Hybrid Classification of European Legislation using Sustainable Development Goals

Michele Corazza, Franco M. T. Gatti, Salvatore Sapienza, Monica Palmirani

University of Bologna, CIRSFID-ALMA AI

Abstract. This study focuses on the automatic classification of European Union legislative documents according to the United Nations Sustainable Development Goals (SDGs) to monitor and improve government policies and legislation. This allows *ex-ante* checks during legal drafting to better align the new proposal with SDG policies and *ex-post* evaluation tools for monitoring the implementation and effectiveness of SDG strategies in the European legislation over time. The research aims to assess the alignment of legislative efforts with these global goals by utilizing an extensive corpus of regulations and directives from the Juncker (2014-2019) and von der Leyen Commission periods (2020-2024). The proposed Hybrid AI methodology employs an unsupervised deep learning approach, leveraging the structure of legislative documents formalized in the Akoma Ntoso XML standard. The research has two primary objectives: first, to examine a novel weighted approach where classifications of the initial articles guide the classification of subsequent articles using an unsupervised sentence embedding model. Second, to monitor document-level classifications over time, tracking legislative evolution and comparing policies under different European Commission presidencies. Initial findings, based on legal expert validation of the technical findings expressed to metrics, reveal that the first articles of legislative documents are crucial in determining the correct SDG classifications and that these classifications may evolve over time with normative modifications and new strategic policies.

Keywords: Hybrid AI · Unsupervised Classification · Temporal Tracking · LegalXML

1 Introduction

Deliberative institutions face an urgent need to detect and track policy implementation in legal acts, measure norm effectiveness, and evaluate societal impact. The Sustainable Development Goals (SDG) program serves as a crucial tool for monitoring global policies, while emerging AI applications in the legislative domain aim to find correspondences between laws and policies defined by deliberative bodies. In 2017, the European Commission's Joint Research Centre (JRC) initiated a policy mapping of Juncker Commission's actions (2014-2019), linking them to the 17 SDGs [6]. This process involved manual mapping and expert analysis of EU legislative documents, followed by automation using text mining and

NLP techniques. A 2022 collaboration between the authors, European DG Informatics, and the European Publication Office resulted in a database of 21991 EU regulations and directives (2010-2019) converted to Akoma Ntoso XML format, and including consolidated versions. Further reviews, and the implementation of text mining and natural language processing techniques, led to an automatisation of the policy mapping process that was implemented starting from the documents of the Von Der Leyen Commission (2020-2024). For our scope we have considered only the manually annotated dataset (Juncker presidency) for creating a ground truth baseline.

However, this dataset classifies the European legislation at the document level. Crucially, legal practitioners - especially at the legislative level, like legal drafters - are interested in knowing precisely the article/portion of the text connected to SDGs in order to track policies and, in case of modifications, to detect improvements over time. Secondly, if the legislator intends to reinforce the SDGs indicators within the legal text, it is important to know exactly what provisions are not sufficient or "weak" in order to reinforce the strategy. Thirdly, legislative documents include modifications/derogations to other documents and the quoted text should be influential in the identification of the correct SDGs.

Some qualified parts of the legislative document express better the scope and the objective of the provision. For instance, the first articles usually describe the objective scope ("what" is regulated), the subjective scope ("who" is regulated) and definitions, which are crucial to contextualise the meaning of the norms. Finally, AI methods are usually applied at a document level. Given limitations in token-handling, legal documents are usually segmented in chunks, with no difference between parts (preamble, definition, articles, final provisions, etc.). This implies that the hierarchical structure of the legal document, which is crucial for the legal significance of the provisions is lost in the AI-friendly representation of the legal text. For these reasons, the main research questions are the following:

- **RQ1**: which parts (Recitals, definition, main body, etc.) of the legislative document are connected with the targets of the SDGs taxonomy? We intend to reach a better granularity of the classification, in order to provide an accurate instrument to the legislator, while maintaining the hierarchical legal significance of the document (e.g., chapters, articles, paragraphs)
- **RQ2**: what is the evolution of SDGs classification over time? We intend to exploit the evolution of the SDGs over time, monitoring the documents, including their consolidated versions, using the temporal metadata of Akoma Ntoso. The temporal spans are in ranges of 3 years (2015-2017, 2018-2020, 2021-2023, 2024) and 5 years (2015-2019, 2020-2024). (e.g., the SDGs 7 related to Affordable and Clean Energy has grown about 1% in the last 5 years monitoring the same documents because the legislator amended them).

We use Akoma Ntoso[1] because it is the official standard in EU institutions for modelling the legislative documents (AKN4EU[2]) and similarly is the standard for the documents in the UN (AKN4UN[3]).

2 Related work

In the Natural Language Processing (NLP) field, the emergence of the transformer models [18] first and subsequently BERT [4] lead to a paradigm where a large model is pre-trained on a self-supervised task, allowing the model to be adapted to a multitude of tasks in a process called fine-tuning. In this context, some models have been fine-tuned from general purpose transformers for the legal domain. One such model is LEGAL-BERT [1], which was fine-tuned on both legislative documents from the EU, UK and US, as well as court documents from the European Court of Justice. In the same family there is another model called custom LEGAL-BERT [21], which was trained on a corpus of case low from the Harvard Law Library. Another dataset is called Pile-of-Law, and it was obtained from 35 different sources in English and used to train a model called PoL [9].

Beside pre-trained models, the application of NLP techniques to the legal domain has produced a multitude of tasks and models. Most of these approaches are related to the judiciary, with a multitude of different tasks such as the prediction of court rulings. In [20], in particular, a pre-trained model is fine-tuned to retrieve similar cases, predict the judgment, and answer legal question in Chinese. Another contribution [5] a transformer model is used to determine which articles have been violated in a given case, by using a global consistency graph which links charge (e.g., abuse) with article and term (e.g., 5 years). Similarly, a model and dataset have been proposed to predict and explain rulings of the Indian supreme court [11]. Beside these models, an evaluation campaign called the Competition on Legal Information Extraction/Entailment [8].

While most of the approaches described in this section are related to the judiciary, the application of NLP approaches to the legislative domain is less explored. However, there is a long standing tradition in the creation of machine-readable representation of legislative documents. In our research, we use the Akoma Ntoso [14, 19] XML standard, which is used across a multitude of institutions around the world [13, 15, 12, 3, 7] and it creates a machine-readable representation that can encode legal definitions, the hierarchical structure of legal documents, temporal aspects (e.g., amendments, consolidation) and nonrmative references.

---

[1] http://docs.oasis-open.org/legaldocml/akn-core/v1.0/akn-core-v1.0-part1-vocabulary.html
[2] https://op.europa.eu/it/web/eu-vocabularies/akn4eu
[3] https://unsceb.org/unsif-akn4un, https://unsceb-hlcm.github.io/

3 Document Collection

The ground truth baseline is composed by 2791 documents along with their associated SDGs, sourced from the JRC portal and derived from the manual policy mapping conducted by the JRC starting in 2017, covering documents produced during the Juncker Commission's mandate (2015-2019). Based on the mapping, each document was linked to at least one SDG, and many documents were connected to multiple SDG.

Our experiment was conducted on a document collection made starting from the 21991 regulations and directives (including consolidated versions) contained in the European legislation database EUR-LEX (2010-2021), converted in Akoma Ntoso format. From this dataset, we have filtered the documents matching the ground truth arriving at a set collection composed of 3846 items. It is important to notice that a consolidation is the version of an act comprising the original act and all the subsequent amendments and corrections to that act. For this reason, the analysis of the documents' evolution over time could reveal the trend of the EU Commission while implementing the SDGs.

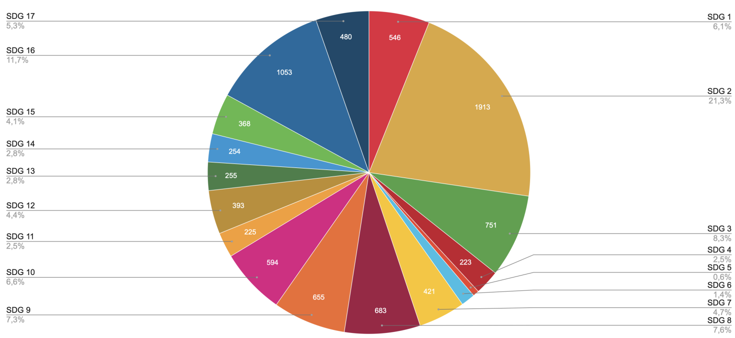

Fig. 1: Number of documents per each SDG in our document collection.

Figure 1 shows how many documents in our dataset are pertaining each SDG by the manual mapping. SDG number 2, which aims at "zero hunger" is the most frequently featured, appearing in 1,913 documents, while goal number 5, focused on "gender equality" is the least featured, appearing in only 5 documents.

Among the 169 targets specified in the 17 SDGs, 39 are absent from all 3846 documents in our dataset. Moreover, 71 targets are found in just 1 to 94 documents, and 31 appear between 103 and 283 times across the entire collection. The 28 most frequently featured targets (appearing between 301 and 1372 times) are depicted in Figure 2. The top two targets (2.3 and 2.1), both pertaining to SDG number 2, are mentioned in 1372 and 1297 documents, respectively.

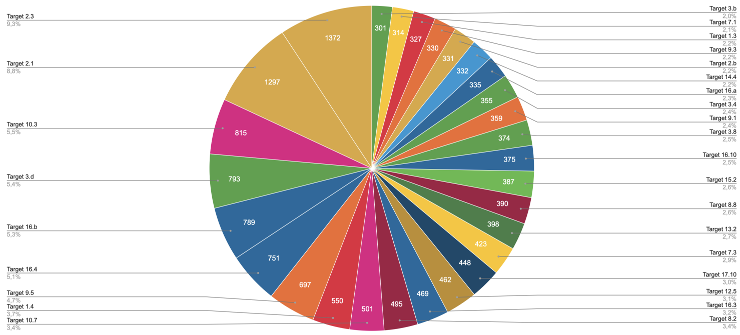

Fig. 2: The 27 most featured targets in our document collection.

4 Methods

In this article, our goal is to asses whether it is feasible to leverage an unsupervised approach in order to determine whether a document is related to one or more SDG targets. To achieve this goal, we use a sentence embedding approach based on Sentence-BERT and the SentenceTransformers library [16]. These models have been fine-tuned to produce vector representations for sentences which can be compared using cosine similarity or euclidean distance as metrics. This way, they can be used in an unsupervised way to assess the semantic similarity between two different portions of text.

In our experiments, we opted to use the "all-distilroberta-v1" model, due to the fact that it shows good performance on various datasets and that it still retains the maximum number of tokens from RoBERTa, namely 512. This model was fine-tuned from the DistillRoBERTa model [17], which in turn was distilled from RoBERTa [10]. This particular model was trained using a contrastive learning procedure, leveraging datasets annotated for semantic similarity. The model M is trained on matched pairs of sentences a, b such that a_i and b_i are semantically related according to the dataset. The two sets of sentences have the same cardinality $|a| = |b| = b_s$ which corresponds to the batch size. In order to proceed with the training, the first step is to obtain the vector representations of these two sets of sentences, normalized so that they have unit norm:

$$\bar{A} = \frac{M(a)}{|M(a)|} \quad \bar{B} = \frac{M(b)}{|M(b)|} \tag{1}$$

Where we assume that each row of the matrices is divided by its own norm. By applying these formulas we obtain two matrices $\bar{A}, \bar{B}$ of size $b_s \times e_s$, where e_s represents the size of the outputs of the model (768 in the case of distillRoBERTa). In order to obtain the cosine similarity between the sentence embeddings in $\bar{A}, \bar{B}$ it is then possible to apply a simple matrix multiplication:

$$S = \bar{A}\bar{B}^T \tag{2}$$

Due to the fact that all vectors $a_i \in \bar{A}, b_i \in \bar{B}$ have unit norm, the normal formula for the cosine similarity can be :

$$sim(a_i, b_i) = \frac{a_i \cdot b_i}{|a_i||b_i|} = a_i \cdot b_i \tag{3}$$

Meaning that the cosine similarity between the vectors can be computed using only the dot product between them. In the context of the similarity matrix S, then, we can observe that:

$$S = \begin{bmatrix} a_1 \cdot b_1 & \dots & a_1 \cdot b_n \\ \vdots & \ddots & \vdots \\ a_n \cdot b_1 & \dots & a_n \cdot b_n \end{bmatrix} \tag{4}$$

Meaning that each cell of the matrix $s_i j \in S$ corresponds to the cosine similarity between a_i and b_j. Once obtained this representation, which acts as the prediction of the model, we can observe that due to the fact that only the sentences with the same index are semantically related. For this reason, we can use as labels a diagonal matrix with the same size as S but with the values of the diagonal set to 1, the others set to 0:

$$y = \mathrm{diag}(1, 2, \dots, b_s) \tag{5}$$

Finally, the complete loss is expressed as a double cross-entropy, as first presented in [2]:

$$\mathcal{L}(S, y) = -\frac{1}{2}\left(\frac{1}{b_s}\sum_{i=0}^{b_s} \log \frac{\exp S_i}{\sum_{j=0}^{b_s} \exp S_j} y + \frac{1}{b_s}\sum_{i=0}^{b_s} \log \frac{\exp S_i^T}{\sum_{j=0}^{b_s} \exp S_j^T} y\right) \tag{6}$$

This double cross entropy, then, has the desired outcome of producing almost orthogonal vectors (with cosine similarity close to 0) for non semantically related sentences, while it should produce almost parallel vectors (with cosine similarity close to 1) for semantically related sentences.

With this model, it is possible to produce semantically aware embeddings for sentences. The maximum length of the normative documents, however, does not allow the application of the model to the entire document. For this reason, we developed an approach that is informed by the structure of the document itself, encoded in the Akoma Ntoso XML tree, and which also considers normative references (see Figure 3). Since our goal is to encode articles and recitals from the preamble, we proceed by first producing a list of inline elements in Akoma Ntoso, meaning those that appear in the middle of text, such as references, dates, etc. With this list, we can populate the leaves of our tree, meaning the elements that have either no children or only inline children. They are also associated with all the textual content that they contain, including text that is inside inline children if they are present. Given a leaf element l of the tree with no references, we can compute its vector representation using the following algorithm:

$$v(l) = M(t(l)) \tag{7}$$

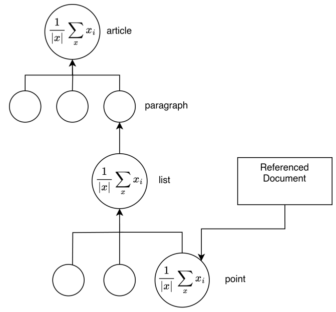

Fig. 3: The strategy used to obtain a vector representation for articles and recitals. Each non-leaf node is represented by a vector obtained from its children, and, if present, the normative references in its text.

Where we denote the textual content of l with $t(l)$. For non-leaf nodes, then, we use a recursive approach to reconstruct their vectors from their children:

$$v(e) = \frac{1}{1 + |c(e)|} \left(M(t(e)) + \sum_i v(c_i(e)) \right) \tag{8}$$

Where $c(e)$ is a tuple containing all the children of the element e and $c_i(e)$ denotes the nth child of the element e. Our approach is also able to represent information contained in documents that are referenced in the text. In particular, we are able to consider both punctual and non punctual normative references, meaning references that point to a specific portion of a document (e.g., an article, a point, a paragraph, etc) and those that indicate an entire document, respectively. In order to represent these references, we use two distinct approaches:

$$R(i) = \begin{cases} v(i) & \text{if } i \text{ is a punctual reference} \\ \frac{1}{2}M(title(i)) + v(article_1(i)) & \text{otherwise} \end{cases} \tag{9}$$

Where $title(i)$ and $article_1(i)$ represent the title and first article of a document, respectively. This approach uses the vector representation of the element in punctual references, while it represents entire document using its title and first article.

Finally, we can produce the vector representation of a non leaf node containing children and one or more references:

$$v(e) = \frac{1}{2 + |c(e)|} \left(M(t(e)) + \sum_i v(c_i(e)) \frac{1}{r(e)} \sum_j R(r_j(e)) \right) \tag{10}$$

Where $r(e)$ is a tuple containing all the references in the text of the node e, while $r_j(e)$ represents the *j-th* reference. This formula computes the mean between the textual content of each element, its children and a vector representing the references contained in its text, allowing us to obtain vector representations for both recitals and articles, which are aware of the structure of the document and of the references.

The final step in our method is the measurement of the semantic similarity between articles and recitals, which are represented by vectors obtained by the aforementioned procedure, and the SDG targets. Each target is associated with a description, so we can represent it using the model used to represent articles and recitals. The cosine similarity between the articles, recitals and each SDG target are used to obtain a ranking, where each portion of the normative document is associated with a list of SDG targets, from the most similar to the least similar.

5 Results and Validation

In order to evaluate the performance of the model, we opted to use the annotated dataset discussed in section 3, which contains matches between documents the relevant SDG targets. This task is a multi-label classification, in which each document can be assigned one or more SDG targets. Since we do not have a precise threshold that can be used to discriminate between matching and non matching SDG targets for a given ranking, we chose to consider only the top 5 ranking SDGs as matching for any given article or recital. This group of SGDs are called the "predicted matching". While it would be possible to leverage some information about the number of SDG for each document in the gold standard, this would amount to a model that is tuned for the same data used to test it, which is methodologically problematic in our unsupervised setting. Secondly, from the legal point of view, we would be agnostic regarding this parameter for discovering new potential legal knowledge.

In our approach, the semantic similarity with the SDGs is measured on articles and recitals, not entire documents. It is important to underline that while articles constitute the binding part of the normative text, recitals serve as an interpretative guidance and cast light on the overall strategy pursued by the normative instrument. In our case, recitals should be able to provide further context to the scope and the definitions mentioned in the articles. Therefore, their use is a valuable approach as they contain the rationale of a normative act, which can be helpful to classify its role within the SDG framework

For this reason, we chose to compare three different strategies to aggregate the results in order to obtain predictions at the document level:

– **All articles**: we use the union of the predicted matching SDGs for all articles of the document;
– **First four articles**: we use the union of the predicted matching SDGs for the first four articles of the document. This choice is motivated by the intuition that the first four articles of documents generally contain the core scope and context of the regulation which are crucial for the classification.
– **First four articles + recitals**: we use the union of the predicted matching SDGs for the first four articles of the document, as well as the predicted

matching SDGs for all the recitals, consistently with observed above with regards to recitals.

With this approach, it is possible to reconstruct, the predicted SDGs for any given document. However, one must consider that this unsupervised approach is very challenging. In particular, our method does not leverage any information about the number of SDG targets that are related to any given document, article or recital. For this reason, the model might produce a higher or a lower number of SDG targets depending on the specific document. Additionally, multiple articles might be predicted as similar to the same SDG target, leading SDGs that are missing from the predictions, especially when the number of ground truth SDGs is high. One other challenging aspect is the fact that, in our dataset, 39 out of 169 targets are not attested. In our evaluation, we chose to consider them despite the fact that they influence the total averages used to compare performance. While it could be possible to remove labels that are not predicted by the model and not annotated at the document level, this would make any comparison between models meaningless, so we opted to keep all the non-attested classes.

In order to evaluate the approach, we present the macro averaged and weighted average precision, recall and F1 score obtained from the 169 SDG targets (Table 1). In addition to the performance metrics obtained from the document-level predictions, we present a random baseline, obtained for each strategy by sampling 5 random SDGs for each article/recital which was used for the evaluation, applied 100 times. We report the mean and standard deviation of the resulting metrics.

While analysing the results of the evaluation, some interesting patterns emerge. In particular, we can see that the difference in terms of macro averaged precision is not substantial when comparing the three models, and that this approach does not produce high precision results. Even if it is a small margin, the best performance in terms of precision is obtained by adopting the "all articles" strategy. However, despite the relatively high number of articles in the document, the performance in terms of recall is quite low, meaning that some of the predicted SDGs might be related to procedural aspects, which are predicted multiple times per document, leading to a low recall value while not impacting the precision. For the first four strategies, the low number of articles seems to penalize the recall values, resulting in a very low overall F1 score. In terms of overall performance, then, the recitals + 4 articles strategy leads to the best overall F1 values, with precision values that are not far from the best ones obtained from the all articles approach, but without the low recall. This strategy is the more promising one and it is the only one to measurably surpass the random baseline F1 scores for both averaging strategies. Incidentally, this finding is also explainable in the light of the legal nature of recitals, which serve as additional context for the binding provisions and, therefore, clarify the scope of the regulation.

In an effort to evaluate the fine-grained performance of the model, we annotated 50 articles with their associated SDGs and we proceeded to measure the resulting top 5,3, and 1 recalls (see Table 2), meaning the ratio between the correctly identified SDG targets in the top 5,3,1 ranking predictions and the total

Strategy	Average	Precision	Recall	F1 Score
All articles	Macro	$\underline{\mathbf{0.11}}$	0.16	$\underline{0.07}$
	Weighted	$\underline{\mathbf{0.37}}$	0.22	0.14
Random (All articles)	Macro	0.03 ± 0.0003	$\underline{0.33} \pm 0.008$	0.06 ± 0.0004
	Weighted	0.12 ± 0.001	$\underline{0.41} \pm 0.003$	$\underline{0.17} \pm 0.001$
First four	Macro	$\underline{0.09}$	$\underline{0.09}$	$\underline{0.04}$
	Weighted	$\underline{0.29}$	$\underline{0.12}$	0.06
Random (First four)	Macro	0.02 ± 0.0006	0.08 ± 0.007	0.03 ± 0.0008
	Weighted	0.09 ± 0.003	0.1 ± 0.003	$\underline{0.08} \pm 0.002$
Recitals + first four	Macro	$\underline{0.09}$	0.36	$\underline{\mathbf{0.10}}$
	Weighted	$\underline{0.27}$	0.44	$\underline{\mathbf{0.22}}$
Random (Recitals + first four)	Macro	0.03 ± 0.0001	$\underline{\mathbf{0.52}} \pm 0.008$	0.05 ± 0.0002
	Weighted	0.10 ± 0.0008	$\underline{\mathbf{0.65}} \pm 0.003$	0.16 ± 0.001

Table 1: Precision, recall and F1 score for the four strategies, obtained from Macro and Weighted averages over individual classes. In bold, the best values for each metric. Underlined, the higher metric when comparing each strategy with its baseline. For each baseline we report the means and standard deviations of the metrics over 100 runs.

Model	Top 5 recall	Top 3 recall	Top 1 recall
Ours	**0.53**	**0.35**	**0.13**
Random Baseline	0.028 ± 0.012	0.017 ± 0.009	0.005 ± 0.006

Table 2: Top 5,3,1 recall for the manually annotated articles using our model and a random baseline. The random baseline has been executed 100 times and we show the mean and standard deviation for the recalls. In bold, the best results.

number of SDGs associated with the articles. We also report a random baseline, obtained by sampling random SDGs for each of the 50 manually annotated articles, which is applied 100 times and reported in terms of means and standard deviations. These results show that, while going from a fine-grained classification to a document-level one is indeed challenging, our model is able to retrieve the SDGs related to a given article with reasonable performance and that the results are markedly better than a random baseline.

6 Discussion and Final Remarks

The results provide answers to the research questions that are the core of this paper:

- **RQ1**: which parts (Recitals, definition, main body, etc.) of the legislative document are connected with the targets of the SDGs taxonomy? We are now able to precisely detect the classification of the SDGs article-by-article, thus contributing to effective SDGs monitoring and better legislation. We are more successful when using the first articles and recitals. In articles where there is only a normative reference, we should be able to navigate the citation for detecting the text of the destination and include it in the model. Instead, for legislative documents focused on procedural aspects of the European Union (e.g., budget), the mapping is more generic and unspecified (e.g., 1.4, 16.10).

– **RQ2**: what is the evolution of SDGs classification over time? Figures 5 and 6 allows for a clear understanding of the evolution over time of the EU committments towards certain SDGs

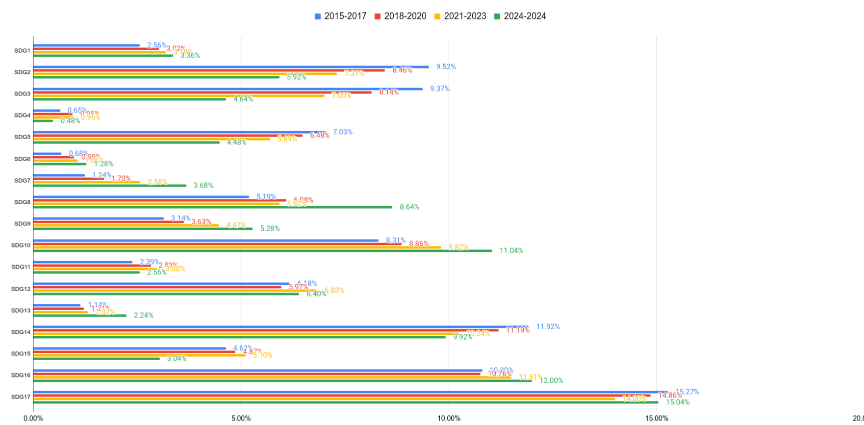

Fig. 4: Percentage distribution of EU legislation across Sustainable Development Goals (SDGs) from 2015 to 2024 - 3-year range

Notable trends include the steady decrease in SDG02 (Zero Hunger) related legislation from about 9.52% in 2018-2020 to 5.92% in 2024. SDG13 (Climate Action) exhibits more modest changes over time, hovering around 2-3% throughout, but with a very recent increase which almost doubles its percentage. Unexpectedly, EU legislation related to SDG03 (Good Health and Well-being) decreases during the 2021-2023 period (4.64%) in comparison to the initial period (9.37%) in the initial period (2015-2017). This decrease in health-related legislation in 2021-2023 seems in contrast to the EU's legislative response to the COVID-19 pandemic.

5-year trends seem to follow a similar distribution with regards to every SDG. Naturally, trends are diluted in a longer period, thus resulting in less evident shifts. However, this analysis allows the comparison between two different Commissions, thus helping legislative drafting and policy monitoring.

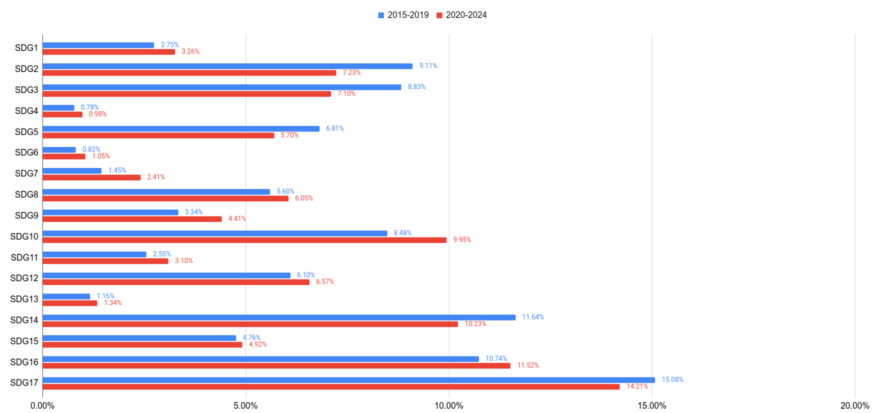

Fig. 5: Percentage distribution of EU legislation across Sustainable Development Goals (SDGs) from 2015 to 2024 - 5-year range

In conclusion, we have used a database of JRC already annotated by human experts with the SDGs classification and an AKN database of legislation for assigning the same classification at the lower granularity (e.g., article vs. document; definition vs. article).

The following findings emerge from the experiments: a) structured documents in AKN-XML are able to handle temporal elements in detecting SDGs; b) documents linked with the main topics of the SDGs work better than documents that describe legal normative procedures of the European or legislative system (e.g., implementation of legislation, modifications, relationship with Member States, derogations, etc.); c) articles describing the scope, definitions, objectives - including Recitals - contribute to the effectiveness of the classifier consistently with their legal function; d) articles that include few sentences and many normative references produce worse results, due to challenges in representing references.

Considering these output the future work will be focused on the following tasks: a) to use EUROVOC for creating a first signal on the good candidates of the SD goals; b) to use definitions and the first five articles for refining the mapping of the targets; c) to navigate the normative references in order to include text from the cited document; d) to navigate the normative references to the updated version, when it is available, allowing us to update the mapping with the evolution of the normative system; e) to map the other articles (from the fifth) using the target only in instances where there are additional SDGs targets as good candidates.

Our work provides better traceability of the SDGs policies in the EU legislation permitting the legislator to detect the articles where the association is weakest. During the legal drafting, our tool could be integrated into the editor to suggest better legal definitions for improving the implementation of the SDGs.

7 Acknowledgement

This project is conducted with the support of the European Commission funds within ERC HyperModeLex. Grant agreement ID: 101055185.

References

1. Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., Androutsopoulos, I.: LEGAL-BERT: The muppets straight out of law school. In: Cohn, T., He, Y., Liu, Y. (eds.) Findings of the Association for Computational Linguistics: EMNLP 2020. pp. 2898–2904. Association for Computational Linguistics, Online (Nov 2020). https://doi.org/10.18653/v1/2020.findings-emnlp.261, https://aclanthology.org/2020.findings-emnlp.261
2. Chen, Y., Qi, X., Wang, J., Zhang, L.: Disco-clip: A distributed contrastive loss for memory efficient clip training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 22648–22657 (June 2023)
3. Cvejić, A., Grujić, K.G., Cvejić, A., Marković, M., Gostojić, S.: Automatic transformation of plain-text legislation into machine-readable format. In: The 11th international conference on information society, technology and management (ICIST 2021) (03 2021)
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
5. Dong, Q., Niu, S.: Legal judgment prediction via relational learning. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 983–992. SIGIR '21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3404835.3462931, https://doi.org/10.1145/3404835.3462931
6. European Commission and Joint Research Centre, Borchardt, S., Barbero Vignola, G., Buscaglia, D., Maroni, M., Marelli, L.: Mapping EU policies with the 2030 agenda and SDGs – Fostering policy coherence through text-based SDG mapping. Publications Office of the European Union (2023). https://doi.org/doi/10.2760/110687
7. Flatt, A., Langner, A., Leps, O.: Model-Driven Development of Akoma Ntoso Application Profiles: A Conceptual Framework for Model-Based Generation of XML Subschemas. Springer Nature (2023)
8. Goebel, R., Kano, Y., Kim, M.Y., Rabelo, J., Satoh, K., Yoshioka, M.: Summary of the competition on legal information, extraction/entailment (coliee) 2023. In: Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law. p. 472–480. ICAIL '23, Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3594536.3595176, https://doi.org/10.1145/3594536.3595176
9. Henderson, P., Krass, M., Zheng, L., Guha, N., Manning, C.D., Jurafsky, D., Ho, D.: Pile of law: Learning responsible data filtering from the law and a 256gb open-source legal dataset. Advances in Neural Information Processing Systems **35**, 29217–29234 (2022)
10. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)

11. Malik, V., Sanjay, R., Nigam, S.K., Ghosh, K., Guha, S.K., Bhattacharya, A., Modi, A.: ILDC for CJPE: Indian legal documents corpus for court judgment prediction and explanation. In: Zong, C., Xia, F., Li, W., Navigli, R. (eds.) Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). pp. 4046–4062. Association for Computational Linguistics, Online (Aug 2021). https://doi.org/10.18653/v1/2021.acl-long.313, https://aclanthology.org/2021.acl-long.313
12. Palmirani, M.: Akoma ntoso for making FAO resolutions accessible. In: Peruginelli, G., Faro, S. (eds.) Knowledge of the Law in the Big Data Age, Conference 'Law via the Internet 2018', Florence, Italy, 11-12 October 2018. Frontiers in Artificial Intelligence and Applications, vol. 317, pp. 159–169. IOS Press (2018). https://doi.org/10.3233/FAIA190018, https://doi.org/10.3233/FAIA190018
13. Palmirani, M.: Lexdatafication: Italian legal knowledge modelling in akoma ntoso. In: Rodríguez-Doncel, V., Palmirani, M., Araszkiewicz, M., Casanovas, P., Pagallo, U., Sartor, G. (eds.) AI Approaches to the Complexity of Legal Systems XI-XII - AICOL International Workshops 2018 and 2020: AICOL-XI JURIX 2018, AICOL-XII JURIX 2020, XAILA JURIX 2020, Revised Selected Papers. Lecture Notes in Computer Science, vol. 13048, pp. 31–47. Springer (2020). https://doi.org/10.1007/978-3-030-89811-3_3, https://doi.org/10.1007/978-3-030-89811-3_3
14. Palmirani, M., Sperberg, R., Vergottini, G., Vitali, F.: Akoma Ntoso Version 1.0 Part 1: XML Vocabulary. OASIS Standard (August 2018), http://docs.oasis-open.org/legaldocml/akn-core/v1.0/akn-core-v1.0-part1-vocabulary.html
15. Palmirani, M., Vitali, F., Bernasconi, A., Gambazzi, L.: Swiss federal publication workflow with akoma ntoso. In: Hoekstra, R. (ed.) Legal Knowledge and Information Systems - JURIX 2014: The Twenty-Seventh Annual Conference, Jagiellonian University, Krakow, Poland, 10-12 December 2014. Frontiers in Artificial Intelligence and Applications, vol. 271, pp. 179–184. IOS Press (2014). https://doi.org/10.3233/978-1-61499-468-8-179, https://doi.org/10.3233/978-1-61499-468-8-179
16. Reimers, N., Gurevych, I.: Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084 (2019)
17. Sanh, V., Debut, L., Chaumond, J., Wolf, T.: Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108 (2019)
18. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
19. Vitali, F., Palmirani, M., Sperberg, R., Parisse, V.: Akoma Ntoso Version 1.0. Part 2: Specifications. OASIS Standard (August 2018), http://docs.oasis-open.org/legaldocml/akn-core/v1.0/akn-core-v1.0-part2-specs.html
20. Xiao, C., Hu, X., Liu, Z., Tu, C., Sun, M.: Lawformer: A pre-trained language model for chinese legal long documents. AI Open **2**, 79–84 (2021). https://doi.org/https://doi.org/10.1016/j.aiopen.2021.06.003, https://www.sciencedirect.com/science/article/pii/S2666651021000176
21. Zheng, L., Guha, N., Anderson, B.R., Henderson, P., Ho, D.E.: When does pretraining help? assessing self-supervised learning for law and the casehold dataset of 53,000+ legal holdings. In: Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law. p. 159–168. ICAIL '21, Association for Computing Machinery, New York, NY, USA (2021)