Data collection method:

To address our research question, we apply the following datasets in a structured, and multi-step approach:

SDG Model

Abdul-Manan Sadick from Deakin University developed a Natural Language Processing (NLP) model for classifying text related to the United Nations Sustainable Development Goals (SDGs). This model utilizes a fine-tuned bert-base-uncased pre-trained model, trained on the publicly available OSDG Community Dataset (OSDG-CD).

Texts within the training set were evaluated by volunteers on the OSDG-CP, who assessed each text’s relevance to specific SDGs. Volunteers either accepted or rejected suggested labels based on their background knowledge. To ensure quality, each text was validated by up to nine different volunteers, with all texts in the public release validated by at least three different volunteers.

The model was applied to 3,092 speeches from the UN General Assembly (2008-2023), scoring the frequency of the first 16 SDGs mentioned in each speech. The probabilities sum to 1, indicating proportions of overall speech mentions.

To improve accuracy, a non-SDG category was added by ensembling a second model. This model, based on bert-base-uncased, included a synthetic sample of societal topics unrelated to SDGs, maintaining the same tokenizer as SDG-BERT.

The refined model demonstrates promising results, achieving an average weighted accuracy of 88%, with the lowest accuracy being 65% for one SDG. This variation is expected due to overlapping topics among certain SDGs.

In addition to developing our model, we enrich the analysis by incorporating two external datasets:

1. SDG Index of 2023:

The 2023 SDG Index offers an annual assessment of progress towards the SDGs across all 193 UN member states. Countries are ranked based on their overall performance score, which reflects accumulative progress towards achieving all 17 SDGs (Sachs et al., 2023). Scores are shown on a scale from 0 to 100 and reflect the percentage of progress made toward the best performance (Sachs et al., 2023). The SDG Index thus acts as a comprehensive metric for assessing each country's relative distance to SDG targets using the most recent data available. To reduce bias due to missing data, countries with more than 20% of indicator data missing are not assigned an overall score or rank. Accordingly, the 2023 SDG Index covers only 166 countries (Sachs et al., 2023).

The index data comes from a combination of official and non-official sources. Roughly two-thirds of the data are drawn from major international organizations such as the World Bank, WHO, UNICEF, etc., which apply rigorous validation procedures. The remaining one-third of the data originated from less conventional sources, including household surveys, for example, Gallup World Poll, and civil society organizations (e.g., Oxfam, Reporters Without Borders, etc.) (Sachs et al., 2023).

To address the issue of data gaps, we applied the most recently available score from the index in cases where a country did not receive a current overall performance score. This dataset provides esteemed insight into the discrepancy between rhetorical commitment to the SDGs (e.g., speech content) and actual implementation progress.

1. GDP Rankings Based on Purchasing Power Parity (PPP) of 2023:

For further analysis, we draw on the World Bank's “GDP Ranking, PPP-based” dataset. This dataset provides country-level rankings of Gross Domestic Product (GDP) adjusted for Purchasing Power Parity (PPP), which accounts for price level differences across countries and enables more accurate international economic comparisons (World Bank, 2024).

While not central to our main research question, the inclusion of GDP (PPP-based) data allows us to consider how economic capacity may relate to the level of SDG implementation. By comparing GDP data with SDG Index scores, we aim to explore potential correlations between economic performance and sustainability outcomes across different regions and countries.

Bibliography:

Sachs, J. D., Lafortune, G., Fuller, G., & Drumm, E. (2023). Sustainable Development Report 2023: Implementing the SDG Stimulus. Cambridge University Press. <https://sdgtransformationcenter.org/reports/sustainable-development-report-2023>

World Bank. (2024, December 17). GDP ranking, PPP based (Version 9) [Data set]. World Bank Group. <https://datacatalog.worldbank.org/search/dataset/0038129/GDP-ranking--PPP-based>