2026

PolarScopEU: Research Note and User Guide

Supplementary Material for the Flask-Based Python Application PolarscopEU

Tiago Silva
(Institute of Social Sciences – University of Lisbon)
SILVA, TIAGO ANDRÉ CASAL DA

Table of Contents

PolarScopEU: Research Note

*What the application is, how it works, and how the social media (**Bluesky**) analysis and Bias Plot components are calculated*

Summary

PolarScopEU (link) is a web-based research application designed to collect, analyze and visualize political communication on social media. The current version only works with ***Bluesky*** (https://bsky.app/), but we hope to expand in the future to work on other text-based social networking sites with public content. The tool allows a user to enter a Bluesky handle, collect posts from that account and/or its network, apply automated text analysis (i.e., topic identification, sentiment analysis and EU-reference detection), and produce interactive summaries that help the users understand how diversified (in terms of topics discussed) and balanced (in terms of sentiment) is the content published by different actors with a presence on social media.

The application combines a *Flask* front-end, *SQLAlchemy*-managed databases, the *atproto client* for ***Bluesky*** data collection, and transformer-based NLP pipelines loaded locally from *Hugging Face*-compatible model directories. Its outputs are intended for exploratory research, comparative political communication analysis and user-friendly demonstrations of bias, diversity and alignment in social media discourse.

A central feature is the **Bias Plot**, a two-dimensional representation of network discourse. It positions accounts according to tone (share of positive and neutral posts) and topic diversity (number of topics that appear above a chosen threshold). The same page also reports top topics per account and (cosine) similarity scores that compare the original handle with other accounts in the network.

This app was inspired in **Voting Advice Applications** and software facilitating the manual coding of text to offer students, researchers and interested citizens, an intuitive tool to map online political communication and analyze text-content in very simple fashion. The current version of the app offers two main features 1) Social Media Analysis and 2) EU Politicization Analysis. The following two sections explain these features in more detail.

Application's primary feature (Social Media Analysis)

1.1. Feature Overview (Social Media Analysis)

PolarScopEU was primarily developed as a no-code analytical interface so that a user can move from a social media (Bluesky) handle to a structured set of results without writing (Python) code or manually cleaning data. The app supports a fuller research workflow, where a user can collect posts, run analyses one step at a time, download the resulting dataset, and inspect a set of descriptive visualizations. It also supports a simplified route in which a user can type a handle and request a **Bias Plot** directly. It is primarily aimed at researchers and students interested, or working, on the topics of (online) political communication and European integration. Additionally, it can also be used by interested citizens or groups to map and

monitor online political communication. Among other things, this app can tell us how diversified and polarized is the content of a single Bluesky account or the entire network of accounts that it follows.

From a software perspective, the application is built around Flask routes, Jinja templates and a modular Python back-end. The system stores user-facing collections in a dynamic database table named after the authenticated user and the platform. It also supports longer-term actor-specific tables for curated handles ('Resources' feature). This architecture allows the app to separate exploratory use from more cumulative research storage.

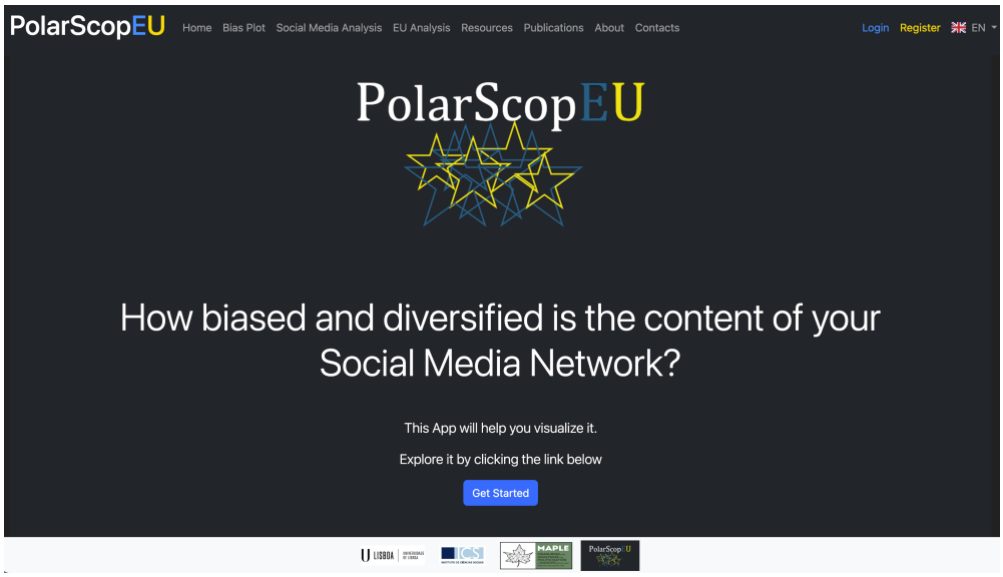


Figure 1 – Starting Page (PolarscopEU).

Access to the application requires user registration and authentication to protect the available data and functionalities and to ensure controlled use of the platform. During registration, passwords are not stored in plain text. Instead, they are converted into *bcrypt* hashes using the *Flask-Bcrypt* library, a widely used solution for secure credential storage. This method automatically applies salting, which strengthens protection against brute-force attacks and the use of precomputed lookup tables. During authentication, the application does not compare the original password itself but rather checks the submitted password against the stored hash, ensuring that user credentials are never directly exposed. At the same time, the registration process is designed to minimize data collection, requiring only a username, password, and email address (which is **not currently requiring a validation step**). In addition, the analytical content stored by the application consists exclusively of publicly available data collected from social media platforms.

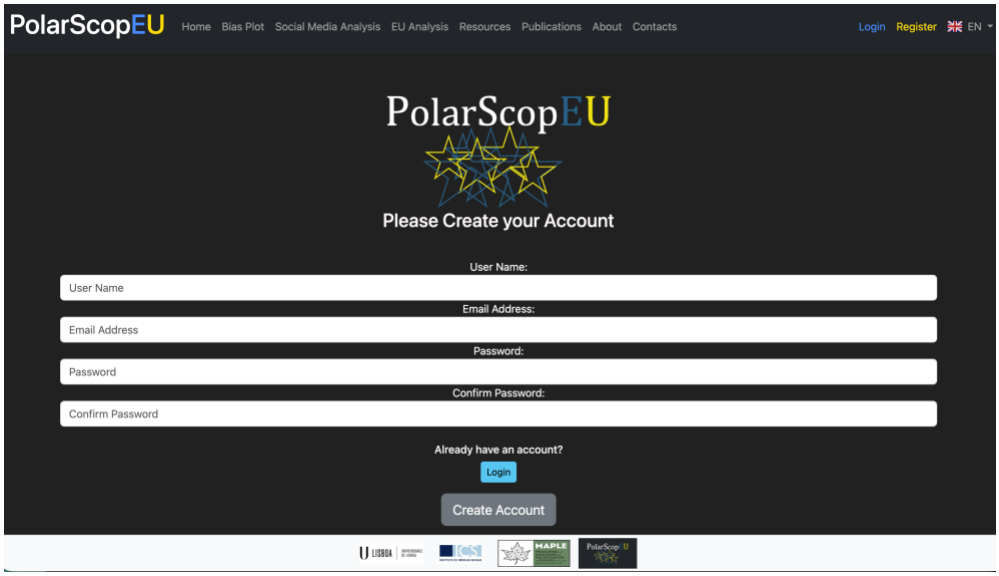


Figure 2 - Registration Page (PolarscopEU).

1.2. Social Media Data Collection (Bluesky)

The current version of the application works with **Bluesky**. For this platform, data collection is carried out through the Bluesky/AT Protocol Python SDK (*atproto*). A *handle* submitted by the user is first resolved to a DID ("Decentralized identifier") via the *IdResolver*. Once the DID is obtained, the app uses it to collect posts from that account and, if network analysis is requested, to identify the accounts followed by that user and retrieve posts from them as well.The collection pipeline supports both user-only and network-level retrieval, as well as sample and full-history modes.

For user-level collection, the app retrieves posts from the selected handle and can optionally include or exclude reposts. Repost filtering is implemented by comparing the post author to the original handle. For network-level collection, the app iterates through followed accounts, retrieves their posts, and can also include the original handle's posts in the same dataset. Data are parsed into a tabular structure containing at least: post identifier (`cid`), `handle`, publication `Date`, post text (`Content`), `Likes`, `Reposts` and `Replies`. Users have the option of collect "**All Data**" or a "**Sample**" (the latest **45**[1] articles). The sample option can be useful for testing the app or doing network analysis of handles that follow many accounts.

[1] We temporarily reduced the original sample size of 200 posts to better accommodate all users under the current virtual machine/server constraints. This value may be adapted or increased in the future if additional computing resources become available.

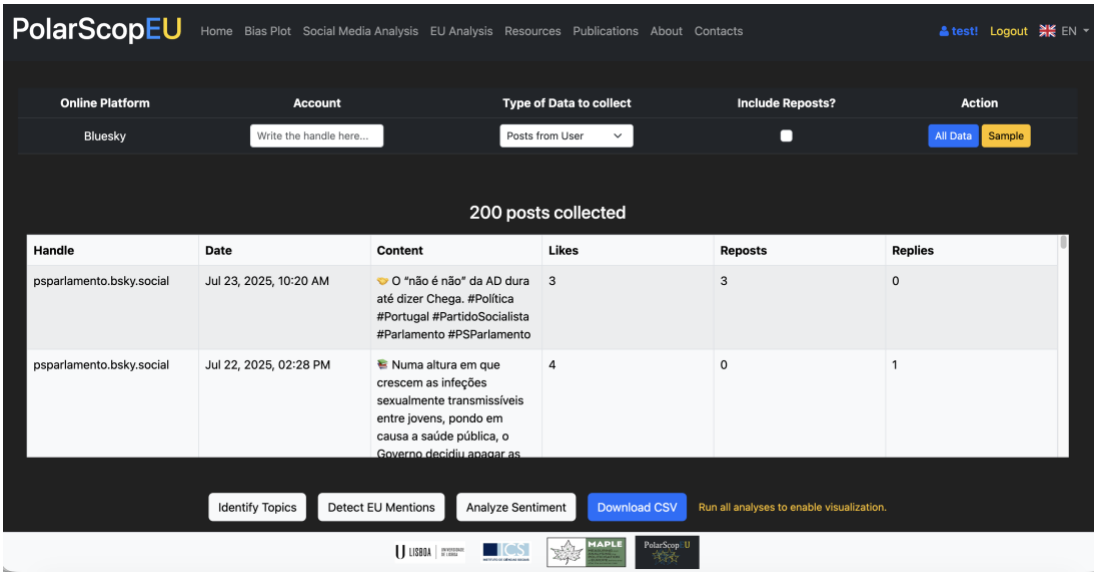


Figure 3 - Example of data collection. A sample of 200 posts from the handle psparlamento.bsky.social.

1.3. Analytical Models and Detection Logic

Topic detection (via the "**Identify Topics**" button) is performed with a fine-tuned transformer model[2] loaded locally through the Hugging Face Transformers library. The topic model and tokenizer are distributed as local model folders, while the underlying classification logic is implemented through a text-classification pipeline. For each post, the model returns a ranked set of candidate labels with associated confidence scores. A post-processing step is then applied to determine the final topic assignment. When the highest-ranked prediction corresponds to "**No Policy Content**", the application does not automatically accept that result. Instead, it checks whether one of the next-ranked candidate topics reaches a sufficient confidence threshold (0.01)[3]. If so, that substantive policy topic is selected; otherwise, the post remains classified as "No Policy Content." This rule was introduced to reduce overly conservative classifications and to better capture posts that do contain policy-related content, even when the model's first prediction is non-policy. The "Identify Topics" updates output table with two new columns, `topic` and `topic_score` with, respectively, the type of policy topic identified in the `Content` column and its confidence score. The `topic` variable has **22** categories: *1) No Policy Content; 2) Agriculture; 3) Banking; Finance and Domestic Commerce; 4) Civil Rights; 5) Culture; 6) Defense; 7) Education; 8) Energy; 9) Environment; 10) Foreign Trade; 11) Government Operations; 12) Health; 13) Housing; 14) Immigration; 15) International Affairs; 16) Labor; 17) Law and Crime; 18) Macroeconomics; 19) Public Lands; 20) Social Welfare; 21) Technology; 22) Transportation*.

[2] The model currently used for this feature is `xlm-roberta-large-portuguese-execorder-cap-v3`. It was fine-tuned by **poltextLAB** (poltextlab.com) on manually coded Portuguese executive orders to identify the 21 **CAP (Comparative Agendas Project)** topics. Although the model was trained on Portuguese-language data, it also performs reasonably well on texts in other languages, which makes it useful for multilingual exploratory analysis. At the same time, this should be understood as an initial solution, and future versions of the application may incorporate alternative or more specialized models to improve performance for this specific task (identify topics on relatively short social media messages).

[3] Due to the short and often context-poor nature of texts on micro-blogging platforms, the model's predictions tend to be relatively conservative. In practice, the highest-ranked label is often **"No policy content"**, even when a post contains some policy-relevant information. To address this, different confidence thresholds were manually tested to improve topic identification. The threshold currently used provided the best balance: if set higher, the model fails to identify some relevant topics; if lowered further, it begins to assign clearly inaccurate labels.

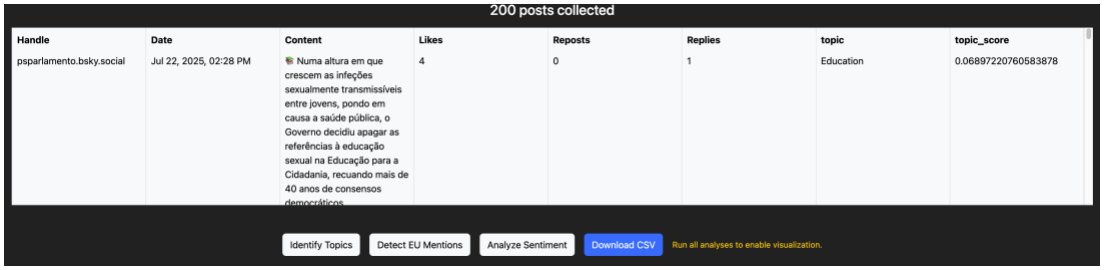


Figure 4 - Example output of the "Identify Topics" button.

References to the European Union (via the "**Detect EU Mentions**" button) are detected with a rule-based dictionary approach derived from the MAPLE project and explained/used in *Silva et al. (2022)*[4]. The app loads multilingual keyword lists and abbreviation patterns from a JSON dictionary, converts them into regular expressions, and flags each post if one or more EU-related patterns are found. This produces an `AboutEU` dichotomous indicator (0=no EU mention; 1= EU mentioned) that can be used independently or in combination with the topic model. For example, users could explore, using the .csv output, what topics are discussed in articles mentioning the EU. Currently, this feature supports only seven languages: **English, Portuguese, Spanish, Greek, German, French, and Dutch**. Support for all official EU languages may be added in future versions. In addition, the process generates a column (`EU_trigger`) that records which language-specific keyword or abbreviation patterns were detected in the text. This can be useful in a post-processing stage to identify possible false positives, for example cases in which Portuguese-language content is triggered by a French keyword or abbreviation.

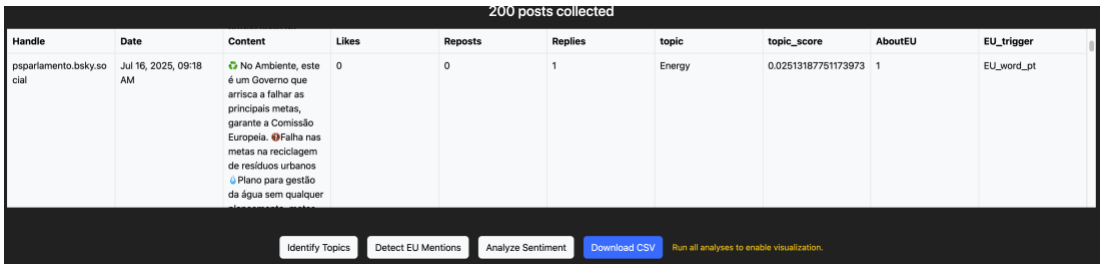


Figure 5 - Example output of the "Detect EU Mentions" button.

[4] Silva, T., Kartalis, Y., & Costa Lobo, M. (2022). Highlighting supranational institutions? An automated analysis of EU politicisation (2002–2017). *West European Politics*, *45*(4), 816-840.

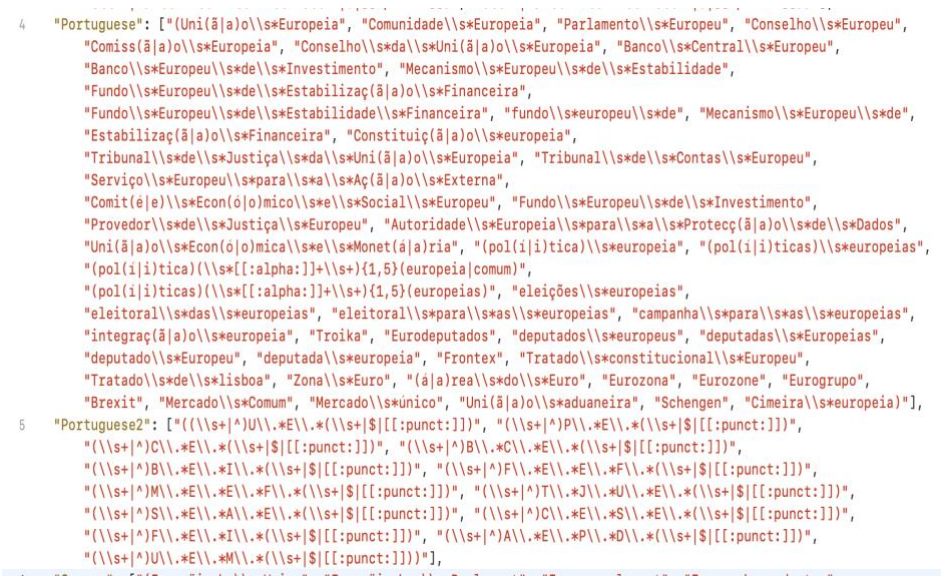


Figure 6 - Example of the JSON dictionary file with EU related expressions

Sentiment analysis (via the "**Analyse Sentiment**" button) is performed through a multilingual transformer model based on XLM-RoBERTa[5], also loaded locally from Hugging Face-compatible files. For each post, the model returns sentiment probabilities, and the application stores the highest-scoring label together with its confidence score. This sentiment layer is especially important for the Bias Plot and the similarity analysis, because it provides the evaluative dimension of the content.

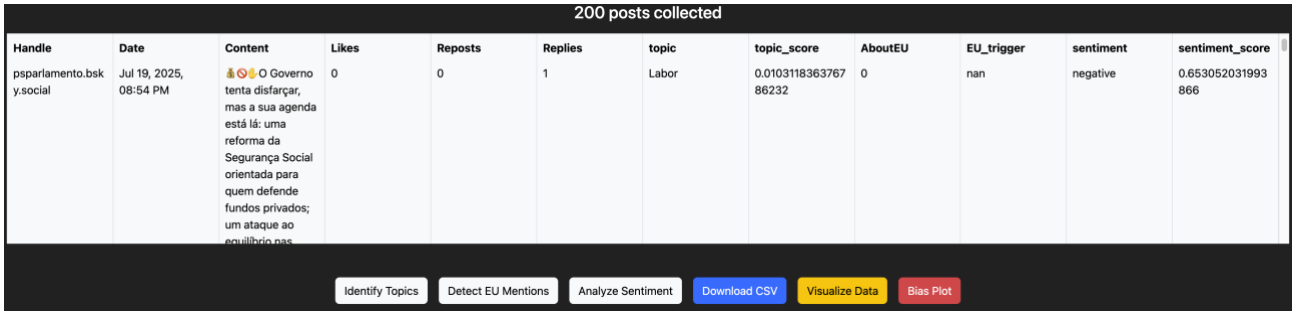


Figure 7 - Example output of the "Analyze Sentiment" button.

After running the three analyses (**policy topic detection**, **EU detection**, and **sentiment analysis**) users can either visualize the results through a set of simple descriptive graphs (via the **"Visualize Data"** button) or generate a **Bias Plot**. Users can also download the dataset at any stage through the **"Download CSV"** button, regardless of whether any of the analyses have been run.

[5] The current version of the application uses the `cardiffnlp/twitter-xlm-roberta-base-sentiment` model for sentiment detection, classifying texts as positive, neutral, or negative.

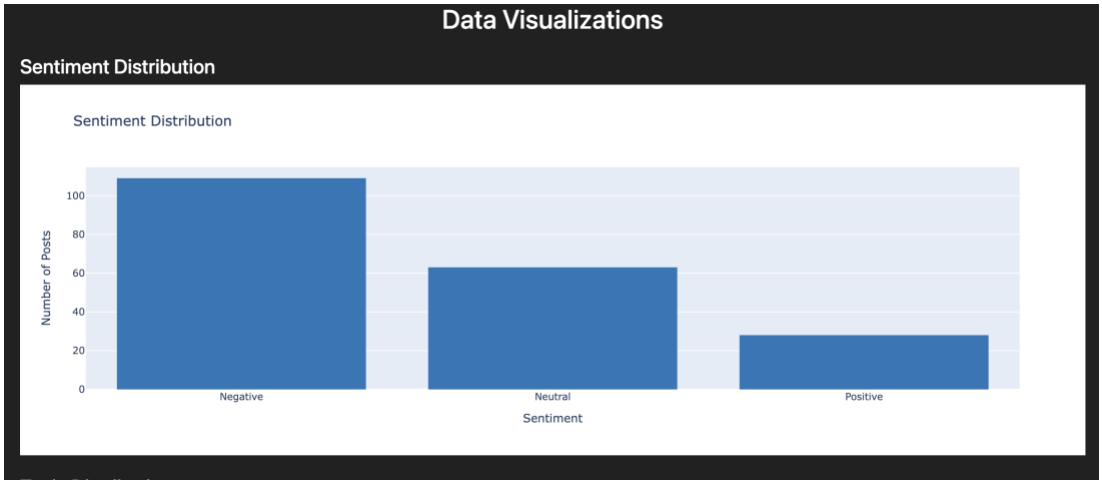


Figure 8 - Example of graph (Sentiment Distribution) available via the "Visualize Data" button.

Policy Topic Detection and Category Overlap

Although the current model and the underlying logic used for topic detection perform well overall, some challenges inevitably remain. Two of the most important are the presence of posts that contain multiple policy topics and cases in which the distinction between categories is inherently blurred. A good example is shown in **Figure 5**, using the text: "No Ambiente, este é um Governo que arrisca a falhar as principais metas, garante a Comissão Europeia. 🚯Falha nas metas na reciclagem de resíduos urbanos 💧Plano para gestão da água sem qualquer planeamento, metas ou financiamento ⛽Aumento do preço dos combustíveis por via do aumento dos impostos". In this case, the app classified the post as **Energy**, although **Environment** could arguably be seen as the more appropriate label, being the main frame of the message ("No Ambiente…").

The model's first three predictions were, in fact: **(1)** *No Policy Content* (0.8593), **(2)** *Energy* (0.0236), and **(3)** *Environment* (0.0184). This example illustrates both the conservative tendency of the model on short social media texts and the difficulty of assigning a single topic when multiple policy dimensions are present. The application is designed to keep the analysis simple and intuitive by assigning one final topic to each post. Users (particualrly the ones using the **PolarscopEU** for research purposes) should therefore be aware of this simplification when interpreting the results and may wish to aggregate related categories in their own analyses.

This is not unusual in applied political text analysis. Indeed, many practical classification schemes and survey-based tools, such as Voting Advice Applications (VAAs), often combine closely related domains into broader categories. For instance, **Energy** and **Environment** are frequently grouped together within a single category such as *Energy, Transport, and Environment*. From this perspective, occasional ambiguity at the boundaries between adjacent topics should be understood as a substantive feature of political texts rather than simply as a model error.

1.4. The Bias Plot

The **Bias Plot,** the core feature of the **PolarscopEU**, was imagined and designed as an intuitive two-dimensional summary of a handle and its network. The x-axis represents tone, operationalized as the proportion of posts with positive or neutral sentiment. A value near 1 therefore indicates a predominantly positive/neutral discourse, while a lower value indicates a more negative one.

The y-axis represents topic diversity. This is not measured simply as the total number of different topics mentioned, but as the number of topics whose relative frequency within a handle's posts exceeds a predefined threshold (this `diversity_threshold` is currently set to 4%). This threshold-based

definition prevents a handle from appearing highly diverse merely because it mentioned many topics once or twice; instead, it highlights topics that occupy a meaningful share of that handle's discourse. It is also a far more straightforward measure of diversity compared to more complex formulas such as the Shannon-Weaver or the inverted Simpson indexes.

Each point in the plot corresponds to one handle. The original handle can be visually distinguished from the network (the single yellow point/dot), and users can hover their mouse above each point for additional information about that specific handle. In addition, users can optionally filter out network accounts with very few posts while always keeping the original handle visible, by selecting a minimum number of posts per handle and using the "**Apply Filter**" button below the plot. This option is useful for generating a cleaner and more analytically robust version of the Bias Plot, since the reliability of the Diversity and Tone indicators tends to increase with the number of posts available for each account.

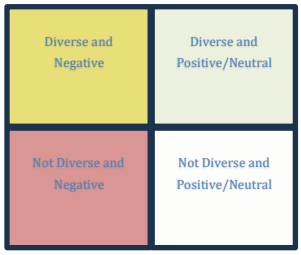
Figure 9 – Example of Bias Plot without (left) and with (right) a filter of 200 minimum posts.

Interpreting the results of the Bias Plot

The interpretation of the Bias Plot is relatively straightforward. Accounts are positioned in a two-dimensional space based on the values of two variables. The first, represented on the **vertical (y-) axis**, is **diversity**. The higher an account is positioned, the more diverse its published content is in terms of policy topics. Given the `diversity_threshold` of **4%**, this axis can theoretically range from **1** to a maximum of **22**, corresponding to the total number of possible topic categories. The **horizontal (x-) axis** represents **tone**. The further to the right an account is placed, the less negative its published content is overall. This value ranges from **0** (completely negative content) to **1** (completely neutral or positive content).

On this basis, handles can be grouped into four broad categories according to their position in the Bias Plot:

1) Diverse and Negative; 2) Diverse and Positive/Neutral; 3) Not Diverse and Negative; 4) Not Diverse and Positive/Neutral.



One-click Bias Plot

To make the app more accessible, we also implemented a one-click feature that generates a Bias Plot for any chosen handle. This feature was designed as the first point of contact with PolarScopEU for new users, and it is the page to which they are first directed after signing in. In order to keep the process simpler and faster, this quick version of the Bias Plot relies on a sample of the **45** most recent posts from each account included in the analysis.

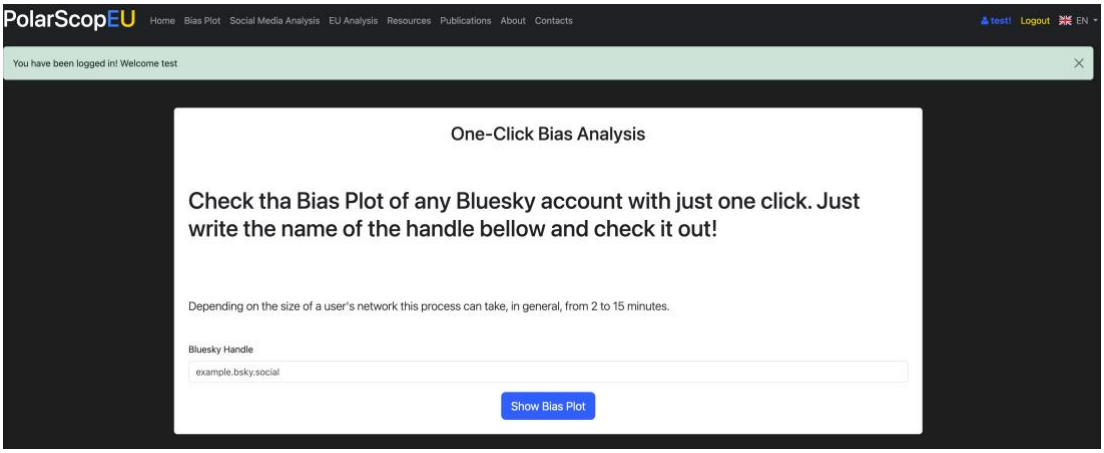


Figure 10 – One-Click Bias Plot feature.

1.5. Similarity Measures

The relative position of two handles in the **Bias Plot**, particularly their proximity to one another, should not be interpreted as a direct measure of similarity in the content they publish. For example, two accounts with the same diversity score of 4 may discuss entirely different policy topics. In this sense, the Bias Plot is useful for comparing general patterns of tone and topic diversity, but not for identifying substantive similarity between accounts. To address this 'limitation', the application also computes similarity between the original handle and the other accounts in the network in two different ways.

The first is **topic-based similarity**. For this, a handle-by-topic matrix is constructed from topic counts, and each row is normalized by the total number of posts from that handle, so that the values represent the proportion of posts devoted to each topic. Cosine similarity is then used to compare the resulting topic-distribution vectors. This means that similarity depends on the relative mix of topics discussed, rather than on raw posting volume.

The second is **topic-plus-tone similarity**. Here, the app constructs a handle-by-topic-sentiment matrix, in which each dimension corresponds to a specific combination such as *Topic A–positive*, *Topic A–neutral*, or *Topic A–negative*. Again, each row is normalized by the total number of posts from that handle, so that the values represent proportions rather than counts, and cosine similarity is computed relative to the original handle. This produces a stricter notion of similarity, because two handles are considered close only if they discuss similar topics in similar tones.

On the **Bias Plot** page, these similarity measures are displayed in ordered tables, allowing the user to identify quickly the accounts that are most similar to, or most distant from, the original handle.

Similarity to Original User (Topic-Based)

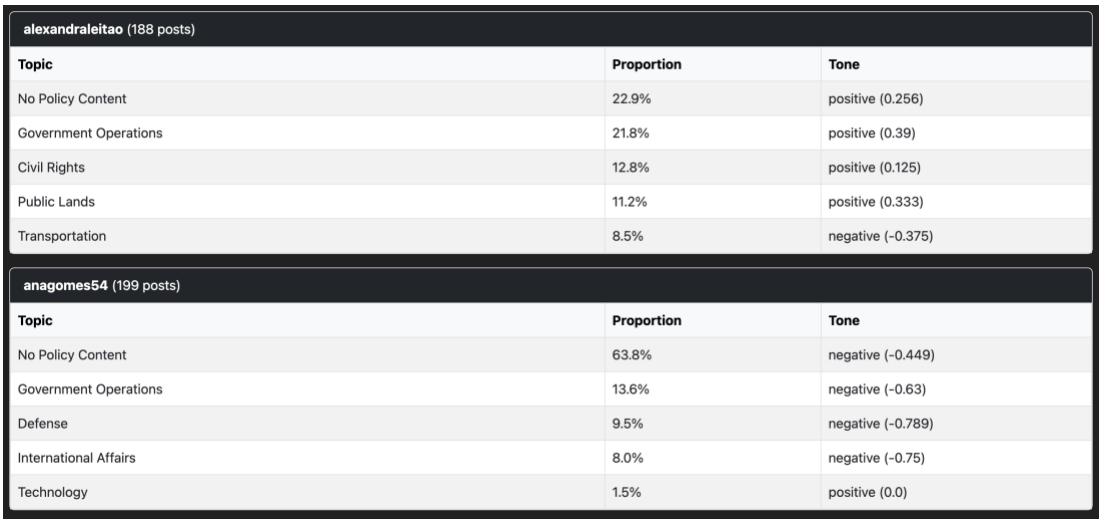
Handle	Cosine Similarity
partidosocialista.bsky.social	0.990
pedronunosantos.bsky.social	0.985
psoeiras.bsky.social	0.951
partisocialiste.bsky.social	0.947
elpais.com	0.943
davideamado.bsky.social	0.922
pedromvaz.bsky.social	0.914
cssdacgtp-in.bsky.social	0.888
ivancgoncalves.bsky.social	0.866

Similarity to Original User (Topic + Tone)

Handle	Cosine Similarity
elpais.com	0.891
politico.eu	0.837
davideamado.bsky.social	0.809
pedromvaz.bsky.social	0.672
pedromtcosta.bsky.social	0.668
psoeiras.bsky.social	0.666
nytimes.com	0.663
washingtonpost.com	0.661
cnn.com	0.644

Figure 11 - Similarity scores of the followed handles relative to the original handle.

In addition to the similarity tables, the Bias Plot page also reports, for each handle, the five most frequent topics in that account's posts. For each of these topics, the application then considers only the subset of posts assigned to that topic and calculates: **(1)** the **dominant tone**, that is, the sentiment category that appears most often among those posts (negative, neutral, or positive), and **(2)** an **average tone value**, shown in parentheses. This value is calculated by assigning numerical scores to the sentiment labels (**negative = -1, neutral = 0, positive = 1**) and then taking the mean across all posts belonging to that topic. The resulting score ranges from **-1** to **+1**, where values closer to **-1** indicate a more negative overall tone, values near **0** indicate a more neutral balance, and values closer to **+1** indicate a more positive overall tone. In this way, the table captures both the most frequent sentiment category and the broader tonal tendency within each topic.



alexandraleitao (188 posts)

Topic	Proportion	Tone
No Policy Content	22.9%	positive (0.256)
Government Operations	21.8%	positive (0.39)
Civil Rights	12.8%	positive (0.125)
Public Lands	11.2%	positive (0.333)
Transportation	8.5%	negative (-0.375)

anagomes54 (199 posts)

Topic	Proportion	Tone
No Policy Content	63.8%	negative (-0.449)
Government Operations	13.6%	negative (-0.63)
Defense	9.5%	negative (-0.789)
International Affairs	8.0%	negative (-0.75)
Technology	1.5%	positive (0.0)

Figure 12 - Tables with the 5 most frequent topics and their respective tone.

1.6. Resources

The application also includes a Resources feature that allows authorized users to collect and store social media data from a curated list of relevant political and media actors, such as parties, politicians, and news organizations. This feature is designed to support the systematic construction of datasets for later analysis. Those authorized users can select handles from the internal directory, retrieve their public posts, and save the results directly in the application's dynamic database. During this process, posts are deduplicated using their unique post identifier whenever available, which helps ensure that repeated collections do not generate duplicate records. In this way, the Resources section functions as a practical data-building tool, enabling users to assemble cleaner and more consistent datasets for comparative analysis across actors, countries, and time periods. While only selected users are authorized to add new entries to the directory, the table itself remains visible to all users for consultation (i.e., a list of relevant political/media, from different countries, and their respective handles).

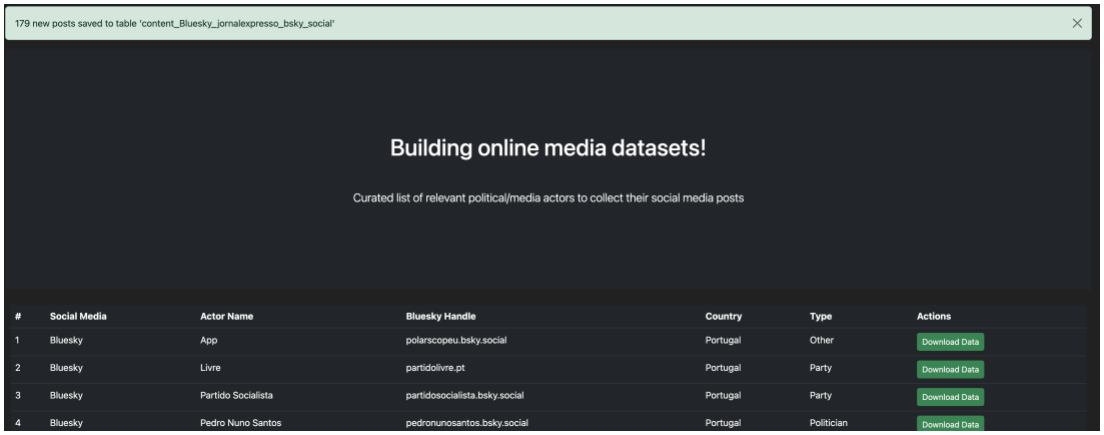


Figure 13 – Table with a curated list of relevant political/media actors in EU.

1.7. Notes on Interpretation and Limitations

The outputs of **PolarScopEU** should be interpreted as **analytical indicators** rather than definitive measures of ideology or political position. Topic classification depends on the quality and coverage of the fine-tuned model, while sentiment classification can be sensitive to sarcasm, ambiguity, multilingual phrasing, and very short posts. Moreover, the purpose of the application is to provide an overall and descriptive picture of the type of political content, tone, and EU salience associated with different accounts, rather than an evaluative or normative assessment of that content. Users are therefore encouraged to validate the results further and to explore additional ways of complementing the data and analyses provided by the app.

The Bias Plot is especially useful as a comparative device. It allows users to observe whether a network clusters around a narrow set of issues, whether discourse is broadly positive or negative in tone, and whether some accounts differ substantially from the rest. Its value lies less in assigning a fixed political score and more in helping researchers and users reflect on discourse patterns, bias and exposure.

Overall, the application offers a tested framework for collecting and structuring **Bluesky data**, applying NLP models to political communication, and transforming those outputs into interpretable visual and relational measures.

1.8. Data Handling and Security Measures

The application was designed to separate clearly between **user authentication data**, **handle metadata**, and **collected social media content**. Authentication information is stored in a dedicated relational database and includes only the minimum fields required to manage access, such as username, (unverified) email address, and password hash. Passwords are not stored in plain text; instead, they are hashed using the **bcrypt** algorithm, which also applies salting automatically. This reduces the risk of credential compromise and follows widely accepted security practices for password storage.

The social media data collected by the application is limited to **publicly available content** retrieved through platform APIs. In the current implementation, the main source is **Bluesky**, from which the app collects posts from a specified handle and, when requested, from the accounts followed by that handle. The collected data are stored separately from authentication data, in a dedicated content database. This database structure distinguishes between **user-specific working tables**, which are overwritten when a new search is run, and **handle-specific cumulative tables**, which can be appended over time for longitudinal analysis. This separation reduces the risk of accidental exposure or mixing of sensitive user information with research data.

Access to the application is protected through authenticated login, and some actions can be restricted to specific users or roles. In practice, the app already supports differentiated permissions such as administrator access for certain management tasks and, as already mentioned, contributions to the **Resources** directory. User-generated datasets and analysis results are associated with the authenticated account that created them, which helps preserve data isolation. In addition, the application does not require unnecessary personal data for its analytical functions, and the processing logic focuses on aggregation, classification, and visualization of public political communication rather than on profiling private individuals.

The system also incorporates several practical safeguards at the processing stage. Duplicate posts are identified and removed using unique post identifiers where available. Temporary or session-level information is not treated as long-term research data. Sensitive credentials required for API access are intended to be stored outside the source code, for example through environment variables or protected configuration files.

1.9. Ethical Considerations and GDPR Alignment

The application is designed to analyze **publicly available political communication** on online platforms, rather than private or restricted user content. In the current implementation, the main data source is **Bluesky**, where content can be accessed without requiring direct interaction with individual users. The app was developed so that users are not required to login with their own personal Bluesky account. The focus is on posts published by actors whose communication is intended for broad audiences, such as political parties, media organizations, and other publicly visible accounts. This reduces the ethical risks associated with covert data collection or the processing of content shared under strong expectations of privacy.

Even when data is technically, and exclusively, public, the application still adopts a cautious research-oriented approach to reuse. Its purpose is not to commercialize or redistribute content, but to support the systematic analysis of issue salience, tone, and network-level patterns in political communication. The app emphasizes aggregated outputs, comparative indicators, and visual summaries rather than

highlighting single posts or exposing individuals unnecessarily. In this sense, it follows the principle that "publicly available" content does not automatically imply unrestricted or context-free reuse.

With regard to data protection, the application is aligned with the general principles of the **General Data Protection Regulation (GDPR)**, especially those relating to **data minimization, purpose limitation, access control, and security of processing**. As we mentioned, user registration data are kept separate from analytical content data, and only the minimum amount of personal information necessary for account management is stored. The architecture also supports future safeguards such as stronger role-based permissions, clearer audit trails, and improved deletion or retention controls. Taken together, these design choices aim to ensure that the platform remains both analytically useful and ethically responsible.

1.10.Technical Snapshot

Component	Implementation
Web framework	Flask + Jinja templates
Database layer	Flask-SQLAlchemy / SQLAlchemy
Bluesky collection	atproto / AT Protocol SDK
Topic model	poltextlab/xlm-roberta-large-portuguese-execorder-cap-v3
Sentiment model	cardiffnlp/twitter-xlm-roberta-base-sentiment
EU detection	Dictionary-based multilingual regex matching (MAPLE)
Interactive plots	Plotly Express
Similarity measure	Cosine similarity on normalized vectors

2. Additional Feature (EU Politicization Analysis)

2.1. Feature Overview (EU Politicization Analysis)

PolarScopEU is a proof-of-concept output of **MAPLE**, a completed ERC-funded project that examined the effects of EU politicization and the eurozone crisis on the voting behaviour of European citizens (Lobo, 2023). One of **MAPLE**'s core research tasks was the assessment of EU politicization in media coverage and parliamentary debates. The second main feature of **PolarScopEU**, **EU Politicization Analysis**, builds directly on the experience, data, and methodological outputs developed within MAPLE in order to make a

new analytical tool publicly available for the study of the *EU dimension* in textual content, including media articles, social media posts, parliamentary speeches, and related sources.

This feature can be used simply by selecting a local dataset. The only requirements are that the file must be in **.csv** format, contain two columns labelled `Unique_ID` and `text`, and have fewer than **5000 rows** (users may analyze multiple files one at the time). The application then uploads the file, runs the analysis automatically, and generates a new **.csv** file for download containing the coded variables. These variables are produced by fine-tuned models trained on political communication datasets developed within **MAPLE**.

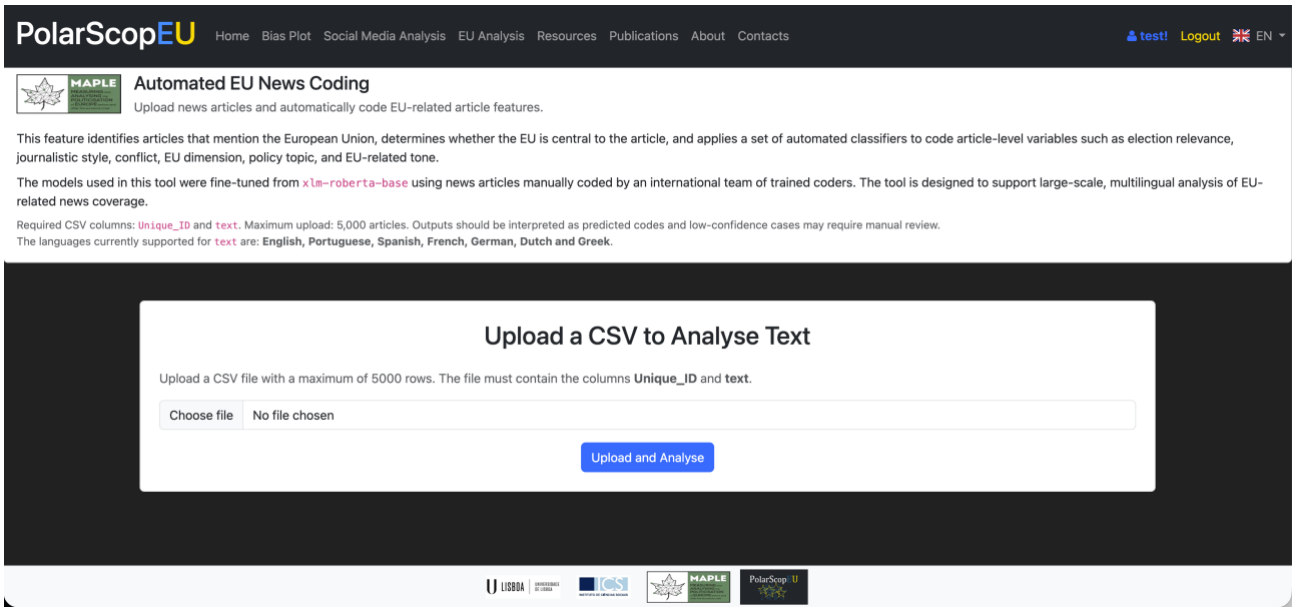


Figure 14 - EU Analysis feature of PolarscopEU.

When the analysis is complete (a process that may take some time) users have two output options. The first is "**Download coded articles**". This is the main output of the feature and corresponds to the original input **.csv** file with the newly generated variables appended to it. The second is "**Download coded sentences**". This produces a new dataset in which each row corresponds to a unique sentence mentioning the EU, extracted from all articles classified as About EU. Each sentence is accompanied by its corresponding tone analysis and the Unique_ID of the corresponding text that it was extracted from.

Figure 15 - Two possible outputs for the EU Analysis feature

This section summarizes the main components of that feature, including the models used, their training logic, interpretation and evaluation metrics. The pipeline combines **EU keyword detection**, **fine-tuned transformer classifiers**, and **sentence-level sentiment analysis**. Its purpose is to support the automated

coding of large collections of texts for EU Politicisation studies (topic salience and contestation) as well as for several article-level variables originally derived from a human-coded codebook, used to create the **MAPLE Media dataset 2: Manual analysis of EU salience and contestation in the media**. For a use case and further details on this dataset, see Silva et al. (2023)[6].

(EU analysis feature) Pipeline overview:

```
Raw article text
→ EU keyword detection

If an EU-related keyword or abbreviation is detected:
    → EU_CENTRAL classification

    If classified as centrally about the EU:
        → ELECT classification
        → NA_NEWSSTYLE classification
        → NA_CONF1 classification
        → NA_EUDIMENS classification

        If NA_EUDIMENS = EU policies or domesticated policies:
            → NA_MAINTOPIC classification

        → Extraction of EU-related sentences
        → Sentence-level sentiment/tone classification
        → Article-level tone aggregation
```

The system intentionally does not apply all classifiers to every single article. Each model is applied only where the variable is conceptually meaningful, following the original coding structure. For example, it only makes sense to identify a policy topic in the articles about "**EU policies**" or "**Domesticated policies**".

2.2. EU keyword detector

Purpose

The EU keyword detector identifies articles that contain at least one explicit EU-related term or abbreviation. Its purpose is not to determine whether an article is centrally about the EU, but rather to create a candidate pool for the EU_CENTRAL model. For this reason, it is important to rely on as complete a list of relevant terms and abbreviations as possible. The list currently used in the application was initially drawn from an existing codebook[7] that employed the same approach to identify EU-related articles, and was later expanded for the purposes of the MAPLE project.

Method

The detector uses a multilingual JSON dictionary of EU-related expressions (shown in Figure 16). The dictionary includes full terms and abbreviations for **seven** languages (**English, Portuguese, German, Greek, Spanish, French, and Dutch**).

[6] Silva, T., Kartalis, Y., & Nina, S. R. (2023). Political implications of the different manifestations of politicization: Examining the news coverage of the EU in Greece and Portugal before and during the Eurozone crisis. In Anne-Marie Houde, Thomas Laloux, Morgan Le Corre Juratic, Heidi Mercenier, Damien Pennetreau, and Alban Versailles (Eds), *The Politicization of the European Union: From Processes to Consequences*, Bruxelles, Éditions de l'Université de Bruxelles.

[7] Maier, Michaela, Adam Silke, Claes H. De Vreese, Melanie Leidecker-Sandmann, Beatrice Eugster, Franzisca Schmidt, and Eva Antl-Wittenberg (2014). 'Politicization of EU Integration. Codebook for a Content Analysis of Media and Party Communication'.

Examples: European Union; European Parliament; European Commission; União Europeia; Unión Europea; Europäische Union; Ευρωπαϊκή Ένωση; Brexit; Schengen; Eurozone; Frontex

Abbreviations are handled separately from full terms. Full terms are matched case-insensitively, whereas abbreviations are matched case-sensitively in order to reduce false positives (e.g., *"eu gosto"* or *"Eu estou"* in Portuguese). The implementation also de-duplicates equivalent regex patterns across languages before matching, thereby avoiding multiple counts for shared terms such as *Brexit* or *Schengen*. In addition, the regex patterns are designed to capture different orthographic variants of abbreviations, including the presence or absence of dots (e.g., E.U., E.U, and EU) as well as social media hashtags (e.g., #EU).

Output columns

```
eu_mentioned_rule
eu_term_match_count
eu_abbr_match_count
eu_total_match_count
eu_term_sources
eu_abbr_sources
```

Interpretation

`eu_mentioned_rule` = True means the article contains at least one EU-related expression. It does not mean the article is about the EU.

2.3. EU_CENTRAL model

Purpose

Determines whether an article that mentions the EU is also about the EU or only mentions the EU in passing.

Training logic

Binary XLM-RoBERTa classifier trained on hand-coded article-level data. It is applied only after the keyword detector identifies an EU-related expression. Coders were asked if EU, EU institutions or EU actors were a central topic/aspect in the article or simply mentioned. The main output variable of this model `pred_EU_Central` has the following codes:

Codes (`pred_EU_Central`)

Code	Interpretation
0	EU mentioned only
1	About EU

Interpretation

"About EU" means that the European Union, its institutions, membership, policies, political processes, or EU-related consequences are central to the article. **"EU mentioned only"** means that the article contains at least one EU-related term, but the EU is not the main focus. The final coded dataset — that is, the downloadable `.csv` file — also includes confidence scores for all possible outcomes of each model, with transparent and intuitive variable names. In the case of the **EU_CENTRAL** model, these are the probability of the article being coded as **EU mentioned only** (`prob_eu_central_0_eu_mentioned_only`) and the probability of it being coded as **About EU** (`prob_eu_central_1_about_eu`).

Model performance

Metric	Value
Accuracy	~0.82
Macro F1	~0.82

2.4. ELECT model

Purpose

This model identifies whether an EU-related article is connected to a national legislative election or to the electoral campaign. This can be interpreted as a measure of **EU–election linkage** or, more broadly, **EU electoral politicization**. More concretely, in the context of national legislative elections, this variable helps us assess how salient the EU dimension is within the campaign. In that sense, it can also be understood as an indicator of the degree to which the EU becomes politicized during electoral competition. The model may also perform less well outside the context of the six MAPLE countries (Belgium, Germany, Greece, Ireland, Portugal, and Spain), as some of the predictive patterns learned during training may reflect country-specific party names, political actors, and campaign dynamics. The same can be the case for non-legislative national elections, such as the European Parliament elections.

Training logic

Binary XLM-RoBERTa classifier trained only on articles where the EU was relevant and the original ELECT variable was valid. Training distribution after filtering: 9,448 not election/campaign related and 2,189 election/campaign related[8].

Codes

Code	Interpretation
0	Not election/campaign related
1	Election/campaign related

Interpretation

Election/campaign related means the EU issue is linked to the legislative election or campaign context, including campaign discussion, party positioning, electoral strategy, voter choice, or EU issues in electoral competition. Coders were simply asked if the article, published in the 30 days before an election, was clearly about an election or electoral campaign.

Model performance

Metric	Value
Validation accuracy	0.894
Validation macro F1	0.826
Validation weighted F1	0.894
Test accuracy	0.88
Test macro F1	0.81
Test weighted F1	0.88
F1: Not election/campaign related	0.93
F1: Election/campaign related	0.68

[8] More information about the dataset and label distribution is given later in a specific section.

2.5. NA_NEWSSTYLE binary model

Purpose

Identifies whether an article is written in a primarily descriptive or interpretive style. Because this variable refers specifically to journalistic style, the model is tailored to news content and should not be expected to perform well, or necessarily to be meaningful, when applied to other types of text.

Training logic

The original variable had three categories, but code 3, cannot be determined, was rare and conceptually less useful for model training (coders were also instructed to avoid that label). The model was trained as a binary XLM-RoBERTa classifier after removing code 3. Binary distribution: 7,284 descriptive and 3,211 interpretive.

Codes

Code	Interpretation
0	Descriptive; maps to original NA_NEWSSTYLE = 1
1	Interpretive; maps to original NA_NEWSSTYLE = 2

Interpretation

Descriptive means the article mainly reports facts, events, or statements. Coders should use this code when the article told ´what happened´ in a rather straightforward, descriptive style and focused on known facts. Interpretive means the article contains substantial explanation, analysis, evaluation, contextualization, or journalistic interpretation. Coders should consider an article interpretative when the article analyzed, evaluated, interpreted a situation, for example ascribing motives for certain actions and decisions.

Model performance

Metric	Value
Validation accuracy	0.823
Validation macro F1	0.782
Validation weighted F1	0.819
Test accuracy	0.81
Test macro F1	0.78
Test weighted F1	0.81
F1: Descriptive	0.85
F1: Interpretive	0.71

2.6. NA_CONF1 model

Purpose

Identifies whether an article contains some elements of **conflict** or **disagreement**. Its main purpose is to capture whether an article about the EU is framed in terms of conflict, and it can therefore be used as a proxy for EU contestation.

Training logic

Binary XLM-RoBERTa classifier trained on articles with valid conflict/disagreement coding. The binary distribution was 6,373 articles coded as “No conflict/disagreement” and 5,340 coded as “Conflict/disagreement present”.

Codes

Code	Interpretation
0	No conflict/disagreement
1	Conflict/disagreement present

Interpretation

Conflict/disagreement present means the article includes explicit disagreement, contestation, conflict, opposition, criticism, controversy, or dispute involving actors, policies, institutions, or political positions. Coders were asked whether an article reflected reflect disagreement between parties, individuals, groups, organizations, or countries, i.e., that different actors have different opinions and disagree with each other.

Model performance

Metric	Value
Validation accuracy after epoch 1	0.728
Validation macro F1 after epoch 1	0.722
Validation weighted F1 after epoch 1	0.726
Test accuracy	0.77
Test macro F1	0.76
Test weighted F1	0.77
F1: No conflict/disagreement	0.79
F1: Conflict/disagreement present	0.74

2.7. NA_EUDIMENS model

Purpose

Classifies the main **EU dimension** addressed in an EU-related article. This typology is based primarily on the work of **Peter Mair**[9], who originally distinguished between a **Europeanization** dimension (dealing with the creation, consolidation, and territorial reach of EU institutions) and a **penetration** dimension, referring to the implications of EU legislation for the domestic sphere. In the original coding, we rely on an updated version of this typology[10] that further divides the penetration dimension into **"EU policies"** and **"Domesticated policies."** While this distinction was particularly useful in the context of the Eurozone crisis and the austerity measures imposed on debtor countries, studied in **MAPLE**, it may be much more blurred in other contexts. For that reason, users may wish to combine these two categories into a single **"policies"** variable in their own analyses.

Training logic

Four-class XLM-RoBERTa classifier. Valid filtered distribution: 1,939 membership; 1,036 constitutional structure; 5,572 EU policies; 3,016 domesticated policies.

Codes

Code	Interpretation
0	Membership; maps to original NA_EUDIMENS = 1
1	Constitutional structure; maps to original NA_EUDIMENS = 2
2	EU policies; maps to original NA_EUDIMENS = 3
3	Domesticated policies; maps to original NA_EUDIMENS = 4

[9] Mair, P. (2004) 'The Europeanization dimension', Journal of European Public Policy, 11(2), pp. 337–348.
[10] Hurrelmann, A., Gora, A., & Wagner, A. (2015). The politicization of European integration: More than an elite affair?. *Political studies*, *63*(1), 43-59.

Interpretation

The **Membership** label refers to discussions about the geographical reach of the EU, including whether a country should join, remain in, or leave the European Union or the Eurozone, as well as the perceived costs and benefits of membership. **Constitutional structure** refers to discussions about the objectives and responsibilities of the EU, its institutional arrangements, decision-making processes, and the functioning of EU institutions more generally, including debates about democratic legitimacy or institutional reform. **EU policies** refers to issues and policies that emerge from EU-level institutions (e.g., legislative, executive, or judicial) and that have implications across member states. These are policies that are on the agenda of EU institutions themselves (e.g., discussion of EU data protection law). **Domesticated policies**, on the contrary, refers to policies emerging from the domestic political bodies that arise as a consequence of EU membership, EU rules, or EU constraints, such as national budget cuts or austerity measures adopted in response to Eurozone requirements.

Model performance

Metric	Value
Validation accuracy	0.732
Validation macro F1	0.693
Validation weighted F1	0.730
Test accuracy	0.72
Test macro F1	0.68
Test weighted F1	0.72
F1: Membership	0.72
F1: Constitutional structure	0.59
F1: EU policies	0.78
F1: Domesticated policies	0.65

2.8. NA_MAINTOPIC model

Purpose

Classifies the substantive policy topic of articles where the EU dimension concerns EU policies or domesticated policies.

Training logic

Trained as a 14-class XLM-RoBERTa classifier only where EUSALIENCE = 2, NA_EUDIMENS is EU policies or domesticated policies, and NA_MAINTOPIC is valid. The variable has 14 classes and is imbalanced.

Original code	Topic
1	Economy and Work
2	Finances and Taxes
3	Health
4	Migration and Immigration
5	National Security
6	Society, Social rights, Religion, and culture
7	Environment protection
8	Transport and Energy
9	Law and Order
10	Foreign Policy
11	Institutional design

12	Welfare and Family
13	Education
14	Other

Model performance

Metric	Value
Validation accuracy	0.716
Validation macro F1	0.570
Validation weighted F1	0.697
Test accuracy	0.72
Test macro F1	0.57
Test weighted F1	0.70

Class	F1
Economy and Work	0.72
Finances and Taxes	0.82
Health	0.78
Migration and Immigration	0.85
National Security	0.44
Society/Social rights/Religion/Culture	0.18
Environment protection	0.78
Transport and Energy	0.72
Law and Order	0.51
Foreign Policy	0.78
Institutional design	0.45
Welfare and Family	0.00
Education	0.75
Other	0.14

Important note: This model performs well on several major and distinctive categories, but weakly on small or heterogeneous classes such as Welfare and Family, Other, and "Society/Social rights/Religion/Culture". These categories should be interpreted cautiously, and future versions may aggregate weak or conceptually related categories.

2019 online news output

The model was applied only to articles predicted as EU policies or domesticated policies. Eligible for `NA_MAINTOPIC`: 3,760.

Predicted topic	Count
Finances and Taxes	1,006
Economy and Work	622
Foreign Policy	600
Migration and Immigration	329
Environment protection	269
Institutional design	262
Law and Order	164
Transport and Energy	147
Health	108
Other	89
Society, Social rights, Religion, and culture	76
National Security	50
Education	38
Welfare and Family	0

2.9. EU sentence-level tone model

Purpose

The original article-level NA_TONEEU classifier was not used in the final pipeline because its weaker performance. Instead, EU tone is derived from sentence-level sentiment predictions applied only to sentences that explicitly mention the EU. This design reflects the fact that tone toward the EU is often localized in a few sentences, while the rest of an article may discuss broader domestic politics, elections, or policy issues. The same design was used in one of MAPLE outputs[11], to assess the politicization of the EU dimension.

MAPLE approach

The implemented tone variable should be understood as EU-related sentence sentiment aggregated to the article level. The procedure first identifies articles predicted as About EU by the EU_CENTRAL classifier, then extracts only sentences containing EU keyword matches, predicts the sentiment of those sentences, and finally aggregates the sentence-level results into an article-level inferred EU tone.

Model used

The sentence-level tone component uses the same model used to assess sentiment in the social media content, a multilingual transformer model based on XLM-RoBERTa[12]. The model has three labels: 0 = negative, 1 = neutral, and 2 = positive.

Procedure

For each article predicted as About EU, the tone script performs the following six steps:

1. Split the article text (i.e., the content in the `text` variable) into sentences using punctuation-based rules and light preprocessing[13].
2. Apply the same de-duplicated multilingual EU keyword detector used in the article-level pipeline to each sentence.
3. Keep only sentences containing at least one EU term or EU abbreviation match.
4. Run the local sentence-level sentiment model on the extracted EU mentioned sentences.
5. Save a sentence-level tone file with sentence IDs, sentence text, predicted tone label, confidence, and class probabilities (can be saved as `.csv` file by the user).
6. Aggregate sentence-level sentiment predictions to a single article-level (inferred) EU tone variable.

Sentence-level output

The sentence-level .csv output (available with the "**download coded sentences**" button) contains one row per EU-mention sentence and includes the following main columns:

- `Unique_ID` (same as the input document. The same value can appear in different rows if one article has multiple EU-mentioned sentences)
- `sentence_id` (unique sentence id)
- `sentence_text`[14] (the content that will be analysed)
- `eu_sentence_term_match_count` (number of EU terms found in the sentence)

[11] Silva, T., Kartalis, Y., & Costa Lobo, M. (2022). Highlighting supranational institutions? An automated analysis of EU politicisation (2002–2017). *West European Politics*, *45*(4), 816-840.
[12] `cardiffnlp/twitter-xlm-roberta-base-sentiment`.
[13] This includes things like normalizing repeated whitespaces or discard short fragments of text.
[14] The user is free to decide what the content is (e.g., the title of a news article, the article body, a press release). The models will perform the best with news media articles. The variable can also combine different parts of the article (title and lead paragraph).

- `eu_sentence_abbr_match_count` (number of EU abbreviations found in the sentence)
- `eu_sentence_total_match_count` (Total number of matches, term + abbreviations)
- `tone_pred` (predicted sentiment. This corresponds to the tone – Negative, Neutral or Positive- with the highest confidence score)
- `tone_pred_label` (predicted sentiment – label)
- `tone_sentence_confidence` (confidence score for tone prediction)
- `prob_tone_negative` (score for negative tone prediction)
- `prob_tone_neutral` (score for neutral tone prediction)
- `prob_tone_positive` (score for positive tone prediction)

Article-level aggregation rule

The article-level tone variable is derived from the distribution of negative, neutral, and positive EU-mention sentences within each article. In line with the tone variable in the training dataset, we account for the possibility of an article being Balanced/Mixed. The aggregation rule is the following:

- If an article has at least one negative EU sentence and at least one positive EU sentence, the article is classified as Balanced/Mixed.
- If an article has at least one negative EU sentence and no positive EU sentence, the article is classified as Negative.
- If an article has at least one positive EU sentence and no negative EU sentence, the article is classified as Positive.
- If all extracted EU sentences are neutral, the article is classified as Neutral.

Article-level output columns

After the sentence-level analysis and aggregation step, the following columns are appended to the main output `.csv` file:

- `eu_tone_sentence_count` (number of sentences mentioning the EU)
- `tone_negative_count` (number of sentences with negative sentiment)
- `tone_neutral_count` (number of sentences with neutral sentiment)
- `tone_positive_count` (number of sentences with positive sentiment)
- `tone_article_aggregate` (tone/valence of the article towards the EU)
- `tone_article_confidence` (average value of confidence scores from the sentence-level analysis)

Important note: The article-level tone confidence (`tone_article_confidence`) is currently calculated as the **mean sentence-level confidence** across the EU-mention sentences extracted from the article. It should therefore **not** be interpreted as the probability of a single article-level tone classifier.

As an example, an article classified as **About EU** contains two EU-mention sentences: one **Negative** with a confidence score of **0.4** and one **Positive** with a confidence score of **0.6**. The aggregated article tone (`tone_article_aggregate`) of that article will be **Balanced/Mixed**, while the article-level confidence (`tone_article_confidence`) will be the mean of the two sentence scores, which is **(0.4 + 0.6) / 2 = 0.5**.

This means that a higher `tone_article_confidence` does not necessarily indicate a more reliable article-level tone classification. For instance, two articles may receive the same aggregated tone label while differing substantially in the distribution of sentence-level predictions. More generally, this variable

should be understood only as a rough summary of the confidence of the sentence-level tone predictions used in the aggregation process, rather than as a direct measure of certainty in the final article-level label. Users may therefore wish to explore alternative ways of constructing `tone_article_confidence`, for example by excluding neutral sentences or by weighting only the dominant valence/tone direction.

2.10. Variables not used

NA_TONEEU article-level classifier

An article-level XLM-R classifier was tested for NA_TONEEU variable, a variable measuring tone towards the EU. The variable's performance was poor, even when recoded as a dichotomous positive/negative variable, and it is not used in the final pipeline.

Metric	Value
Accuracy	0.49
Macro F1	0.27
Weighted F1	0.41
F1: Negative	0.00
F1: Positive	0.00

NA_ECONCONDI

The original 5-class NA_ECONCONDI model (does the article evaluates the economic situation, comparing it to a previous point in time) was tested but did not perform well. A binary economic-worsening version was also tested but missed most true worsening cases, so the variable was dropped from the final pipeline.

Model	Metric	Value
5-class	Accuracy	0.53
5-class	Macro F1	0.30
5-class	Weighted F1	0.50
Binary worsening	Accuracy	0.78
Binary worsening	Macro F1	0.57
Binary worsening	Weighted F1	0.73
Binary worsening: Worse class	Precision	0.65
Binary worsening: Worse class	Recall	0.17
Binary worsening: Worse class	F1	0.28

2.11. Final model set used in the app

Used in final pipeline	Not used in final pipeline
EU keyword detector	NA_TONEEU article-level classifier
EU_CENTRAL	NA_ECONCONDI
ELECT	
NA_NEWSSTYLE binary	
NA_CONF1	
NA_EUDIMENS	
NA_MAINTOPIC	
EU sentence-level tone aggregation	

2.12. Variables Included in the CSV Output

Unique_ID (retained from original input file: unique identifier of the text/article)
text *(retained from original input file: original text content of the article/post)*
pred_EU_CENTRAL (predicted EU_CENTRAL class code)
pred_EU_CENTRAL_label (predicted EU_CENTRAL class label)
confidence_EU_CENTRAL (confidence score for EU_CENTRAL prediction)
pred_ELECT (predicted ELECT class code)
pred_ELECT_label (predicted ELECT class label)
confidence_ELECT (confidence score for ELECT prediction)
pred_NA_NEWSSTYLE (predicted news style class code)
pred_NA_NEWSSTYLE_label (predicted news style class label)
confidence_NA_NEWSSTYLE (confidence score for news style prediction)
pred_NA_CONF1 (predicted conflict class code)
pred_NA_CONF1_label (predicted conflict class label)
confidence_NA_CONF1 (confidence score for conflict prediction)
pred_NA_EUDIMENS (predicted EU dimension class code)
pred_NA_EUDIMENS_label (predicted EU dimension class label)
confidence_NA_EUDIMENS (confidence score for EU dimension prediction)
pred_NA_MAINTOPIC (predicted main topic class code)
pred_NA_MAINTOPIC_label (predicted main topic class label)
confidence_NA_MAINTOPIC (confidence score for main topic prediction)
tone_article_aggregate (aggregated article-level tone)
tone_article_confidence (average confidence of sentence-level tone predictions)
eu_tone_sentence_count (number of EU-related sentences used for tone aggregation)
review_recommended (flag indicating cases recommended for manual review)

2.13. Confidence scores and review flags

Model confidence is prediction-level, not model-level. A model can have good overall validation/test metrics while still producing low-confidence predictions for some individual articles. The pipeline stores confidence columns for each classifier, for example confidence_EU_CENTRAL, confidence_ELECT, confidence_NA_EUDIMENS, and confidence_NA_MAINTOPIC.

The current review flag is based on the minimum available model confidence for an article: **review_recommended = True** whenever at least one model prediction has a confidence score below 0.60. In other words, the flag identifies cases in which one or more predictions fall below the threshold, not cases in which all predictions do.

As it was previously explained, tone confidence is different from classifier confidence. tone_article_confidence is currently the mean confidence across EU sentence-level sentiment predictions, not a single article classifier probability.

2.14. General caution for documentation

The models reproduce selected codebook variables using supervised machine learning trained on articles coded by trained team. Outputs should be interpreted as predicted codes. For high-stakes analysis, low-

confidence cases and theoretically important subsets should be manually reviewed. To assist with this task, the pipeline generates the boolean variable `review_recommended`.

Most classifiers produced high-confidence predictions for the large majority of coded cases. The main exception is NA_MAINTOPIC, where 31.5% of predictions had confidence below 0.60. This is expected because NA_MAINTOPIC is a 14-class very imbalanced topic classification task. Users should treat low-confidence topic predictions as review candidates and consider, as already mentioned, combining some of the less frequent labels.

confidence_EU_CENTRAL n= 19127 below_0.6= 234 share= 0.012
confidence_ELECT n= 8424 below_0.6= 180 share= 0.021
confidence_NA_NEWSSTYLE n= 8424 below_0.6= 408 share= 0.048
confidence_NA_CONF1 n= 8424 below_0.6= 587 share= 0.07
confidence_NA_EUDIMENS n= 8424 below_0.6= 691 share= 0.082
confidence_NA_MAINTOPIC n= 3760 below_0.6= 1184 share= 0.315

2.15. Testing the Models

Results from applying the pipeline to a news dataset collected around the 2019 European Parliament elections (MAPLE Media Dataset 4: Online Media Dataset). The input in the **`text`** field consisted of the article title and lead paragraph.

Metric		Value
Total articles		111,397
EU mentioned articles (EU keyword detector)		19,127 (17.2% of all articles)
Predicted About EU (EU_CENTRAL)		8,424 (44.04% of EU-mentioned articles)
Predicted context of election/campaign (ELECT)		1,590 (18.9% of About EU articles)
Predicted interpretive style (NA_NEWSSTYLE)		3,486 (41.4% of About EU articles)
Predicted conflict (NA_CONF1)		4,148 (49,2% of About EU articles)
Dimensions of EU (NA_EUDIMENS)	Membership	3,743 (44.4% of About EU articles)
	EU Policies	2,429 (28.8%of About EU articles)
	Domesticated Policies	1,331 (15.8% of About EU articles)
	Constitutional Structure	921 (10,9% of About EU articles)
Policy topic (NA_MAINTOPIC)	Finances and Taxes	1,006 (26.8% of articles with Policies)
	Economy and Work	622 (16.5% of articles with Policies)
	Foreign Policy	600 (16% of articles with Policies)
	Migration and Immigration	329 (8.8% of articles with Policies)
	Environment protection	269 (7.2% of articles with Policies)
	Institutional design	262 (7% of articles with Policies)
	Law and Order	164 (4.4% of articles with Policies)
	Transport and Energy	147 (3.9% of articles with Policies)
	Health	108 (2.9% of articles with Policies)
	Other	89 (2.4% of articles with Policies)
	Society, Social rights, Religion, and culture	76 (2% of articles with Policies)
	National Security	50 (1.3% of articles with Policies)
	Education	38 (1% of articles with Policies)
	Welfare and Family	0
Tone of the article (tone_article_aggregate)	Negative	3,808 (45.2% of About EU articles)
	Balanced/Mixed	2,054 (24.4% of About EU articles)
	Neutral	1,717 (20.4% of About EU articles)

	Positive	842 (10.0% of About EU articles)

2019 online news output

In the full 2019 online news dataset, 8,424 articles were predicted as About EU by the EU_CENTRAL classifier. The tone procedure extracted 50,960 EU-mention sentences from these articles, corresponding to an average of approximately 6.0 EU-mention sentences per About EU article.

Sentence-level tone distribution:

- Neutral: 27,460
- Negative: 18,877
- Positive: 4,623

2.16. Additional information about the Training data

Training corpus: MAPLE Media Dataset 2

The supervised models were trained using MAPLE's Media Dataset 2. This dataset contains 22,616 newspaper articles mentioning the European Union (EU) or EU-related terms. The articles come from 12 newspapers across six EU member states: Belgium, Germany, Greece, Ireland, Portugal, and Spain.

The articles were published during the 30 days preceding 29 legislative elections held in these six countries between 2002 and 2017. The dataset therefore captures EU-related news coverage in national media during pre-election campaign periods.

All articles in the dataset were manually coded by a team of native-speaking coders. In total, 14 coders contributed to the dataset. Coding was carried out using a Windows application developed specifically for the project. The coders applied a shared codebook to identify whether the EU was mentioned, whether the EU was central to the article, and, where applicable, additional article-level variables such as EU dimension, election relevance, journalistic style, conflict or disagreement, policy topic, and evaluations of the EU.

Several procedures were used to ensure coding quality. All coders completed individual training on both the codebook and the coding tool before beginning the main coding process. Coders only started coding articles after demonstrating acceptable levels of inter-coder reliability. When inter-coder reliability was below the acceptable threshold, additional training was conducted.

To reduce potential coder bias, each coder received a fully randomized set of articles in their language, drawn from different years and newspapers. During coding, coders were shown only the title and body text of each article. Metadata such as publication date and newspaper name were omitted from the coding interface.

These manually coded labels were then used as supervised learning targets for fine-tuning the automated classifiers. Because the models were trained to reproduce these human coding decisions, the model **outputs should be interpreted as predicted codebook labels** rather than direct measurements of article content.

Belgium	Election			2003	2007	2010	2014	Totals
	# of articles			*334*	*401*	*500*	*496*	*1731*
Germany	Election		**2002**	**2005**	**2009**	**2013**	**2017**	
	# of articles		*914*	*852*	*858*	*867*	*1022*	*4513*
Greece	Election	**2004**	**2007**	**2009**	**2012**	**2015(i)**	**2015(ii)**	
	# of articles	914	749	987	1928	1305	1171	7054
Ireland	Election			**2002**	**2007**	**2011**	**2016**	
	# of articles			607	584	979	777	2947
Portugal	Election		**2002**	**2005**	**2009**	**2011**	**2015**	
	# of articles		573	676	612	1045	649	3555
Spain	Election		**2004**	**2008**	**2011**	**2015**	**2016**	
	# of articles		494	474	811	444	593	2816
Total number of articles coded:								22616

Training data and sample sizes

The supervised models were trained using MAPLE's Media Dataset 2. This dataset contains 22,616 newspaper articles mentioning the European Union (EU) or EU-related terms. The articles come from 12 newspapers across six EU member states: Belgium, Germany, Greece, Ireland, Portugal, and Spain.

The articles were published during the 30 days preceding 29 legislative elections held in these six countries between 2002 and 2017. The dataset therefore captures EU-related news coverage in national media during pre-election campaign periods.

All articles in the dataset were manually coded by a team of native-speaking coders. In total, 14 coders contributed to the dataset. Coding was carried out using a Windows application developed specifically for the project. The coders applied a shared codebook to identify whether the EU was mentioned, whether the EU was central to the article, and, where applicable, additional article-level variables such as EU dimension, election relevance, journalistic style, conflict or disagreement, policy topic, economic evaluations, and evaluations of the EU.

Several procedures were used to ensure coding quality. All coders completed individual training on both the codebook and the coding tool before beginning the main coding process. Coders only started coding articles after demonstrating acceptable levels of inter-coder reliability. When inter-coder reliability was below the acceptable threshold, additional training was conducted.

To reduce potential coder bias, each coder received a fully randomized set of articles in their language, drawn from different years and newspapers. During coding, coders were shown only the title and body text of each article. Metadata such as publication date and newspaper name were omitted from the coding interface.

These manually coded labels were then used as supervised learning targets for fine-tuning the automated classifiers. Because the models were trained to reproduce these human coding decisions, the model outputs should be interpreted as predicted codebook labels rather than direct measurements of article content.

The usable labelled sample differed across variables because each model was trained only on articles for which the corresponding human-coded label was valid and conceptually applicable. For example, `EU_CENTRAL` used the broadest labelled sample, while `NA_MAINTOPIC` was trained only on articles where the EU dimension concerned either EU policies or domesticated policies. For this reason, the number of training, validation, and test cases differs across models.

Model / variable	*Valid labelled cases*	*Training*	*Validation*	*Test*	*Notes*
EU_CENTRAL	21,714	15,199	3,257	3,258	Binary model distinguishing articles where the EU is only mentioned from articles where the EU is central.
ELECT	11,637	8,145	1,746	1,746	Binary model identifying whether the article is election/campaign related.
NA_NEWSSTYLE binary	10,495	7,346	1,574	1,575	Binary model distinguishing descriptive from interpretive articles; excludes original code `3 = cannot be determined`.
NA_CONF1	11,713	8,199	1,757	1,757	Binary conflict/disagreement model.
NA_EUDIMENS	11,563	8,094	1,734	1,735	Four-class model classifying the EU dimension of the article.
NA_MAINTOPIC	8,563	5,994	1,284	1,285	Fourteen-class policy-topic model; trained and applied only when the EU dimension concerns EU policies or domesticated policies.
NA_ECONCONDI	4,478	3,134	672	672	Five-class economic-condition model; tested but excluded from the final pipeline.
ECON_WORSE binary	4,478	3,134	672	672	Binary "economic situation got worse" model; tested but excluded from the final pipeline.

Label distributions by model

The saved split files contain the following sample sizes:

```
Training examples:     15,199
Validation examples:    3,257
Test examples:          3,258
Total labelled cases:  21,714
```

EU_CENTRAL

```
Training examples:  15,199
Validation examples:       3,257
Test examples:             3,258
Total labelled cases:      21,714
```

Label distribution:

```
Train:
0 = EU mentioned only: 6,946
1 = About EU: 8,253

Validation:
0 = EU mentioned only: 1,488
1 = About EU: 1,769
```

```
Test:
0 = EU mentioned only: 1,489
1 = About EU: 1,769
```

ELECT

```
Training examples:      8,145
Validation examples:    1,746
Test examples:          1,746
Total labelled cases:  11,637
```

Label distribution:

```
Train:
0 = Not election/campaign related: 6,613
1 = Election/campaign related:     1,532

Validation:
0 = Not election/campaign related: 1,417
1 = Election/campaign related:       329

Test:
0 = Not election/campaign related: 1,418
1 = Election/campaign related:       328
```

NA_NEWSSTYLE binary

```
Training examples:      7,346
Validation examples:    1,574
Test examples:          1,575
Total labelled cases:  10,495
```

Label distribution:

```
Train:
0 = Descriptive:  5,098
1 = Interpretive: 2,248

Validation:
0 = Descriptive:  1,093
1 = Interpretive:   481

Test:
0 = Descriptive:  1,093
1 = Interpretive:   482
```

The original NA_NEWSSTYLE variable included a third category, cannot be determined. This category was excluded from the final training data because it was rare and reflected coding uncertainty rather than a substantive news-style category.

NA_CONF1

```
Training examples:      8,199
Validation examples:    1,757
Test examples:          1,757
Total labelled cases:  11,713
```

Label distribution:

```
Train:
0 = No conflict/disagreement:       4,461
1 = Conflict/disagreement present:  3,738
```

```
Validation:
0 = No conflict/disagreement:          956
1 = Conflict/disagreement present:     801

Test:
0 = No conflict/disagreement:          956
1 = Conflict/disagreement present:     801
```

NA_EUDIMENS

```
Training examples:       8,094
Validation examples:     1,734
Test examples:           1,735
Total labelled cases:   11,563
```

Label distribution:

```
Train:
0 = Membership:                    1,357
1 = Constitutional structure:        725
2 = EU policies:                   3,901
3 = Domesticated policies:         2,111

Validation:
0 = Membership:                      291
1 = Constitutional structure:        156
2 = EU policies:                     835
3 = Domesticated policies:           452

Test:
0 = Membership:                      291
1 = Constitutional structure:        155
2 = EU policies:                     836
3 = Domesticated policies:           453
```

NA_MAINTOPIC

```
Training examples:       5,994
Validation examples:     1,284
Test examples:           1,285
Total labelled cases:    8,563
```

Labels distribution:

```
Train:
0  = Economy and Work:                                1,379
1  = Finances and Taxes:                              2,038
2  = Health:                                            148
3  = Migration and Immigration:                         418
4  = National Security:                                 116
5  = Society, Social rights, Religion, and culture:     153
6  = Environment protection:                            232
7  = Transport and Energy:                              166
8  = Law and Order:                                     174
9  = Foreign Policy:                                    531
10 = Institutional design:                              264
11 = Welfare and Family:                                 46
12 = Education:                                          60
13 = Other:                                             269

Validation:
0  = Economy and Work:                                  295
1  = Finances and Taxes:                                437
2  = Health:                                             32
3  = Migration and Immigration:                          89
4  = National Security:                                  25
5  = Society, Social rights, Religion, and culture:      33
6  = Environment protection:                             49
```

```
7  = Transport and Energy:                              36
8  = Law and Order:                                     37
9  = Foreign Policy:                                   114
10 = Institutional design:                              56
11 = Welfare and Family:                                10
12 = Education:                                         13
13 = Other:                                             58

Test:
0  = Economy and Work:                                 296
1  = Finances and Taxes:                               437
2  = Health:                                            32
3  = Migration and Immigration:                         90
4  = National Security:                                 25
5  = Society, Social rights, Religion, and culture:     33
6  = Environment protection:                            50
7  = Transport and Energy:                              35
8  = Law and Order:                                     37
9  = Foreign Policy:                                   114
10 = Institutional design:                              57
11 = Welfare and Family:                                 9
12 = Education:                                         12
13 = Other:                                             58
```

The NA_MAINTOPIC model is the most difficult included classifier because it has 14 classes and several categories have very small numbers of training examples. This explains why its macro F1 is lower than the binary classifiers and why low-confidence predictions are more frequent for this variable.

NA_ECONCONDI

```
Training examples:      3,134
Validation examples:      672
Test examples:            672
Total labelled cases:   4,478
```

Label distribution:

```
Train:
0 = Economic situation mentioned without evaluation:   680
1 = Economic situation got better:                     196
2 = Economic situation got worse:                      745
3 = Economic situation stayed the same:                 43
4 = Economic situation not mentioned:                1,470

Validation:
0 = Economic situation mentioned without evaluation:   146
1 = Economic situation got better:                      42
2 = Economic situation got worse:                      160
3 = Economic situation stayed the same:                  9
4 = Economic situation not mentioned:                  315

Test:
0 = Economic situation mentioned without evaluation:   146
1 = Economic situation got better:                      42
2 = Economic situation got worse:                      160
3 = Economic situation stayed the same:                  9
4 = Economic situation not mentioned:                  315
```

This model was tested but excluded from the final pipeline because the multi-class performance was not sufficient for reliable applied coding.

ECON_WORSE binary

```
Training examples:      3,134
Validation examples:      672
```

```
Test examples:            672
Total labelled cases:   4,478
```

Label distribution:

```
Train:
0 = Economic situation not coded as worse: 2,389
1 = Economic situation got worse:           745

Validation:
0 = Economic situation not coded as worse:   512
1 = Economic situation got worse:            160

Test:
0 = Economic situation not coded as worse:   512
1 = Economic situation got worse:            160
```

This binary version was also tested but excluded because recall for the "worse" class remained too low for reliable applied use.

3. Features and improvements to be soon implemented.

EU politicization models (training feature)

A future development planned for the application is the addition of a training feature that would allow users to fine-tune the existing EU politicization models and potentially train new ones. This would make it possible to incorporate manually reviewed or newly coded datasets into the modelling workflow, thereby improving performance for specific tasks or extending the app to new research areas or communication contexts. In the longer term, such a feature could also support a more iterative workflow in which users review uncertain cases, correct model outputs, and use those corrected examples to improve the models.

Language detection/selection

At present, EU mention detection works only for a limited set of languages. While the current version supports seven languages (English, Portuguese, Spanish, German, Greek, Dutch and French), future versions will aim both to expand this coverage and to improve the detection logic itself. One possible direction is the integration of **automatic language detection**, so that the appropriate language-specific EU dictionary can be applied dynamically. Another option would be to allow users to **select the language of the dataset manually**, which could improve precision, speed (compared to automatic language detection), and reduce false positives in multilingual or mixed-language material.

Adaptative Thresholds

Another feature planned for future versions is to offer some users more control of the application's outputs such as the possibility for users to define their own analytical thresholds. This could include, for example, adjusting the **topic frequency threshold** used for the **diversity** measure in the **Bias Plot** or setting different confidence thresholds for flagging observations that may require manual review. Such flexibility would allow users to tailor the analysis more closely to their own data and research objectives, while also making the app more transparent in terms of how classifications are operationalized.

More available languages

The application is currently available only in **English** and **Portuguese**. Future versions aim to expand the number of supported interface languages in order to make the platform accessible to a broader range of users in EU member-states. Given the comparative and multilingual focus of the project, extending the interface to additional European languages would be an important step toward wider usability and dissemination.