Analyzing Sustainability Reports Using Natural Language Processing

Item Type	Preprint
Author	Alexandra Luccioni
Author	Emily Baylor
Author	Nicolas Duchene
Abstract	Climate change is a far-reaching, global phenomenon that will impact many aspects of our society, including the global stock market \cite{dietz2016climate}. In recent years, companies have increasingly been aiming to both mitigate their environmental impact and adapt to the changing climate context. This is reported via increasingly exhaustive reports, which cover many types of climate risks and exposures under the umbrella of Environmental, Social, and Governance (ESG). However, given this abundance of data, sustainability analysts are obliged to comb through hundreds of pages of reports in order to find relevant information. We leveraged recent progress in Natural Language Processing (NLP) to create a custom model, ClimateQA, which allows the analysis of financial reports in order to identify climate-relevant sections based on a question answering approach. We present this tool and the methodology that we used to develop it in the present article.
Date	2020-11-17
Library Catalog	arXiv.org
URL	http://arxiv.org/abs/2011.08073
Accessed	2/1/2024, 4:21:38 PM
Extra	arXiv:2011.08073 [cs]
Repository	arXiv
Archive ID	arXiv:2011.08073
Date Added	2/1/2024, 4:21:38 PM
Modified	2/1/2024, 4:21:38 PM

Tags:

/unread, Computer Science - Computation and Language, Computer Science - Machine Learning

Attachments

- arXiv.org Snapshot
- Full Text PDF

Arabic Mini-ClimateGPT : A Climate Change and Sustainability Tailored Arabic LLM

Item Type	Conference Paper

Author Sahal Mullappilly
Author Abdelrahman Shaker
Author Omkar Thawakar
Author Hisham Cholakkal
Author Rao Anwer
Author Salman Khan
Author Fahad Khan
Editor Houda Bouamor
Editor Juan Pino
Editor Kalika Bali

Abstract Climate change is one of the most significant challenges we face together as a society. Creating awareness and educating policy makers the wide-ranging impact of climate change is an essential step towards a sustainable future. Recently, Large Language Models (LLMs) like ChatGPT and Bard have shown impressive conversational abilities and excel in a wide variety of NLP tasks. While these models are close-source, recently alternative open-source LLMs such as Stanford Alpaca and Vicuna have shown promising results. However, these open-source models are not specifically tailored for climate related domain specific information and also struggle to generate meaningful responses in other languages such as, Arabic. To this end, we propose a light-weight Arabic Mini-ClimateGPT that is built on an open-source LLM and is specifically fine-tuned on a conversational-style instruction tuning curated Arabic dataset Clima500-Instruct with over 500k instructions about climate change and sustainability. Further, our model also utilizes a vector embedding based retrieval mechanism during inference. We validate our proposed model through quantitative and qualitative evaluations on climate-related queries. Our model surpasses the baseline LLM in 88.3% of cases during ChatGPT-based evaluation. Furthermore, our human expert evaluation reveals an 81.6% preference for our model's responses over multiple popular open-source models. Our open-source demos, models and curated instruction sets are available here : https://github.com/mbzuai-oryx/ClimateGPT

Date 2023-12
Short Title Arabic Mini-ClimateGPT
Library Catalog ACLWeb
URL https://aclanthology.org/2023.findings-emnlp.941
Accessed 2/1/2024, 3:50:52 PM
Place Singapore
Publisher Association for Computational Linguistics
Pages 14126–14136
Proceedings Title Findings of the Association for Computational Linguistics: EMNLP 2023
Conference Name Findings 2023
DOI 10.18653/v1/2023.findings-emnlp.941
Date Added 2/1/2024, 3:50:53 PM
Modified 2/1/2024, 3:50:55 PM

Tags:

/unread

Attachments

- Full Text PDF

---

chatClimate: Grounding Conversational AI in Climate Science

Item Type Preprint
Author Saeid Ashraf Vaghefi
Author Qian Wang
Author Veruska Muccione
Author Jingwei Ni
Author Mathias Kraus
Author Julia Bingler
Author Tobias Schimanski
Author Chiara Colesanti-Senni
Author Nicolas Webersinke
Author Christrian Huggel
Author Markus Leippold

Abstract Large Language Models (LLMs) have made significant progress in recent years, achieving remarkable results in question-answering tasks (QA). However, they still face two major challenges: hallucination and outdated information after the training phase. These challenges take center stage in critical domains like climate change, where obtaining accurate and up-to-date information from reliable sources in a limited time is essential and difficult. To overcome these barriers, one potential solution is to provide LLMs with access to external, scientifically accurate, and robust sources (long-term memory) to continuously update their knowledge and prevent the propagation of inaccurate, incorrect, or outdated information. In this study, we enhanced GPT-4 by integrating the information from the Sixth Assessment Report of the Intergovernmental (IPCC AR6), the most comprehensive, up-to-date, and reliable source in this domain. We present our conversational AI prototype, available at www.chatclimate.ai and demonstrate its ability to answer challenging questions accurately in three different QA scenarios: asking from 1) GPT-4, 2) chatClimate, and 3) hybrid chatClimate. The answers and their sources were evaluated by our team of IPCC authors, who used their expert knowledge to score the accuracy of the answers from 1 (very-low) to 5 (very-high). The evaluation showed that the hybrid chatClimate provided more accurate answers, highlighting the effectiveness of our solution. This approach can be easily scaled for chatbots in specific domains, enabling the delivery of reliable and accurate information.

Date 2023-04-28

Short Title	chatClimate
Library Catalog	arXiv.org
URL	http://arxiv.org/abs/2304.05510
Accessed	2/1/2024, 3:46:38 PM
Extra	arXiv:2304.05510 [cs]
Repository	arXiv
Archive ID	arXiv:2304.05510
Date Added	2/1/2024, 3:46:38 PM
Modified	2/1/2024, 3:46:38 PM

Tags:

/unread, Computer Science - Computation and Language

Attachments

- arXiv.org Snapshot
- Full Text PDF

CHATREPORT: Democratizing Sustainability Disclosure Analysis through LLM-based Tools

Item Type	Conference Paper
Author	Jingwei Ni
Author	Julia Bingler
Author	Chiara Colesanti-Senni
Author	Mathias Kraus
Author	Glen Gostlow
Author	Tobias Schimanski
Author	Dominik Stammbach
Author	Saeid Ashraf Vaghefi
Author	Qian Wang
Author	Nicolas Webersinke
Author	Tobias Wekhof
Author	Tingyu Yu
Author	Markus Leippold
Editor	Yansong Feng
Editor	Els Lefever
Abstract	In the face of climate change, are companies really taking substantial steps toward more sustainable operations? A comprehensive answer lies in the dense, information-rich landscape of corporate sustainability reports. However, the sheer volume and complexity of these reports make human

	analysis very costly. Therefore, only a few entities worldwide have the resources to analyze these reports at scale, which leads to a lack of transparency in sustainability reporting. Empowering stakeholders with LLM-based automatic analysis tools can be a promising way to democratize sustainability report analysis. However, developing such tools is challenging due to (1) the hallucination of LLMs and (2) the inefficiency of bringing domain experts into the AI development loop. In this paper, we introduce ChatReport, a novel LLM-based system to automate the analysis of corporate sustainability reports, addressing existing challenges by (1) making the answers traceable to reduce the harm of hallucination and (2) actively involving domain experts in the development loop. We make our methodology, annotated datasets, and generated analyses of 1015 reports publicly available. Video Introduction: https://www.youtube.com /watch?v=Q5AzaKzPE4M Github: https://github.com/EdisonNi-hku/chatreport Live web app: reports.chatclimate.ai
Date	2023-12
Short Title	CHATREPORT
Library Catalog	ACLWeb
URL	https://aclanthology.org/2023.emnlp-demo.3
Accessed	2/1/2024, 8:08:11 PM
Place	Singapore
Publisher	Association for Computational Linguistics
Pages	21–51
Proceedings Title	Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations
DOI	10.18653/v1/2023.emnlp-demo.3
Date Added	2/1/2024, 8:08:11 PM
Modified	2/1/2024, 8:08:11 PM

Tags:

/unread

Attachments

- Full Text PDF

ESG information extraction with cross-sectoral and multi-source adaptation based on domain-tuned language models

Item Type	Journal Article
Author	Jaeyoung Lee
Author	Misuk Kim

Abstract Along with future sustainability, factors such as environmental protection, co-prosperity, and ethical management are also gaining attention, resulting in the increased significance of the non-financial performance of companies. Accordingly, many institutions are evaluating the non-financial activities of companies using environmental, social, and governance (ESG) as evaluation indicators. However, the use of the current ESG rating as a reliable indicator has a limitation as the criteria are different for each evaluation institution and the ESG rating for the same company may differ when provided by different evaluation institutions. Accordingly, it is important to directly extract non-financial information generated by various media to provide objective ESG information, but the studies that focus on the need for automatic and direct extraction and classification of ESG information are limited. Therefore, this paper proposes a classifier that can discriminate ESG information. To train the ESG classifier, a dataset was constructed by manually labeling ESG data, and the ESG classifier recorded a classification accuracy of 86.66% for the 4-class classification problem of the constructed dataset. In addition, three application experiments were conducted to verify the usability of the proposed model. First, the results confirmed that the ESG classifier showed a significant performance of 83.96% accuracy for sectors not included in the training data. Second, the results qualitatively confirmed that the proposed model properly extracts ESG-related information from multi-source text data. Finally, the ESG dataset was additionally constructed through the pseudo-labeling technique, and the performance improvement of the data augmentation and classifier was verified. The ESG classifier proposed in this paper will provide appropriate ESG information to stakeholders. In future studies, we intend to extract and classify ESG-related information from more diverse types of data through an advanced ESG classifier. Further, we will use the information for investigating methods for reflecting the non-financial performance of companies in ESG ratings.

Date 2023-07-01
Library Catalog 8.5
Call Number 1
URL https://www.sciencedirect.com/science/article/pii/S0957417423002270
Accessed 2/1/2024, 4:11:34 PM
Volume 221
Pages 119726
Publication Expert Systems with Applications
DOI 10.1016/j.eswa.2023.119726
Journal Abbr Expert Systems with Applications
ISSN 0957-4174
Date Added 2/1/2024, 4:11:34 PM
Modified 2/1/2024, 4:11:35 PM

Tags:

/unread, ESG, Language model, Pseudo-labeling, Sustainability report, Transfer learning

Attachments

- Lee and Kim - 2023 - ESG information extraction with cross-sectoral and.pdf
- ScienceDirect Snapshot

ESGReveal: An LLM-based approach for extracting structured data from ESG reports

Item Type Preprint

Author Yi Zou
Author Mengying Shi
Author Zhongjie Chen
Author Zhu Deng
Author ZongXiong Lei
Author Zihan Zeng
Author Shiming Yang
Author HongXiang Tong
Author Lei Xiao
Author Wenwen Zhou

Abstract ESGReveal is an innovative method proposed for efficiently extracting and analyzing Environmental, Social, and Governance (ESG) data from corporate reports, catering to the critical need for reliable ESG information retrieval. This approach utilizes Large Language Models (LLM) enhanced with Retrieval Augmented Generation (RAG) techniques. The ESGReveal system includes an ESG metadata module for targeted queries, a preprocessing module for assembling databases, and an LLM agent for data extraction. Its efficacy was appraised using ESG reports from 166 companies across various sectors listed on the Hong Kong Stock Exchange in 2022, ensuring comprehensive industry and market capitalization representation. Utilizing ESGReveal unearthed significant insights into ESG reporting with GPT-4, demonstrating an accuracy of 76.9% in data extraction and 83.7% in disclosure analysis, which is an improvement over baseline models. This highlights the framework's capacity to refine ESG data analysis precision. Moreover, it revealed a demand for reinforced ESG disclosures, with environmental and social data disclosures standing at 69.5% and 57.2%, respectively, suggesting a pursuit for more corporate transparency. While current iterations of ESGReveal do not process pictorial information, a functionality intended for future enhancement, the study calls for continued research to further develop and compare the analytical capabilities of various LLMs. In summary, ESGReveal is a stride forward in ESG data processing, offering stakeholders a sophisticated tool to better evaluate and advance corporate sustainability efforts. Its evolution is promising in promoting transparency in corporate reporting and aligning with broader sustainable development aims.

Date 2023-12-25

Short Title	ESGReveal
Library Catalog	arXiv.org
URL	http://arxiv.org/abs/2312.17264
Accessed	2/1/2024, 3:53:29 PM
Extra	arXiv:2312.17264 [cs]
Repository	arXiv
Archive ID	arXiv:2312.17264
Date Added	2/1/2024, 3:53:30 PM
Modified	2/1/2024, 3:53:32 PM

Tags:

/unread, Computer Science - Computation and Language, Computer Science - Information Retrieval

Attachments

- arXiv.org Snapshot
- Full Text PDF

EXAMINING LLM'S AWARENESS OF THE UNITED NA- TIONS SUSTAINABLE DEVELOPMENT GOALS (SDGS)

Item Type	Journal Article
Author	Mehdi Bahrami
Author	Ramya Srinivasan
Date	2023
Language	en
Library Catalog	Zotero
Date Added	2/1/2024, 3:59:18 PM
Modified	2/1/2024, 3:59:19 PM

Tags:

/unread

Attachments

- Bahrami and Srinivasan - 2023 - EXAMINING LLM'S AWARENESS OF THE UNITED NA- TIONS .pdf

---

FABULA: Intelligence Report Generation Using Retrieval-Augmented Narrative Construction

Item Type	Preprint
Author	Priyanka Ranade
Author	Anupam Joshi
Abstract	Narrative construction is the process of representing disparate event information into a logical plot structure that models an end to end story. Intelligence analysis is an example of a domain that can benefit tremendously from narrative construction techniques, particularly in aiding analysts during the largely manual and costly process of synthesizing event information into comprehensive intelligence reports. Manual intelligence report generation is often prone to challenges such as integrating dynamic event information, writing fine-grained queries, and closing information gaps. This motivates the development of a system that retrieves and represents critical aspects of events in a form that aids in automatic generation of intelligence reports. We introduce a Retrieval Augmented Generation (RAG) approach to augment prompting of an autoregressive decoder by retrieving structured information asserted in a knowledge graph to generate targeted information based on a narrative plot model. We apply our approach to the problem of neural intelligence report generation and introduce FABULA, framework to augment intelligence analysis workflows using RAG. An analyst can use FABULA to query an Event Plot Graph (EPG) to retrieve relevant event plot points, which can be used to augment prompting of a Large Language Model (LLM) during intelligence report generation. Our evaluation studies show that the plot points included in the generated intelligence reports have high semantic relevance, high coherency, and low data redundancy.
Date	2023-10-20
Short Title	FABULA
Library Catalog	arXiv.org
URL	http://arxiv.org/abs/2310.13848
Accessed	2/1/2024, 3:09:38 PM
Extra	arXiv:2310.13848 [cs]
DOI	10.1145/3625007.3627505
Date Added	2/1/2024, 3:09:38 PM
Modified	2/1/2024, 3:09:39 PM

Tags:

/unread, Computer Science - Information Retrieval

Attachments

- arXiv.org Snapshot
- Full Text PDF

Glitter or Gold? Deriving Structured Insights from Sustainability Reports via Large Language Models

Item Type	Preprint
Author	Marco Bronzini
Author	Carlo Nicolini
Author	Bruno Lepri
Author	Andrea Passerini
Author	Jacopo Staiano
Abstract	Over the last decade, several regulatory bodies have started requiring the disclosure of non-financial information from publicly listed companies, in light of the investors' increasing attention to Environmental, Social, and Governance (ESG) issues. Publicly released information on sustainability practices is often disclosed in diverse, unstructured, and multi-modal documentation. This poses a challenge in efficiently gathering and aligning the data into a unified framework to derive insights related to Corporate Social Responsibility (CSR). Thus, using Information Extraction (IE) methods becomes an intuitive choice for delivering insightful and actionable data to stakeholders. In this study, we employ Large Language Models (LLMs), In-Context Learning, and the Retrieval-Augmented Generation (RAG) paradigm to extract structured insights related to ESG aspects from companies' sustainability reports. We then leverage graph-based representations to conduct statistical analyses concerning the extracted insights. These analyses revealed that ESG criteria cover a wide range of topics, exceeding 500, often beyond those considered in existing categorizations, and are addressed by companies through a variety of initiatives. Moreover, disclosure similarities emerged among companies from the same region or sector, validating ongoing hypotheses in the ESG literature. Lastly, by incorporating additional company attributes into our analyses, we investigated which factors impact the most on companies' ESG ratings, showing that ESG disclosure affects the obtained ratings more than other financial or company data.
Date	2024-01-16
Short Title	Glitter or Gold?
Library Catalog	arXiv.org
URL	http://arxiv.org/abs/2310.05628
Accessed	2/1/2024, 4:15:27 PM
Extra	arXiv:2310.05628 [cs]
Repository	arXiv
Archive ID	arXiv:2310.05628
Date Added	2/1/2024, 4:15:27 PM
Modified	2/1/2024, 4:15:29 PM

Tags:

/unread, Computer Science - Computation and Language, Computer Science - Computational Engineering, Finance, and Science, Computer Science - Computers and Society

Attachments

- arXiv.org Snapshot
- Full Text PDF

GreenAI – An NLP approach to ESG financing

Item Type	Conference Paper
Author	Nicolaas Ruberg
Author	Rafael Pereira
Author	Mauro Stein
Abstract	Environmental, Social, and Governance (ESG) factors are critical for investors and financing institutions like the Brazilian Development Bank (BNDES). Such institutions are currently working on setting up a framework to assess companies' ESG factors in their financing evaluation. In this study, we identify an opportunity to use Natural Language Processing (NLP) to improve the framework. This opportunity stems from the fact that the key documents for ESG analysis, such as the company's activity report (RAA), Environmental Impact Study (EIA), and Environmental Impact Report (RIMA), undergo manual screening and decomposition whilst being analyzed by specialists. By incorporating NLP, we aim to automate the classification of text passages from these reports and enhance the efficiency of the analysis process.
Date	2023-08-06
Library Catalog	ResearchGate
Pages	37-48
DOI	10.5753/bwaif.2023.229922
Date Added	2/1/2024, 4:36:06 PM
Modified	2/1/2024, 4:36:07 PM

Tags:

/unread

Attachments

- Full Text
- ResearchGate Link

Harnessing Large Language Models with Neo4j

Item Type	Blog Post
Author	Oskar Hane
Abstract	Episode 1—Exploring Real-World Use Cases
Date	2023-06-15T06:48:04.569Z
Language	en
URL	https://medium.com/neo4j/harnessing-large-language-models-with-neo4j-306ccbdd2867
Accessed	2/1/2024, 3:55:02 PM
Blog Title	Neo4j Developer Blog
Date Added	2/1/2024, 3:55:02 PM
Modified	2/1/2024, 3:55:04 PM

Tags:

/unread

Attachments

- Snapshot

Harnessing NLP for Effective ESG Reporting: Insights for SMEs

Item Type	Web Page
URL	https://positivethinking.tech/insights/nlp-esg-reporting-smes/
Accessed	2/1/2024, 4:17:35 PM
Date Added	2/1/2024, 4:17:35 PM
Modified	2/1/2024, 4:17:35 PM

Tags:

/unread

Attachments

- Harnessing NLP for Effective ESG Reporting: Insights for SMEs

---

How Generative AI can build an organization’s ESG roadmap

Item Type	Web Page
Abstract	EY highlights how Gen AI offers transformative opportunities for organizations in their sustainability journey. Learn more about role of AI in ESG.
Language	en-IN

URL	https://www.ey.com/en_in/ai/how-generative-ai-can-build-an-organization-s-esg-roadmap
Accessed	2/1/2024, 4:25:30 PM
Date Added	2/1/2024, 4:25:30 PM
Modified	2/1/2024, 4:25:32 PM

Tags:

/unread

How will AI text generation and processing impact sustainability reporting? Critical analysis, a conceptual framework and avenues for future research

Item Type	Journal Article
Author	Charl De Villiers
Author	Ruth Dimes
Author	Matteo Molinari
Abstract	Purpose The ability of generative artificial intelligence (AI) tools such as ChatGPT to produce convincing, human-like text has major implications for the future of corporate reporting, including sustainability reporting. As the importance of sustainability reporting continues to grow, this study aims to critically analyse the benefits and pitfalls of automated text generation and processing. Design/methodology/approach This study develops a conceptual framework to delineate the field, assess the implications and form the basis for the generation of research questions. This study uses Alvesson and Deetz's critical framework, considering insight (a review of literature and practice in the field), critique (consideration of the influences on the production and use of non-financial information and the implications for assurers of such information) and transformative redefinition (considering the implications of generative AI for sustainability reporting and proposing a research agenda). Findings This study highlights the implications of generative AI for sustainability accounting, reporting, assurance and report usage, including the risk of AI facilitating greenwashing, and the importance of more research on the use of AI for these matters. Practical implications The paper highlights to stakeholders the implications of AI for all aspects of sustainability reporting, including accounting, reporting, assurance and usage of reports. Social implications The implications of AI need to be understood in society, which this paper facilitates. Originality/value This study critically analyses the potential use of AI for sustainability reporting, construct a conceptual framework to delineate the field and develop a research agenda.
Date	2024-01-02
Language	en
Short Title	How will AI text generation and processing impact sustainability reporting?
Library Catalog	DOI.org (Crossref)

URL	https://www.emerald.com/insight/content/doi/10.1108/SAMPJ-02-2023-0097/full/html
Accessed	2/1/2024, 4:31:13 PM
Volume	15
Pages	96-118
Publication	Sustainability Accounting, Management and Policy Journal
DOI	10.1108/SAMPJ-02-2023-0097
Issue	1
Journal Abbr	SAMPJ
ISSN	2040-8021, 2040-8021
Date Added	2/1/2024, 4:31:13 PM
Modified	2/1/2024, 4:31:14 PM

Tags:

/unread

Attachments

- De Villiers et al. - 2024 - How will AI text generation and processing impact .pdf

---

Introducing BloombergGPT, Bloomberg's 50-billion parameter large language model, purpose-built from scratch for finance | Press | Bloomberg LP

Item Type	Newspaper Article
Abstract	BloombergGPT is a 50-billion parameter large language model that was purpose-built from scratch for finance.
Language	en-US
Library Catalog	www.bloomberg.com
URL	https://www.bloomberg.com/company/press/bloomberggpt-50-billion-parameter-llm-tuned-finance/
Accessed	2/1/2024, 4:00:48 PM
Section	Bloomberg Professional Services
Publication	Bloomberg L.P.
Date Added	2/1/2024, 4:00:48 PM
Modified	2/1/2024, 4:00:49 PM

Tags:

/unread

Microsoft Cloud for Sustainability: Generative AI, Scope 3 Data Capabilities for ESG Reporting Coming Soon

Item Type	Web Page
Author	Joe Panettieri
Abstract	Microsoft Cloud for Sustainability preps generative AI (artificial intelligence) capabilities for ESG reporting. Microsoft Copilot & Scope 3 data support coming soon.
Date	2024-01-31T14:10:18+00:00
Language	en-US
Short Title	Microsoft Cloud for Sustainability
URL	https://sustainabletechpartner.com/topics/cloud-services/microsoft-cloud-for-sustainability-generative-ai-scope-3-data-capabilities-for-esg-reporting-coming-soon/
Accessed	2/1/2024, 4:18:08 PM
Website Title	Sustainable Tech Partner for Green IT Service Providers
Date Added	2/1/2024, 4:18:08 PM
Modified	2/1/2024, 4:18:09 PM

Tags:

/unread

Attachments

- Snapshot

ReportDVSJan2023_AnalysingSustainabilityReportsUsingML.pdf

Item Type	Attachment
URL	https://www.administration-numerique-suisse.ch/application/files/8316/8129/0720/ReportDVSJan2023_AnalysingSustainabilityReportsUsingML.pdf
Accessed	2/1/2024, 4:02:32 PM
Date Added	2/1/2024, 4:02:32 PM
Modified	2/1/2024, 4:02:32 PM

Top 7 Sustainability Reporting Standards - Comparing Disclosure Frameworks

Item Type	Web Page
Author	Brightest

Abstract	Brightest outlines key sustainability reporting frameworks and standards for disclosing your environmental footprint and ESG performance. Understand and compare sustainability reporting standards with this easy-to-follow guide
Language	en-US
URL	https://www.brightest.io/sustainability-reporting-standards
Accessed	2/1/2024, 4:19:58 PM
Website Title	Brightest
Date Added	2/1/2024, 4:19:58 PM
Modified	2/1/2024, 4:19:59 PM

Tags:

/unread