

# Climate Finance Text Summarization

## AUTHORS

Sophie Logan, MSCAPP 2023  
Rodrigo Salazar, MSCAPP 2023

University of Chicago  
Department of Computer Science

Climate finance is an important and complex topic that involves the mobilization and deployment of financial resources to address climate change. Given the complexity and the speed with which the climate crisis must be addressed, having ways to summarize the climate finance landscape is beneficial. A broad understanding of the different climate financing mechanisms would help policymakers and investors optimize decision making. In 2009, the world committed to \$100 billion in climate financing per year at the UN Summit in Copenhagen. Allowing stakeholders to quickly summarize key climate and financial data will help them make informed decisions about how to deploy that capital going forward.

Climate finance texts can be technical and difficult to understand, especially for people who are not experts in the field. Summarizing climate finance texts can be beneficial for the following reasons:

- **Accessibility:** Summarizing climate finance texts can make the information more accessible to a wider audience, including policymakers, practitioners, and the general public
- **Engagement:** Summarizing climate finance texts can help to engage a broader range of stakeholders, including those who may not have a technical background in the field
- **Action:** Summarizing climate finance texts can help to translate complex information into actionable insights that can be used to inform policy decisions and actions
- **Transparency:** Summarizing climate finance texts can help to increase transparency and accountability in the management and deployment of climate finance resources, which can be especially important in engaging people who historically have not been part of the climate finance process

Given our concern about the climate crisis, we would like to use our project as a way to combat some of the issues facing many stakeholders working on climate who may not have the time and bandwidth to engage with the volume of climate texts available. Some key players might be policymakers, media, and people at multilateral development banks.

## METHODOLOGY

We leverage data from the *Climate Policy Initiative*, an organization that provides analysis on climate finance to governments and financial institutions. We ingest approximately 500 texts on various climate finance topics. Additionally, we collect the human summaries for those texts.

We use two pre-trained summarization models to create our own summaries of the texts. They are both BART models from Hugging Face. The first is a Distilbart CNN 6-6 model and the second is a CNN SAMSUM model. The CNN 6-6 model is trained with over 500k news articles with summarizations. The SAMSUM dataset contains 1dk messenger-like conversations with summaries and is one of the most popular summarization models with almost 1 million downloads last month.

We leverage another model, topic modeling, to evaluate our summarization. We use LDA to find the topics generated for the full text, the human summary, and the model generated summaries and compare the intersections of those topics for one measure of summary accuracy.

To evaluate our results, we treat the human summaries as "ground truth" data. We assume that these summaries are what a good summary should be like. We use several methods to evaluate our models' performance including:

- **Bleu:** *Bilingual Evaluation Understudy* measures precision and how much the word (and/or n-grams) appeared in the text vs. the summary
  - **Meteor:** *Metric for Evaluation of Translation with Explicit Ordering* considers unigram matching, stemming and word order similarity
- These are different ways of assessing how well the summary captures the most important topics and themes in a text.

## RELATED LITERATURE

- Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stryanov, V., & Zettlemoyer, L. (Year). "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension." 2019.
- Giliva, B., Mochol, I., Besek, M., & Wawer, A. "SAMSUM Corpus: A Human-annotated Dialogue Dataset for Abstractive Summarization." Proceedings of the 2nd Workshop on New Frontiers in Summarization, Association for Computational Linguistics, 2019.
- Abdoljilbi, Samir, and Houda Bouazma. "An exploration of automatic text summarization of financial reports." Proceedings of the Third Workshop on Financial Technology and Natural Language Processing. 2021.

## RESULTS

We find that overall, SAMSUM performs better by both meteor and bleu accuracy measures. Human summaries perform better on all measures, which makes sense. We corroborate these findings with our qualitative assessment that the SAMSUM summaries are better than the CNN 6-6 summaries, with one example being detailed here. Despite being from a smaller dataset (1dk vs 500k), the SAMSUM model summaries are more succinct and clear.

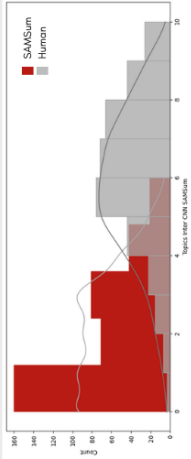
The bleu and meteor accuracies measure different aspects of the summaries. Since bleu measures prevalence, it might miss some synonym measures or other broad themes. Since meteor measures complex semantic topics it can be augmented with the simple bleu measures. These differences help us get a big picture.

We use the intersection of modeled topics as another way to estimate how close are texts are. We find that human summaries have higher intersections of topics with the original texts, which we would expect. We find a healthy amount of overlap between SAMSUM and the original text.

COUNT OF MAXIMUM SCORE FOR SUMMARIES BY MODEL AND ACCURACY TYPE



TOPIC PREVALENCE AFTER SUMMARIZATION



## EXAMPLE MODEL SUMMARY

PUBLICATION TITLE: **IMPACT AND HIGHLIGHTS FROM CPI'S CLIMATE FINANCE TRACKING PROGRAM**

### Full report excerpt

Climate Finance Tracking Program Impact and Highlights in 2022 How is CPI's Climate Finance Tracking Program Unique? CPI's Climate Finance Tracking Program provides the most comprehensive climate finance flows data across the public and private sectors. **We bring together data from 20+ public and private data sources.** Advanced data science techniques are applied across the board for data harmonization and consistency. Our impartial, rigorously researched evidence is trusted to shape international agendas.

How is CPI's Climate Finance Tracking Program Unique? 70+ Knowledge Products 12+ Sectors/Themes 8+ Countries Landscapes, methodological briefs, taxonomies, case studies, data dashboards, consulting reports, technical submissions, deep dives Renewables, energy access, efficiency, coal, infrastructure, methane, air pollution, cooling, AFOLU, nature, biodiversity, adaptation India, Brazil, Indonesia, South Africa, Kenya, Burkina Faso, Nigeria, Ethiopia

### Human Summary

Our 2022 Highlights and Impact Report outlines the impact and key milestones achieved by our climate finance tracking program in 2022 and beyond. It also highlights our contribution in advancing data-driven analysis, influencing climate finance policy, and amplifying knowledge.

Highlights include:

- In 2022, we published 23 climate finance tracking knowledge products across six sectors and in more than seven countries and regions leading to a 3x increase in projects compared to 2019.
- **We brought together data from 20+ public and private sources** and applied robust methodology, rigor for data standardization and analysis.
- 70+ organizations including high-level policy makers, research organizations, and top-tier media outlets cited CPI's landscape reports.
- We worked directly alongside public and private financial institutions, coalitions, governments, and climate champions to enable investments where they are needed most.

### CNN 6-6 Summary

Climate finance tracking program impact and highlights in 2022/23 years of data how is cpi's climate finance tracking program . The program provides the most comprehensive climate finance flows data across the public and private sectors **we bring together data from 20+ public and private data sources** . advanced data science techniques are applied across the board for data harmonization

### SAMSUM Summary

CPI's climate finance tracking program provides the most comprehensive climate finance flows data across the public and private sectors. **It brings together data from 20+ data sources.** CPI signed new data exchange deals with industry leaders like global water intelligence and accessed data from more than 20 sources.