On the performativity of SDG classifications in large bibliometric databases

Matteo Ottaviani[1] Stephan Stahlschmidt[12]

[1]German Centre for Higher Education Research and Science Studies (DZHW), Schützenstr. 6a, 10117 Berlin (Germany)

[2]Universidad de Granada, Gran Vía 48, 18071, Granada, Spain

Motivation & Approach

Research Objective

Our study aims to identify the underlying reasons for the observed mismatch between the diverse SDG classifications in bibliometric databases [1]. Large Language Models (LLMs) are applied to learn the particular perspective each classification imposes onto the science system.

LLMs as magnifiers of data biases

LLMs are known to be affected by biases present in the underlying data, which we utilise to learn about the different perspectives each SDG classification entails. As LLMs add a further layer of semantic associations they serve as a magnifier of the structural differences in the SDG classifications.

Introduction

Bibliometric databases, as crucial digital infrastructures, enable bibliometric analyses and impact assessments. However, it is imperative to realize that these data structures are not neutral. They embody a performative nature, rooted in distinct understandings of the science system and value attributions [2, 3].

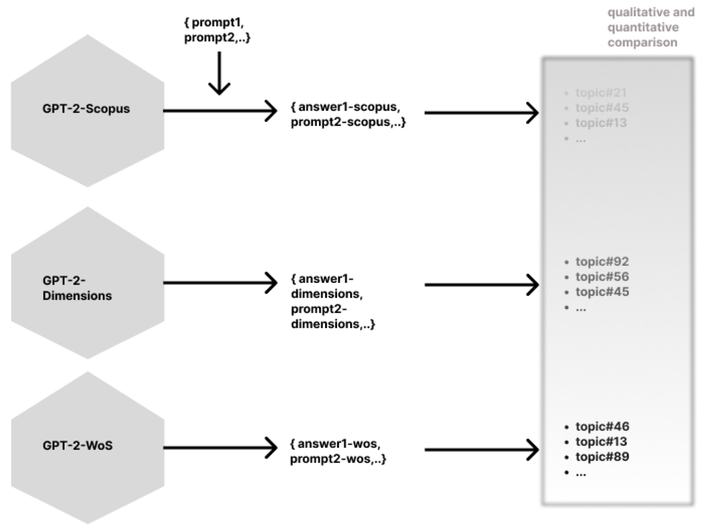
Recently political interest has focused on the science system's contribution to the United Nations' Sustainable Development Goals (SDGs). As a means to assess societal relevance and attribute research impact, the SDGs serve as a powerful framework [4, 5] and major bibliometric databases providers have followed suit in linking their indexed publications to the SDGs.

However, these classifications have been observed to be riddled with discrepancies [1] posing questions on what research is attributed a societal impact via a SDG classification and what research is not.

Research Design

Our methodology encompasses four pivotal stages:

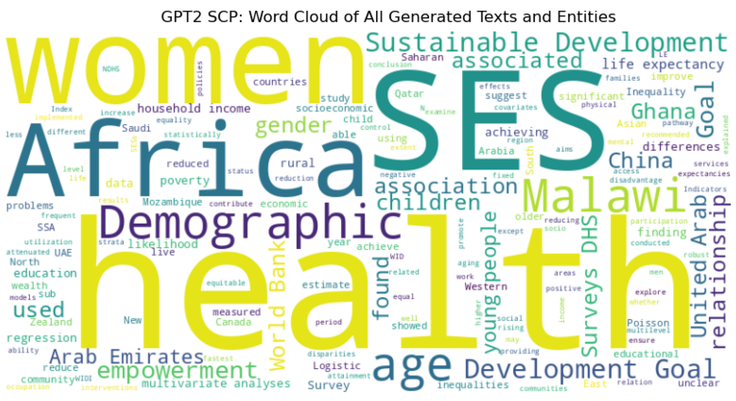
1. **Data Collection**: We draw a corpus of jointly indexed publications in Web of Science, Scopus, and Dimensions (14 Million publications from 2012 to 2020) to control for differences in the overall data coverage among the databases.
2. **Subdivision by SDG classification**: Within the jointly indexed publications we extract three varying subsets of publications attributed to SDG 10 by each of three SDG classifications to focus on the classification effect.
3. **LLM Fine-tuning**: A generic GPT2 model is fine-tuned in parallel on the three subsets via the respective abstracts to encode the semantic associations implied by each classification.
4. **Text Analysis**: The three fine-tuned LLMs are subjected to a series of prompts. Analyzing the generated responses, we endeavor to discern linguistic features and differences characteristic of each classification.



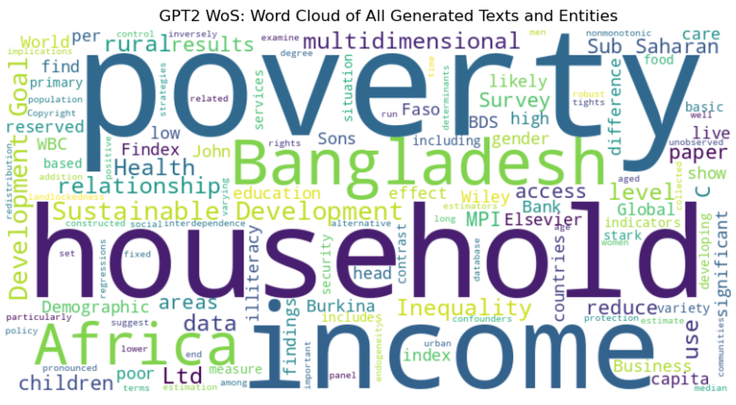
Schematic illustration of the research design.

Illustrative findings

Decoding for generative summarization has been implemented via *contrastive search*[6] balancing likelihood and context for text generation. The LLMs for Scopus and WoS were prompted by: "*The Sustainable Development Goal 10 is*" and the summaries converted into word clouds:



SDG 10 foci in Scopus classification.



SDG 10 foci in WoS classification.

Conclusions

The SDG classifications of WoS and Scopus provide each a different perspective on what constitutes SDG 10. Bibliometric classifications, while striving to offer objective measures, seem to present a specific focus, which is crucial in the attribution of social relevance via SDG classifications. Depending on the applied classification, scientists and institutions working in the aforementioned fields might, or might not, be able to empirically underline their impact to policy makers.

Key Findings

Large Language Models are sensitive to differences in SDG classifications and provide insights into the varied perspectives inscribed into the classifications. They further show high sensitivity to model architecture, fine-tuning process and decoding strategy.

Ceteris paribus the two SDG classifications not only differ in the number of publications assigned to SDG 10 (40 k for Scopus and 140 k for Web of Science), but also differ on the classified substance. WoS focuses predominately on an economic perspective, while Scopus offers a more diverse picture emphasising also health, gender and geographical aspects.

LLMs have been instrumental in unearthing and understanding these perspectives. The pre-trained GPT2 model, while smaller, offers several computational advantages; it is suitable for fine-tuning and generating scientific text. Moreover, contrary to more complex LLMs (e.g. *Falcon*), it owns no knowledge about SDGs in general by default. That is an essential feature for informing about data biases after fine-tuning.

References

[1] Caroline S Armitage, Marta Lorenz, and Susanne Mikki. Mapping scholarly publications related to the sustainable development goals: Do independent bibliometric approaches get the same results? *Quantitative Science Studies*, 1(3):1092–1108, 2020.

[2] Richard Whitley. *The intellectual and social organization of the sciences.* Oxford University Press, USA, 2000.

[3] Peter Vinkler. Bibliometric features of some scientific subfields and the scientometric consequences therefrom. *Scientometrics*, 14(5-6):453–474, 1988.

[4] Manoranjan Mishra et al. A bibliometric analysis of sustainable development goals (sdgs): a review of progress, challenges, and opportunities. *Environment, Development and Sustainability*, pages 1–43, 2023.

[5] Christine Meschede. The sustainable development goals in scientific literature: A bibliometric overview at the meta-level. *Sustainability*, 12(11):4461, 2020.

[6] Yixuan Su, Tian Lan, Yan Wang, Dani Yogatama, Lingpeng Kong, and Nigel Collier. A contrastive framework for neural text generation. *Advances in Neural Information Processing Systems*, 35:21548–21561, 2022.

Contact Information

ottaviani@dzhw.eu