Sentiment Analysis

## Introduction

We have experimented with 4 sentiment analysis (SA) models. In order to compare them, a sample of 140 publications has been generated.

This sample contains 20 publications randomly selected for each of the following SDG 3,5,7,8,10,11,13. Each publication was manually annotated as either positive, neutral or negative depending on the impact of the digital applications mentioned in the abstract.

Two of the 4 models are of general purpose, which means that they compute a sentiment score for a given text. We make the assumption that the sentiment of text is correlated to the impact of the digital technology on a given SDG.

These models are the following:

* [Vader](https://github.com/cjhutto/vaderSentiment" \l "features-and-updates) (Hutto & Gilbert, 2014), a rule-based tool using a lexicon, designed primarily for social media.
* [Base](https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base-sentiment) (Barbieri et al., 2022), a pretained-model based on neural network language model RoBERTa (Liu et al., 2019) , trained on 198M tweets.

The last 2 models are aspect-based, which means that the sentiment score is computed for a given noun-phrase. These models allow theoretically for a more precise analysis since it is then, possible to compute the sentiment on the digital keywords present in the abstract.

These models are the following:

* [SentiBigNomics](https://github.com/sergioconsoli/SentiBigNomics) (Barbaglia L et al., 2021), a sentiment analysis tool designed for economic texts utilizing a lexicon.
* [deberta-asba](https://huggingface.co/yangheng/deberta-v3-base-absa-v1.1?text=%5BCLS%5D+when+tables+opened+up%2C+the+manager+sat+another+party+before+us.+%5BSEP%5D+manager+%5BSEP%5D) (Yang et al., 2021), a pre-trainted model based on the neural-network language model DeBERTa.

None of these models are trained specifically to perform a sentiment analysis on scientific publications.

## Evaluation

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Tool** | **Sentiment** | **Precision** | **Recall** | **F1score** | **correct** | **manual** | **annotated** | **accuracy** |
| **Vader** | **Neg** | 0.08 | 0.25 | 0.12 | 1 | 4 | 13 | 0.52 |
|  | **Neu** | 0.32 | 0.13 | 0.19 | 7 | 52 | 22 |
|  | **Pos** | 0.62 | 0.77 | 0.69 | 65 | 84 | 105 |
| **deberta\_asba** | **Neg** | 0.33 | 0.5 | 0.4 | 2 | 4 | 6 | 0.69 |
|  | **Neu** | 0.6 | 0.54 | 0.57 | 28 | 52 | 47 |
|  | **Pos** | 0.76 | 0.79 | 0.77 | 66 | 84 | 87 |
| **base** | **Neg** | 0.4 | 0.5 | 0.44 | 2 | 4 | 5 | 0.37 |
|  | **Neu** | 0.37 | 0.96 | 0.54 | 50 | 52 | 134 |
|  | **Pos** | 0 | 0 | 0 | 0 | 84 | 1 |
| **SentiBigNomics** | **Neg** | 0 | 0 | 0 | 0 | 4 | 3 | 0.46 |
|  | **Neu** | 0.41 | 0.87 | 0.55 | 45 | 52 | 111 |
|  | **Pos** | 0.77 | 0.24 | 0.36 | 20 | 84 | 26 |

Precision (for a given sentiment): total correct annotation / total annotated

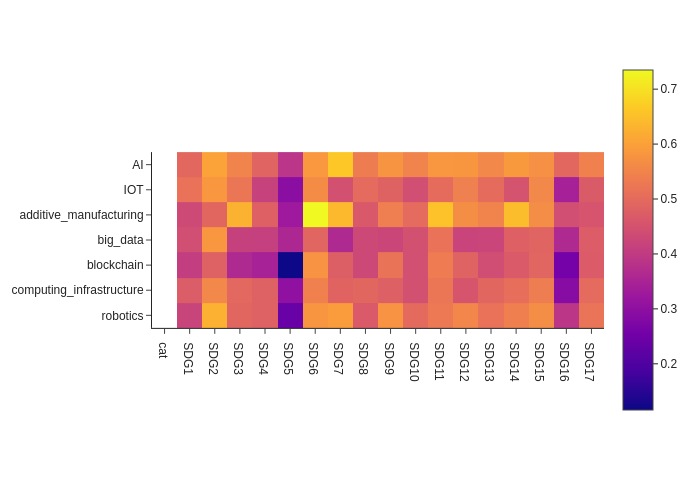
Recall (for a given sentiment): total correct annotation / total manually annotated

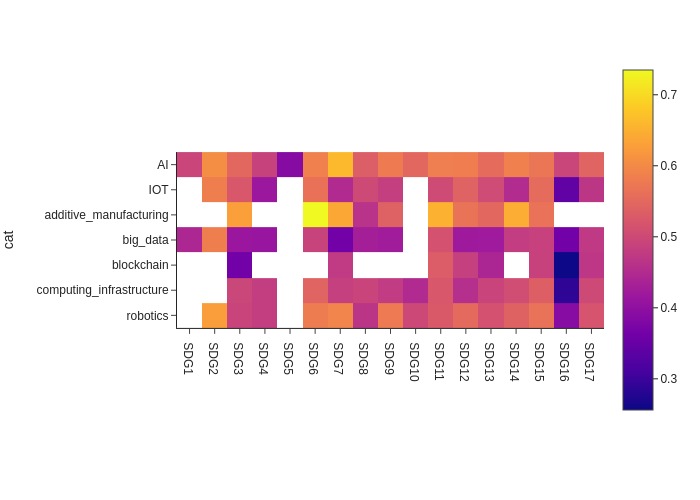
F1-score: 2\*precision\*recall / (precision+recall)

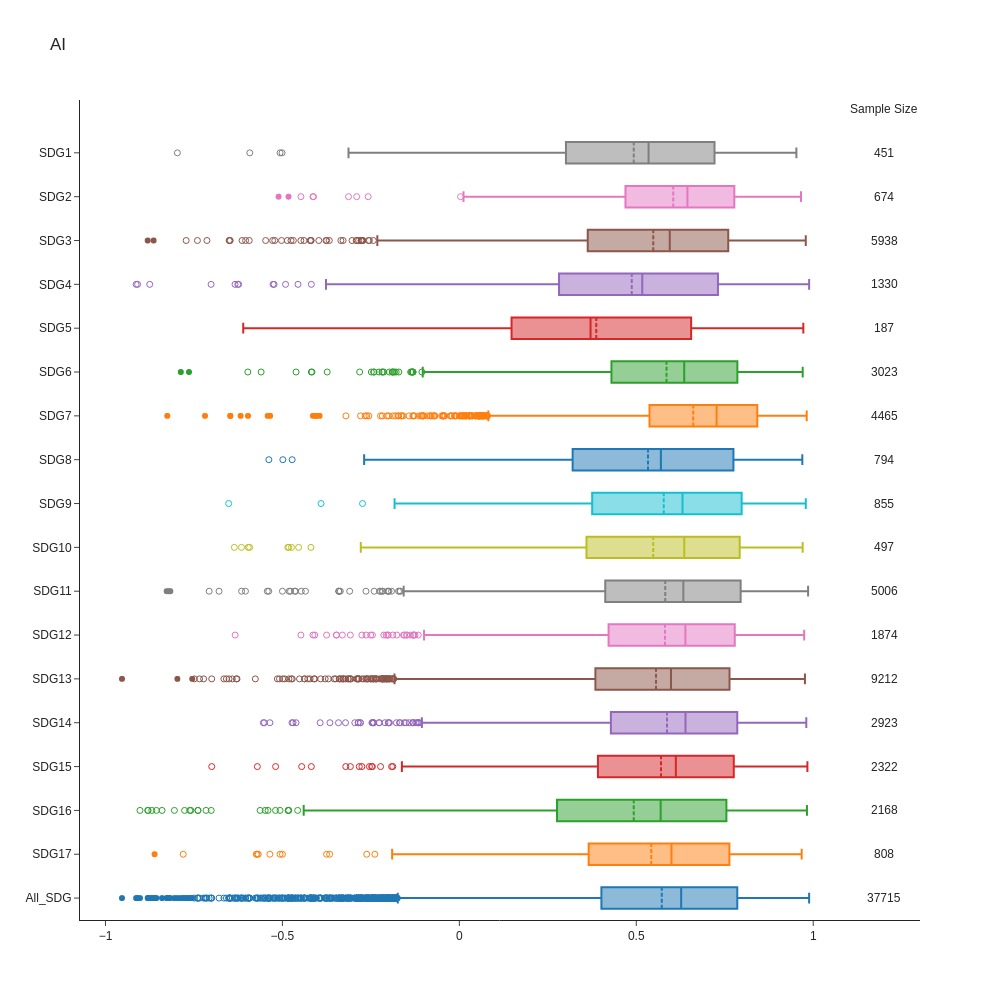
Accuracy (overall): total correct annotation / total annotations

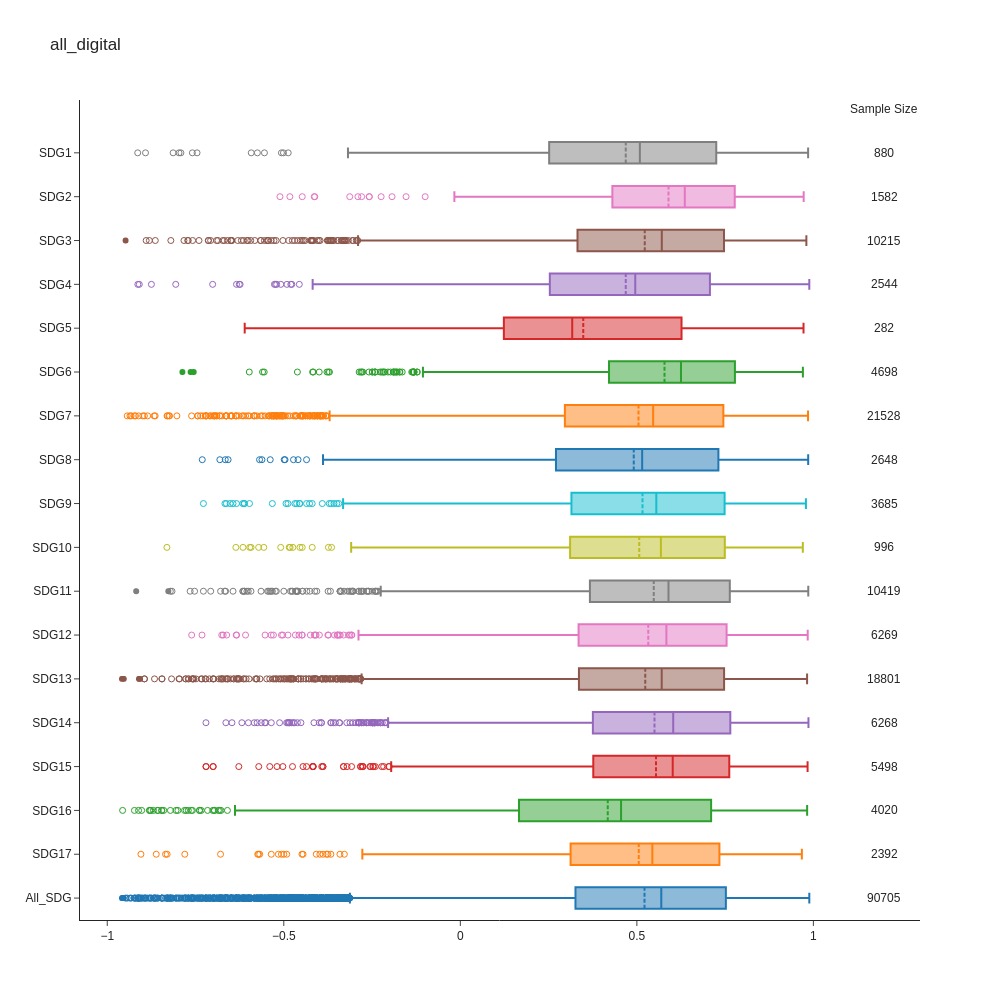
As we can see, the neural network aspect-based model deberta\_asba outperforms the other models both in accuracy and f1-score.

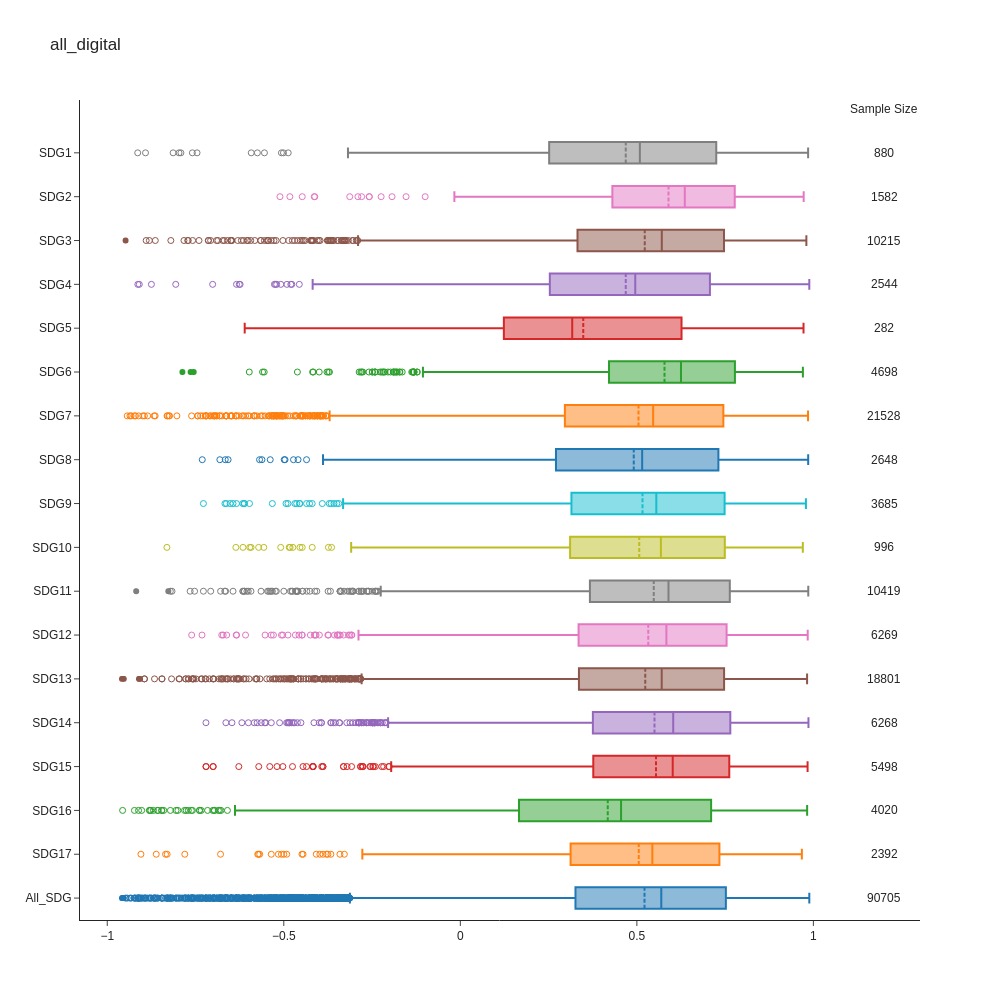
## Preliminary Results

Figure 1: Heatmap of sentiments (Pos\_sent - Neg\_sent)

Figure 2: Heatmap of sentiments for intersection with more than 100 keywords.



Figure 3: Box plot comparing the sentiment of AI-related keywords aggregated across 17 SDGs.

Figure 4: Box plot comparing the sentiment of digital keywords aggregated across 17 SDGs.

Barbieri, F., Anke, L. E., & Camacho-Collados, J. (2022). *XLM-T: Multilingual Language Models in Twitter for Sentiment Analysis and Beyond* (arXiv:2104.12250). arXiv. https://doi.org/10.48550/arXiv.2104.12250

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), Article 1. https://doi.org/10.1609/icwsm.v8i1.14550

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach* (arXiv:1907.11692). arXiv. https://doi.org/10.48550/arXiv.1907.11692

Yang, H., Zeng, B., Xu, M., & Wang, T. (2021). *Back to Reality: Leveraging Pattern-driven Modeling to Enable Affordable Sentiment Dependency Learning*.