the period 2021-2023 by different stakeholders and identify commonly suggested assessment indicators (see section S.3 for an overview). The indicators signal whether companies disclose information related to specific topics or not. For example, indicators are related to the disclosure of emission targets or reported emissions. In the second step, we assess the importance of the identified indicators by the frequency of appearance in the assessed frameworks. In the third step, we assign weights to the indicators based on the discussion with experts. We share our list of indicators with more than 50 selected researchers and practitioners who formed our advisory board.[7] The experts were asked to provide comments, suggest amendments, and refine the indicators.

Finally, we classify the indicators into 'walk' (W) and 'talk' (T). The distinction is made based on whether a specific indicator relates to future targets and/or general transition monitoring and management approaches (T) or to specific and already verifiable transition activities (W).

[6] Climate Action 100+ is an investor-led initiative that aims to incentivize the world's largest corporate greenhouse gas emitters to take climate action.

[7] The advisory board included financial industry representatives, central bankers, and financial supervisors. It was established to provide feedback throughout the entire project. For more details on the advisory board, see section S.2.

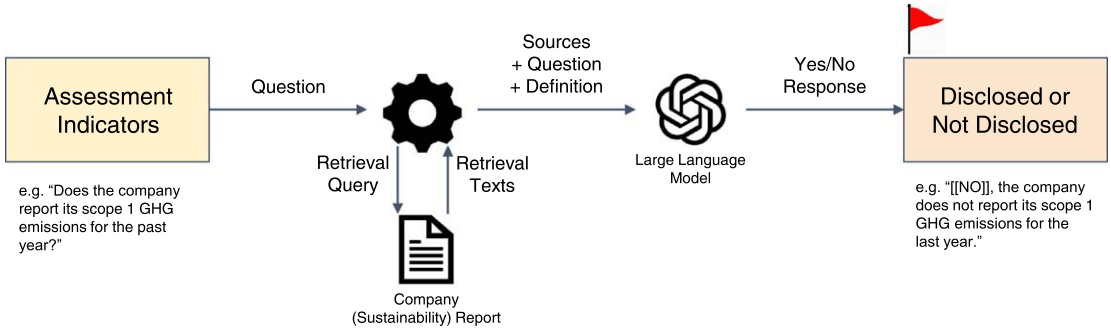

Figure 1. Retrieval Augmented Generation (RAG) pipeline for analyzing the assessment indicators.

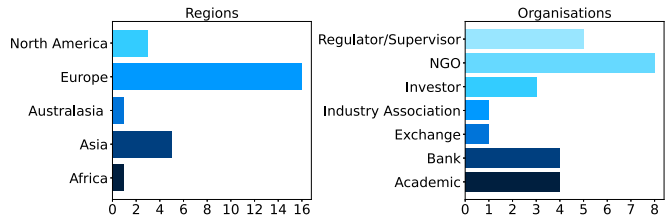

Figure 2. Overview of regions and organization types of the 26 participating organizations to the study.

2.2. LLM-based tool

To translate our framework into an automated analysis tool, we rely on Large Language Models (LLMs). LLMs have shown vast capabilities in reasoning, understanding, and generation of text [18–21]. However, LLMs also face hallucination, i.e., producing non-factual output [22]. As a remedy, practitioners and researchers rely on a technique called Retrieval Augmented Generation (RAG) [23]. RAG systems aim to include external knowledge in the prompts provided to the LLM and force the model to rely only on this information when answering a given question. Thus, RAG systems make use of the strong capabilities of LLMs to summarize and reason over the provided content and try to minimize the dependence on the internal knowledge of the LLM.

Figure 1 displays the RAG system used in this project. We use the assessment indicators as search queries applied to company sustainability reports. The information retrieved is used to answer the corresponding question (for exact details of implementation, see section S.5). Importantly, the tool produces a structured answer that indicates whether information on the indicator is available. The output is a yes or no answer, followed by an explanation of the decision and the source references. These explanations can enable a more holistic and, importantly, transparent understanding of the evaluation made. They allow the user to understand the reasoning behind the choice and can provide starting points for more detailed investigations. The sources allow the user to cross-check whether the model has extracted the relevant information and whether it has been correctly summarized (see section S.7 for more details).

2.3. Expert evaluation

We evaluate our tool in a pilot study. We choose an evaluation design that helps us obtain quantitative and qualitative feedback from domain experts. This adds to previous research, which has mainly addressed the quantitative evaluation of RAG systems in artificial or theoretical setups [24–27]. Only a few analyses have considered expert-based evaluations [28, 29]. However, none of these previous papers attempted a comprehensive evaluation of quality dimensions, such as correct sourcing and response, as well as the potential use of such tools for stakeholders.

Our evaluation incorporates feedback from domain experts from 26 organizations. The experts include financial regulators and supervisors, investors, exchanges, NGOs and industry associations representatives, bankers, and academics. The participants are predominantly based in Europe, followed by participants from Asia, North America, Australasia, and Africa (see figure 2 for more details). An organization participated with three people. Thus, we have 28 participants in total.

The domain experts assess the tool along the following qualitative dimensions:

1. **System quality**, which aims at understanding if relevant information is retrieved, answers are accurate and faithful, and reasoning capabilities are solid.
2. **Trustworthiness**, which yields insights into users' perceived trustworthiness and expertise of the model.
3. **Usage**, which aims at understanding the usefulness of the tool for the respective stakeholder, as well as future possible use cases.

In the evaluation, we distinguish between feedback on the tool's assessment of individual indicators, that is, the responses and sources for specific yes / no questions (Q1-Q9), and overall feedback about the tool in general (Q10-Q15). The participants provided an answer based on a given set of options (multiple choice) and could provide additional explanations (free text). For the exact questions and details of the setup, see section S.8.

3. Results

The results span three dimensions. First, we present the results of the assessment indicators selection process. Second, we present the results of the evaluation of the LLM-based tool. Finally, we apply the tool to the sustainability reports of the most emitting companies in the world, those addressed by the Climate Action 100+ initiative.

3.1. Assessment indicators

As a result of the analysis of the transition plan frameworks, we obtain a list of 64 indicators along which we assess transition-related disclosures. Section S.4 lists all the indicators selected, covering the broad categories 'Target', 'Governance', 'Strategy', and 'Tracking' (as used in e.g. TCFD). In addition, we show each indicator's classification as primarily a walk (W) or talk (T) indicator.

An important advantage of the proposed framework is its flexibility: While we suggest a list of indicators deemed appropriate to assess transition strategies, users can modify and extend this depending on their needs. For instance, financial institutions lending to specific sectors might want to add more details about risks that are specifically important for their assessment. Similarly, users interested in broader nature-related risks might want to include additional questions about supply chains and the locations of companies.

3.2. LLM-based tool evaluation

After the definition of the indicators, we create the corresponding LLM-based tool and evaluate it with 28 experts from 26 institutions. As a result, we obtain 28 assessments of the tool and 396 evaluations of individual indicators. The results of the human evaluation offer several insights.

Regarding **system quality**, participants display a very high level of satisfaction. For example, 81% of the participants found that the model summarizes relevant content correctly, without making up information not contained in the report (Q4). Furthermore, most of the participants perceive that the tool captures the most relevant information for the indicator analyzed (Q3) and provides the correct references (Q6-Q9). Negative feedback included too judgmental responses or occasionally missing out on sources. Regarding the main area of improvement for the model, part of the experts suggested a more detailed assessment, while another part highlighted that a broader overview would be desirable (Q14). Our results align with previous research that outlines that models can achieve very satisfactory results in terms of faithfulness [27]. However, our analysis also reveals that adaptation to individual needs is critical for specific users.

Regarding **trusthworthiness**, the user's first impression of the individual answers is largely positive (Q2), and the sources provided in an answer largely support the trust in the model (Q5). This is also mirrored in the general assessment of the tool, where only three respondents claimed that they did not trust the responses when using the model (Q15). Only two participants find the tool unsatisfactory (Q10). Generally, the answers of the tool were perceived as comparable with an expert with 1–2 or more years of experience in assessing companies' disclosures (Q11). However, it also becomes apparent that a large share of participants is not sure if they would fully trust the information provided by the tool (11 of 28 respondents 'partially' trust the model (Q15)). The optional explanations help us understand where these concerns stem from. Some participants outlined the lack of ability of the tool to handle sector-specific issues based on the current set of indicators. Others highlighted the need to compare the answers produced by the tool with third-party data.

Finally, the **usage** dimension sheds light on the ease of understanding the indicators themselves, use cases to apply the tool in practice, and potential for improvement. Concerning understanding the information requested by the indicators, the dominant feedback from the users is positive. Most of the participants would know what information they would look for if they had to answer the question themselves (Q1). Furthermore, stakeholders would utilize the tool for a wide set of use cases, ranging from the evaluation of corporate risk and opportunities

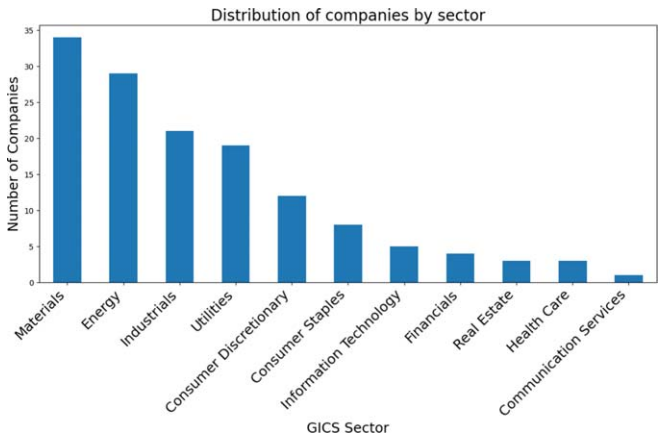

Figure 3. Distribution of the CA100+ companies by sector.

to a high- and deep-level understanding of transition plans (Q13). Again, enhancements reflect individual needs (see Q14). A common qualitative feedback is that the tool can be valuable for quick assessment and understanding, but future improvements could include deeper and more actionable insights or sector-specific adjustments (Q12, Q16).

Overall, human evaluation provides insights into understanding that the LLM-based tool represents a valuable asset for end users. However, the introduction of more user-specific adaptions of the tool could foster both trust and usage. For a detailed overview of the results, see section S.8.

3.3. CA100+ companies

To illustrate the functioning of our tool, we apply it to analyze corporate sustainability reports from Climate Action 100+ companies and assess the transition-related information contained in these reports along the 64 assessment indicators. CA100+ is an investor-led initiative to track the most emitting companies in the world. Our sample covers 143 companies and the corresponding corporate sustainability reports for the fiscal year 2022. The distribution of the companies across the different sectors is shown in figure 3.

For all the reports in our sample, we assess how many indicators are disclosed in the report of the company under consideration (i.e., the tool assigned a 'yes' answer to the prompt question of whether the information requested in the indicator is available in the report).

We find that the average count of indicators disclosed per report is 23 out of the 64 required indicators, while the best-performing report reaches a value of 43 (see section S.9 for an overview of all indicators). The distribution of the average share of indicators disclosed by the company shows that a value of almost 40% is achieved by more than 35% of the companies (see figure 4). However, it is important to understand the type of indicators for which information is available. Similarly, there might be sectoral patterns in the disclosure.

3.3.1. Most and least disclosed indicators

In our analysis of corporate disclosures, we observe a clear divergence in the quality of disclosure depending on the types of indicators considered. This highlights areas where companies excel and fall short in their disclosures. The ranking of the disclosed indicators shows a pattern for the 10% most disclosed and the 10% least disclosed indicators (see tables 1 and 2). For results by sector, refer to section S.10.

We first turn our attention to the top 10% disclosed indicators in table 1. These are related to GHG emission reduction interim targets, the structure of governance for environmental initiatives, oversight by corporate boards, and the specification of quantitative sub-targets. In addition, companies tend to disclose assessment indicators related to operational adjustments toward sustainability, including strategies related to adopting renewable energy and reporting Scope 1 GHG emissions from the past year.

In contrast, the disclosure of indicators that cover a more comprehensive and ambitious implementation of transition strategies in all operations and activities is less frequent (table 2). These least disclosed indicators are related to the responsible use of carbon credits and offsets, alignment of executive remuneration with climate goals, and holistic integration of climate strategies across all business operations. The lack of disclosure in these dimensions might suggest a lack of fully developed strategies or a reluctance to reveal comprehensive details.

In addition, the least disclosed assessment indicators focus on policy engagement transparency and ending the use of fossil fuels. These include examining the company's strategies for engagement in net-zero initiatives, policies on quitting support for additional fossil fuel exploration activities and plans to decommission fossil fuel

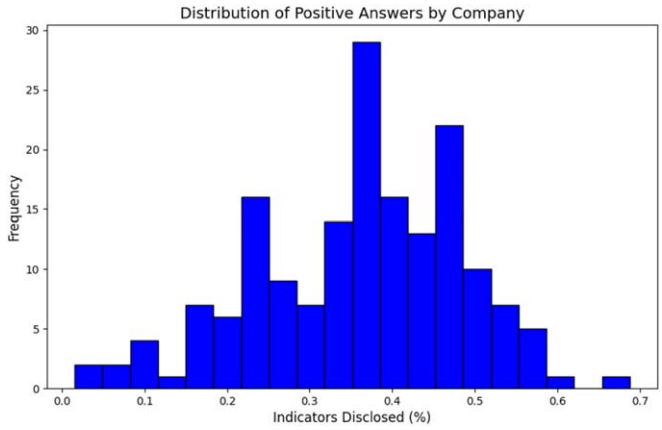

Figure 4. Distribution of the share of disclosed indicators by company.

Table 1. 10% Most disclosed indicators. The last column provides the suggested classifier whether the question is more related to 'talk' (T) or 'walk' (W).

Indicator	Question	
9	Does the company report its GHG emission reduction interim targets for achieving the overall goal?	T
13	Does the company explain its governance structure for managing the climate transition?	T
17	Does the company report how its board oversees the climate transition plan implementation?	T
23	Does the company report quantitative or quantifiable sub-targets in line with their climate targets and their climate key performance indicators?	T
26	Does the company report a renewable energy strategy and activities, covering renewable energy build-out, procurement, and consumption?	W
46	Does the company report its scope 1 GHG emissions for the past year?	W
50	Does the company report its annual progress of reducing GHG emissions to achieve its emission reduction or net zero targets?	W

Table 2. 10% Least disclosed indicators. The last column provides the suggested classifier whether the question is more related to 'talk' (T) or 'walk' (W).

Indicator	Question	
12	If carbon credits and offsets are reported to be used by the company, does the company state explicitly that carbon credits and offsets will be only used when the company can ensure that the emission reduction or emission avoidance is sustained permanently?	T
19	Does the company provide a higher share of remuneration and bonuses that are linked to the successful implementation of the climate transition plan interim targets compared to the general part of variable compensation for executives and managers?	W
22	Does the company provide comprehensive evidence that it fully and completely integrates its climate strategy into its business strategy, product development, operations, financial and human resources, asset management, and asset decommissioning?	W
36	Does the company report serious consequences and escalation strategies if net zero engagement is ineffective at upstream, downstream, policymaker, and industry association levels?	T
37	Does the company state explicitly that it stopped or will immediately stop any support or activities in new additional fossil fuel exploration and extension of fossil fuel supply?	W
38	Does the company report a strategy and activities for the decommissioning and canceling of planned or existing fossil fuel exploration and supply infrastructure?	W

infrastructure. The lack of disclosure in these areas could indicate a significant gap between companies' stated targets and the actual practices.

Using our classification, we observe that the least disclosed indicators are 'walk' indicators, while the most disclosed indicators are 'talk' related.

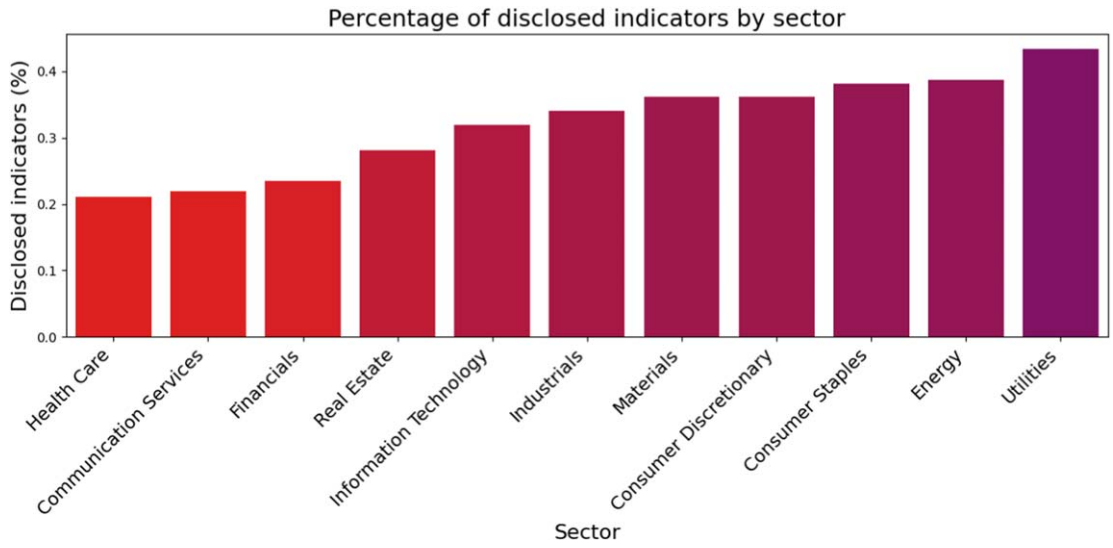

Figure 5. Average share of disclosed indicators by sector.

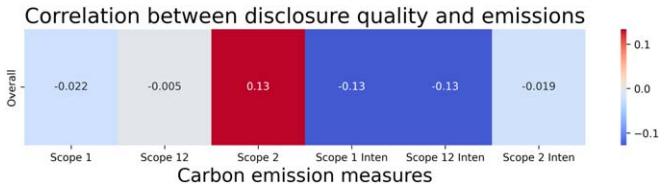

Figure 6. Correlation between amount of indicators disclosed and carbon emissions.

3.3.2. Sectoral analysis

Other than by the specific indicators, heterogeneity in disclosures could also stem from sectoral characteristics. Hence, we investigate whether companies in some sectors disclose, on average, more information than others, according to our indicators.

Interestingly, we find that companies in sectors that are more likely to be exposed to transition risk also have a higher number of disclosed indicators (see figure 5). This could mean either that these companies are more prepared for the transition, that they are under greater pressure from investors and stakeholders to disclose their strategies, or that they are potential greenwashers.

3.3.3. Carbon emissions

To continue our exploratory understanding of greenwashing, we move beyond the sole assessment of disclosures and provide initial evidence of how the quality of disclosures relates to real activity measures. Hence, we obtain company-level data on total emissions and emission intensities from MSCI. We observe that the correlation between disclosure quality and all the emission measures is negative. The only exception is total scope 2 emissions (see figure 6). Hence, companies with higher-quality transition plans tend to have lower total emissions and emission intensities. The positive correlation with total scope 2 emissions, that is, indirect emissions from the generation of purchased electricity, steam, heating, and cooling consumed by the company, might suggest complexities in the transition process, such as increased reliance on electricity which may be sourced from non-renewable grids, or improved transparency in reporting. When we investigate the individual sectors, we observe a greater heterogeneity (see section S.11 for details). This confirms the general intuition of the paper that sector-specific investigations are a useful step for further explorations.

However, the preliminary evidence provided by our results should be further supported by future studies using larger samples and more in-depth analysis.

4. Discussion

Our results have important implications for a variety of stakeholders in the research and practice community.

The tool introduced bridges an important gap between climate change and LLM research. On the one hand, we address the call for a comprehensive evaluation of transition strategies [1–5]. On the other hand, we respond to the calls for more fine-granular LLM-based analyses [17] by extending prior work in the intersection of climate change and LLMs [14, 15, 28]. We also provide unique and novel expert-centric insights on tool evaluation, flexible usage, and possible extension. In this way, we highlight potential barriers and pathways to facilitate the use of automatic evaluations.

Our results for the CA100+ companies align with previous manual investigations of company transition planning. In line with previous work, we show that companies' transition strategies lack decision-useful information related to concrete action [30] and highlight the importance for companies and the financial sector to walk the talk [31, 32]. The negative correlation between company disclosure quality and emissions is both logical and alarming. It is logical because companies with lower emissions have societal, reputational, and potentially financial incentives to disclose their transition. They are likely aware of their position and will use this to advertise themselves. However, it is also alarming because it seems to confirm that companies with higher responsibility (that is, more polluting) seem to be more reluctant to disclose their transition efforts. A similar trend is found by [33] who show that the most emitting companies in the world usually do not participate in carbon reduction projects such as the Science Based Targets initiative (SBTi). See section S.1 for more insights.

5. Conclusion

To the best of our knowledge, this is the first paper to show how a structured expert-centric LLM-based tool can help assess transition strategies. This is of central importance given the growing amounts of unstructured transition-related information disclosed by companies. Furthermore, an automated, scalable, and consistent method can help company reporting, enhancing information availability and transparency for financial markets and financial regulators.

Future work can build on this flexible approach, focusing on sector-specific needs or applying it to more targeted problems. Thus, this work presents a basis for further enhancements of the automatic analysis of transition strategies.

Acknowledgments

This paper has received funding from the Swiss National Science Foundation (SNSF) under the project 'How sustainable is sustainable finance? Impact evaluation and automated greenwashing detection' (Grant Agreement No. 100018_207800), as well as partial funding from WWF Switzerland.

Data availability statement

The data that support the findings of this study and the code are openly available at the following URL/DOI: https://github.com/tobischimanski/transition_NLP.

Ethical statement

Human evaluation: In this work, all human evaluators are domain experts who have good knowledge about transition planning and gave their consent to participate in the study. The experts have full knowledge of the context and utility of the collected data. We adhered strictly to ethical guidelines, respecting the dignity, rights, safety, and well-being of all participants. The ethics committee of the University of Zurich qualified us as exempt from needing ethical approval.

Data privacy or bias: There are no data privacy issues or biases against certain demographics about the data collected from real-world applications and LLM generations. All artifacts we use are under a Creative Commons license. We also notice no ethical risks associated with this work.

Reproducibility statement: To ensure full reproducibility, we will disclose all codes and data used in this project, as well as the LLM generations. For OpenAI models, we use 'gpt-4-1106-preview'. We always fix the temperature to 0 when using APIs.

ORCID iDs

Chiara Colesanti Senni https://orcid.org/0000-0001-8029-4106

References

[1] Dikau S, Robins N, Smoleńska A, vant Klooster J and Volz U 2022 Net zero transition plans: a supervisory playbook for prudential authorities
[2] Rappe L, Haddad N and Jeanmart H 2023 Credibility analysis of the companys transition plan
[3] Reséndiz J L and Shrimali G 2023 Assessing the credibility of climate transition plans in the aviation sector *Discussion Paper* Oxford Sustainable Finance Group
[4] Zhou X and Shrimali G 2023 Assessing the credibility of climate transition plans in the power sector *Technical Report*
[5] Chan K J D, Cheung B and Shen L Y 2023 How to assess the credibility of corporate transition pathways? Application of the TCFDframework based on economic theory *Technical Report* Oxford Sustainable Finance (https://doi.org/10.2139/ssrn.4504017)
[6] Dunz N, Naqvi A and Monasterolo I 2019 Climate transition risk, climate sentiments, and financial stability in a stock-flow consistent approach *Climate Sentiments, and Financial Stability in a Stock-Flow Consistent Approach (April 1, 2019)* (https://doi.org/10.2139/ssrn.3520764)
[7] Battiston S, Dafermos Y and Monasterolo I 2021 Climate risks and financial stability (https://doi.org/10.1016/j.jfs.2021.100867)
[8] Diluiso F, Annicchiarico B, Kalkuhl M and Minx J C 2021 Climate actions and macro-financial stability: the role of central banks *Journal of Environmental Economics and Management* **110** 102548
[9] Duchin R, Gao J and Xu Q 2022 Sustainability or greenwashing: evidence from the asset market for industrial pollution *Available at SSRN 4095885* (https://doi.org/10.2139/ssrn.4095885)
[10] Daumas L 2023 Financial stability, stranded assets and the low-carbon transition-a critical review of the theoretical and applied literatures *Journal of Economic Surveys* **38** 601–716
[11] Gourier E and Mathurin H 2024 A greenwashing index *Available at SSRN* (https://doi.org/10.2139/ssrn.4715053)
[12] Hhne N *et al* 2021 Wave of net zero emission targets opens window to meeting the paris agreement *Nat. Clim. Change* **11** 820–2
[13] Luccioni A, Baylor E and Duchene N 2020 *Analyzing Sustainability Reports using Natural Language Processing* arXiv:2011.08073
[14] Bingler J A, Kraus M, Leippold M and Webersinke N 2024 How cheap talk in climate disclosures relates to climate initiatives, corporate emissions, and reputation risk *Journal of Banking & Finance* **164** 107191
[15] Schimanski T, Bingler J, Hyslop C, Kraus M and Leippold M 2023 *Climatebert-netzero: Detecting and Assessing net Zero and Reduction Targets.* arXiv:2310.08096
[16] Toetzke M, Stünzi A and Egli F 2022 Consistent and replicable estimation of bilateral climate finance *Nat. Clim. Change* **12** 897–900
[17] Sietsma A J, Ford J D and Minx J C 2023 The next generation of machine learning for tracking adaptation texts *Nat. Clim. Change*
[18] Brown T *et al* 2020 Language models are few-shot learners *Advances in Neural Information Processing Systems* **33** 1877–901
[19] Ouyang L *et al* 2022 Training language models to follow instructions with human feedback *Advances in Neural Information Processing Systems* **35** 27730–44
[20] Touvron H *et al* 2023 Llama: open and efficient foundation language models arXiv:2302.13971
[21] OpenAI. Gpt-4 technical report, 2023 (https://cdn.openai.com/papers/gpt-4.pdf)
[22] Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, Ishii E, Bang Y J, Madotto A and Fung P 2023 Survey of hallucination in natural language generation *ACMComputing Surveys* **55** 1–38
[23] Lewis P *et al* 2020 Retrieval-augmented generation for knowledge-intensive nlp tasks ed H Larochelle *et al Advances in Neural Information Processing Systems* 33, 9459–74(Curran Associates, Inc)
[24] Zhang T, Patil S G, Jain N, Shen S, Zaharia M, Stoica I and Gonzalez J E 2024 *Raft: Adapting Language Model to Domain Specific Rag*
[25] Saad-Falcon J, Khattab O, Potts C and Zaharia M 2024 Ares: an automated evaluation framework for retrieval-augmented generation systems
[26] Asai A, Wu Z, Wang Y, Sil A and Hajishirzi H 2023 *Self-rag: Learning to Retrieve, Generate, and Critique Through Self-Reflection*
[27] Schimanski T, Ni J, Kraus M, Ash E and Leippold M 2024 Towards faithful and robust llm specialists for evidence-based question-answering
[28] Ni J *et al* 2023 CHATREPORT: Democratizing sustainability disclosure analysis through LLM-based tools *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* ed Y Feng and E Lefever (Singapore) 21–51 December 2023. Association for Computational Linguistics.
[29] Leippold M *et al* 2024 Automated fact-checking of climate change claims with large language models (https://doi.org/10.2139/ssrn.4731802)
[30] Sokolowski A, Twigg S, McHugh E and Shamir A 2024 *The State of Play: 2023 Climate Transition Plan Disclosure*Carbon Disclosure Project
[31] Zentralbank Europäische 2022 Walking the talk: banks gearing up to manage risks from climate change and environmental degradation results of the 2022 thematic review on climate-related and environmental risks publications office
[32] Chen S 2022 Green investors and green transition efforts: talk the talk or walk the walk? *Available at SSRN 4254894* (https://doi.org/10.2139/ssrn.4254894)
[33] Watanabe K, Panagiotopoulos A and He S 2023 Assessing science-based corporate climate target-setting *MSCI*