Assembleia da República roll-call votes dataset

Corpora of 15,621 bills presented in the Portuguese Parliament (1980-2024) with annotated topics

NELSON SANTOS

University of Namur, Transitions Institute

Nelson-leonardo.rodrigues@unamur.be

SOFIA SERRA-SILVA

University of Lisbon, Institute of Social Sciences

Sofia.silva@ics.ulisboa.pt

TIAGO SILVA

University of Lisbon, Institute of Social Sciences

Tiago.silva@ics.ulisboa.pt

28 October 2025

Data Summary

This dataset was collected as part of the empirical strategy of the article *The Role of Radical Right Parties in Escalating Parliamentary Conflict* (Santos, Serra-Silva, and Silva, 2025). For a detailed account of the methodological and empirical decisions underpinning data collection and analysis, please refer to that publication [hyperlink].

The data were retrieved from the official open data platform of the Portuguese Parliament (*Assembleia da República*), in accordance with its open data policy [hyperlink]. All legislative initiatives (15,666) were obtained in XML format and subsequently processed through Natural Language Processing (NLP) techniques to extract and structure a wide range of information, as described in the variable list below.

The XML files were collected on **19 April 2024**, covering all legislative initiatives introduced between the **II Legislature (1980–1983)** and the **XV Legislature (2022–2024)**.

Legislature	Number of Bills
II	160
III	376
IV	199
V	691
VI	584
VII	857
VIII	880
IX	724
X	1,337
XI	792
XII	2,622
XIII	3,178
XIV	1,974
XV	1,292
Total	**15,666**

Table 1 - Number of bills by legislature

Although the Parliament's open data portal provides information on all initiatives since the second legislature (1980), some entries are incomplete. Notably, the XML files include the *full text* of bills only from the **VIII Legislature (1999–2002)** onwards.

To identify the policy area of each bill, we employed the **CAP Babel Machine** developed by the HUN-REN Centre for Social Sciences (Sebők et al., 2024), specifically the *xlm-roberta* model the authors have fine-tuned on Portuguese data[1]. This multilingual model classifies textual data into the **21 major policy categories** defined by the *Comparative Agendas Project* (CAP). Since the classification relies on the full text of the bills, topical categorization is available only from the VIII Legislature onwards.

CAP Babel Predictions (21 Categories)

Bill_Babel_prediction	Frequency
1.0	507
2.0	453
3.0	1,392
4.0	662
5.0	781
6.0	1,150
7.0	722
8.0	251
10.0	827
12.0	1,115
13.0	336
14.0	725
15.0	410
16.0	174
17.0	154
18.0	126
19.0	875
20.0	1,278
21.0	399
23.0	268
Total	**12,605**

Table 2 - Frequency of CAP Babel model predictions

[1] The model used is available on the following site: https://huggingface.co/poltextlab/xlm-roberta-large-portuguese-cap-v3

For analytical purposes, as explained in Santos et al. (2025), these 21 CAP categories were merged into **11 broader policy domains**:

Bill_topic	Frequency
Civil rights and liberties	453
Education	1,150
Environment, agriculture, and energy	1,635
Government operations, public lands, and transportation	2,504
Health	1,392
Housing	725
International affairs and foreign trade	1,001
Labor and social welfare	1,117
Law, crime, and defense	1,289
Macroeconomic and financial issues	917
Science, Communication, and Cultural issues	422
Total	**12,605**

Table 3 - Frequency of the 11 broader policy domains

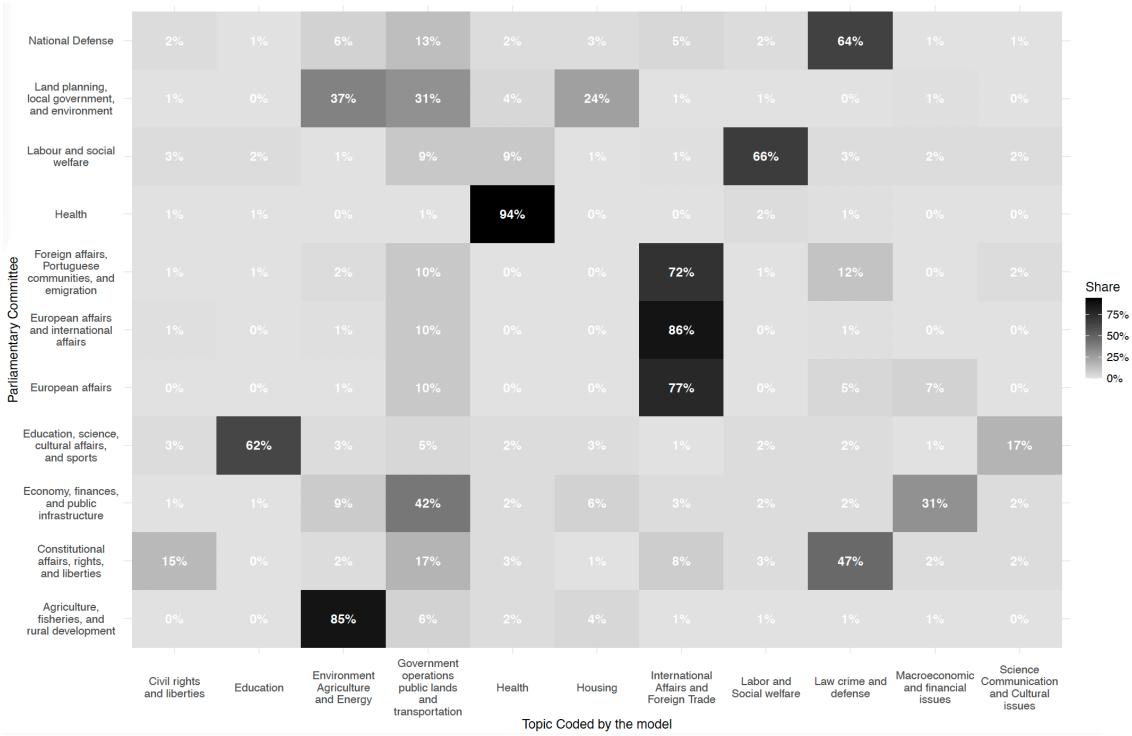
To evaluate the **validity and accuracy** of the **CAP Babel Machine** model within our dataset, we conducted additional robustness tests. Specifically, we compared the model's classification across the **11 policy topics** with the **corresponding permanent parliamentary committees** to which each bill had been considered[2]. Since these committees are specialized by subject matter, we can proceed to a qualitative assessment on the extent the committees and the topic identified overlap.

Figure 1 shows the distribution of bills for each topic across parliamentary committees (percentages by column), while Figure 2 illustrates the distribution of bills considered within each parliamentary committee across topics (percentages by row). Based on a qualitative overview of both figures, we can confirm that, to a significant extent, the model successfully identifies the primary topic of each bill.

[2] For parliamentary committees that underwent slight name changes while addressing the same matters, we standardized the names to ensure a single, consistent observation.

Figure 1 – Distribution of each topic's content across the parliamentary committees.

Figure 2 – Distribution of each parliamentary committee's content across the topics.



The final dataset contains **15,666 observations** and **44 variables**. Below we describe the content of each variable.

Variable	Description
ID	Unique identifier for each observation.
Bill_type	Category of legislative proposal based on initiator and subject matter (e.g., project of law, proposal of law).
Legislature	The legislature in which the bill was introduced.
Legislative_session	The specific session of the legislature in which the bill was presented.
Start_date_legislature	Starting date of the corresponding legislature.
End_date_legislature	Ending date of the corresponding legislature.
Bill_date	Date of submission of the bill.
Bill_link	URL linking to the bill's PDF on the parliamentary website.
Bill_author	Type of initiator (e.g., Parliamentary Group, Government).
Bill_author_individual	Name(s) of the MP(s) or other initiators (e.g., President of the Assembly), when applicable.
Bill_author_Party	Political party affiliation of the initiator(s).
Bill_committee	Committee to which the bill was referred.
Bill_Babel_prediction	Policy category assigned by the CAP Babel Machine model.
Bill_topic	Aggregated policy domain based on merged CAP categories (see Santos et al., 2025).
Bill_title	Official title of the bill.
Bill_text	Full legislative text of the bill.
Proposed_amendments	Text or references to proposed amendments, where available.
Bill_vote_type	Type of vote held (e.g., general, final).
Bill_supporting_parties_MPs	Parties that voted in favor.
Bill_opposing_parties_MPs	Parties that voted against.
Bill_abstention_parties_MPs	Parties that abstained.
Bill_absence_parties_MPs	Parties absent during the vote.
Bill_result	Outcome of the vote (approved or rejected).
Bill_unanimous_vote	Binary indicator for unanimous votes.
Columns with Party Names	Party-specific vote indicator (1 = Support; -1 = Reject; 0 = Abstain/Absent).

Table 4 - Dataset variables' description

References

Santos, N., Serra-Silva, S., & Silva, T. (2025). The role of radical right parties in escalating parliamentary conflict: policy issues and party responses in Portugal. *Frontiers in Political Science*, *7*, 1553921. https://doi.org/10.3389/FPOS.2025.1553921/BIBTEX

Sebők, M., Máté, Á., Ring, O., Kovács, V., & Lehoczki, R. (2024). Leveraging Open Large Language Models for Multilingual Policy Topic Classification: The Babel Machine Approach. *Social Science Computer Review*. https://doi.org/10.1177/08944393241259434/ASSET/IMAGES/LARGE/10.1177_08944393241259434-FIG8.JPEG