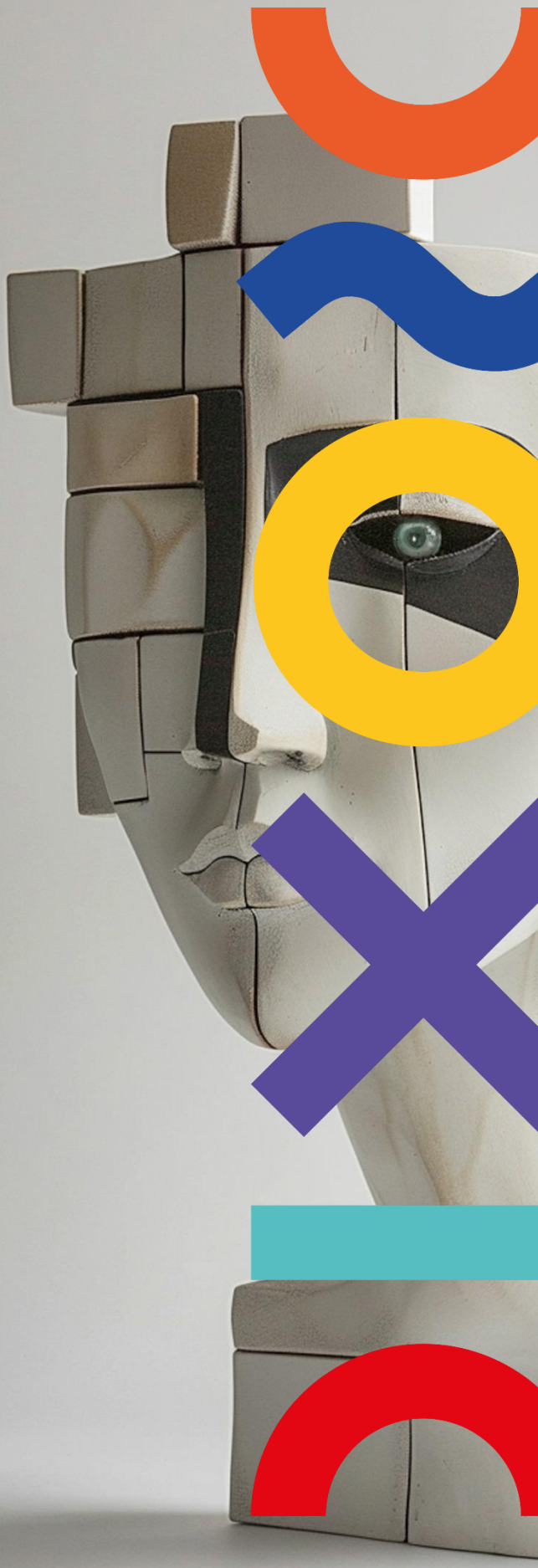
ENCODE

D3.3

Sentiment analysis



Funded by
the European Union

This project has received funding from the European Union under
the Horizon Europe Research & Innovation Programme
(Grant Agreement no. 101132698 ENCODE).

D3.3 Sentiment analysis

Dissemination Level: PU -Public
Lead Partner: PBY
Due date: 30/06/2025
Actual submission date: 19/09/2025
Actual submission date after review meeting: 9/01/2026

PUBLISHED IN THE FRAMEWORK OF

ENCODE - Unveiling emotional dimensions of politics to foster European democracy consumers

AUTHORS

Rodrigo Ortega Izquierdo, *PBY*
Frans Folkvord , *PBY*
Jim Ingebretsen Carlson, *PBY*

REVISION AND HISTORY CHART

VERSION	DATE	EDITORS	COMMENT
0.1	03/09/2025	PBY	First version submitted to review
0.2	13/09/2025	UWR	Reviewed
0.3	15/09/2025	PBY	Comments addressed and sent for final review
1.0	19/09/2025	ASM	Submission to Participant Portal
1.1	22/10/2025	PBY	Revisions after the review meeting
1.2	11/12/2025	ASM	Review and final comments
1.3	19/12/2025	PBY	Final version
2.0	9/01/2026	ASM	Submission to Participant Portal

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

DISCLAIMER

The information in this document is subject to change without notice. Company or product names mentioned in this document may be trademarks or registered trademarks of their respective companies.

All rights reserved

The document is proprietary of the ENCODE consortium members. No copying or distributing, in any form or by any means, is allowed without the prior written agreement of the owner of the property rights.

This document reflects only the authors' view. The European Community is not liable for any use that may be made of the information contained herein. Responsibility for the information and views expressed in the therein lies entirely with the author(s).

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

TABLE OF CONTENT

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

LIST OF FIGURES:

LIST OF TABLES:

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

EXECUTIVE SUMMARY

This deliverable presents the methodology, implementation, and findings of the ENCODE project's sentiment analysis, which aims to decode the emotional dimensions of political discourse and their implications for European democracy. The analysis focuses on understanding how emotions and values shape political communication across social media, particularly during electoral periods, and provides evidence-based insights to foster constructive engagement and democratic resilience.

The study employed a multilingual, multi-stage approach combining manual annotation and Large Language Models (LLMs). Data were collected from X (formerly Twitter) across six European countries (Austria, Bosnia and Herzegovina, Bulgaria, Denmark, North Macedonia, and Poland) covering three user categories: general public, politicians, and media outlets. The final dataset comprised over 2.1 million tweets and comments spanning 2022–2024, including periods around national and EU elections. Manual coding of a stratified sample ensured high-quality ground truth for model training, with inter-coder reliability validated through Krippendorff's alpha thresholds. Transformer-based models (XLM-RoBERTa) were fine-tuned for three core tasks: political content detection, emotion recognition (fear/anxiety, happiness/enthusiasm, anger, hate, neutrality), and human values identification (benevolence, security, universalism, tradition, self-direction, power). Evaluation metrics demonstrated robust multilingual performance, enabling large-scale automated annotation. Statistical analyses included chi-square tests for emotion-value associations, regression models for user category and electoral period effects, and engagement modelling to assess the interplay between emotional tone and audience reactions.

Key findings reveal systematic differences in emotional expression across actors and contexts. Politicians and media exhibit distinct emotional profiles compared to the general public, while electoral periods amplify certain emotions: anger and enthusiasm rise during national elections, whereas EU elections show slight increases in fear and reduced happiness. Hate speech does not escalate during elections, suggesting either moderation effects or its independence from electoral dynamics. Engagement analysis confirms that emotionally charged content drives higher interaction, particularly when aligned with salient values such as security or power. These insights inform ENCODE's broader objectives: developing best practices to counter disinformation, mapping emotional gaps, and supporting the co-creation of positive narratives for democratic discourse. The findings will feed into subsequent deliverables, including the Catalogue of Best Practices (D3.4), and underpin policy recommendations aimed at strengthening trust, inclusivity, and resilience in European democracies.

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

1. INTRODUCTION

1.1 THE ENCODE PROJECT

The ENCODE project, titled "Unveiling Emotional Dimensions of Politics to Foster European Democracy," aims to explore and decode the role of emotions in political discourse and their impact on democratic processes. Recognising that emotional appeals have significantly influenced political movements and voter behaviour, ENCODE seeks to understand the interplay between emotions, values, and identities. The project's primary goal is to create new positive narratives that can foster trust and engagement in European democratic processes, thereby counteracting the negative emotions that often dominate political discussions. Through innovative methodologies, including social media sentiment analysis, biometric research, and surveys, ENCODE aims to provide policymakers with tools and strategies to better incorporate the emotional needs of citizens into governance, ultimately enhancing democratic resilience and fostering a more inclusive political environment.

1.2 OBJECTIVES OF DELIVERABLE

The objective of this deliverable is to present the methodology, implementation, and results of the emotional analysis conducted within the ENCODE project, in full alignment with the project's overarching aim to decode the emotional dimensions of politics and foster European democracy. This deliverable seeks to develop and apply advanced, multilingual large language models (LLMs) for political content detection, emotion recognition, and human values identification in social media discourse. It aims to provide a comprehensive analysis of emotional dynamics in political communication across different user categories, including the general public, politicians, and media, as well as across electoral periods. The deliverable is designed to generate actionable insights that inform strategies to foster constructive political dialogue, counteract negative emotions in public discourse, and strengthen democratic engagement and resilience in Europe. Furthermore, it supports the project's key exploitable results, such as the development of best practices for tackling disinformation, the creation of emotional gap maps, and the validation of survey instruments and democratic resilience heatmaps.

1.3 STRUCTURE OF THE DOCUMENT

This document is structured to provide a logical and transparent flow from context to methodology, results, and implications, facilitating both scientific rigour and practical usability:

Section 1 - Introduction. Presents the ENCODE project context, rationale, and the specific objectives of this deliverable.

Section 2 – Methodology. Details the multilingual data collection, annotation processes, and the machine learning models used for classification tasks.

Section 3 - Data collection and annotation results. Summarises the dataset composition, including country and category breakdowns, and provides descriptive statistics to contextualise the analysis.

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Section 4 – Analysis methodology. Details the analysis methods used to respond to the research questions posed in Deliverable 3.2.

Section 5 – Results. Presents the main findings, including statistical analyses, regression models, and engagement patterns related to emotions and values in political discourse.

Section 6 – Conclusions. Highlights key findings, their implications for democratic discourse, and recommendations for future research and policy.

Finally, the annexes include codebooks, model performance metrics and detailed statistical tables.

1.4 RELATION TO OTHER TASKS

This deliverable is closely interlinked with other tasks and work packages within the ENCODE project. It is a direct output of WP 3, "Analysing Social Media Communication," and builds upon the methodological frameworks and state-of-the-art reviews developed in previous deliverables.

The key interconnections between WP3 and other work packages are as follows:

- WP2 - The conceptual frameworks and theoretical models of affective pluralisation and emotional politics developed in WP2 provide the foundation for the analytical approaches used in this deliverable, as described in D3.2.
- WP3 - The analyses, methodologies, and results presented in this deliverable will serve as the primary basis for the preparation and writing of D3.4, the "Catalogue of Best Practices," ensuring that the recommendations and guidelines developed for tackling disinformation and promoting positive emotional narratives are firmly grounded in the empirical evidence and analytical work carried out here.WP4
- WP4 - The annotated datasets and analytical outputs generated here feed into WP4, which focuses on biometric and qualitative research. WP3 data will serve to generate synthetic posts used as experimental stimuli in biometric research to explore emotional and cognitive responses to different communication patterns.
- WP5 - which is dedicated to experimental validation and survey development will complement its survey design with the findings from the social media engagement, and subsets of the data from WP3 could be used to explore specific topics addressed in the survey research. This enables cross-method triangulation and the development of validated instruments for measuring democratic resilience.
- WP6 - Insights from WP3's analysis of emotional dynamics guide WP6 in the co-creation of positive emotional narratives and counter-disinformation strategies that align with observed audience behaviours.
- WP7 - The evidence and analytical results produced in WP3 support WP7's foresight and policy workshops, providing empirically grounded recommendations for strengthening democratic resilience and inclusive communication practices.

In summary, this deliverable acts as a bridge between the project's conceptual, methodological, and empirical components, ensuring coherence and integration across tasks and work packages, and directly supporting the achievement of ENCODE's scientific and societal objectives.

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

2. METHODOLOGY

This deliverable presents a multilingual approach to analyse political behaviour on social media through automated classification of posts across three key dimensions: (1) political content detection, (2) emotion recognition, and (3) human values identification. The methodology employs transformer-based machine learning models to annotate social media content from seven languages (Polish, German, Danish, Bulgarian, Serbian, Macedonian, Albanian) across multiple European contexts.

The complete analytical pipeline, encompassing data extraction, preprocessing, and model training, was implemented in Python 3.12.4. Core functionality was provided by established libraries, including requests for API interactions, pandas for data manipulation, and NumPy for numerical computations (Harris et al., 2020; McKinney et al., 2010; Reitz, 2022; Van Rossum & Python Software Foundation, n.d.). These were complemented by custom-developed functions designed to address the specific methodological requirements of this multilingual political content analysis.

2.1 DATA COLLECTION

For this study, data was collected from X (formerly Twitter), focusing on six countries (Austria, Bosnia and Herzegovina, Bulgaria, Denmark, North Macedonia, Poland) over the period from 2022 to 2024. The intent was to analyse political content across three distinct user categories: the general public, politicians, and media outlets.

The identification and selection of politicians and media outlets were conducted through a structured process designed to ensure both national representativeness and methodological rigour. This process was grounded in the collaborative engagement of local partners, within each participating country, leveraging their contextual expertise to curate relevant actors in the political and media landscapes.

For the political actor sample, local partners were tasked with compiling comprehensive lists of both political parties and individual politicians. The inclusion criteria for political parties stipulated that only those entities which had secured more than 5% of the vote in the most recent national parliamentary elections were eligible. This threshold was established to guarantee the inclusion of parties with demonstrable public support and substantive influence on national discourse. Within each qualifying party, partners identified between 10 and 20 representative political figures who maintained active accounts on X. These individuals typically encompassed party leaders, prominent parliamentarians, and other figures of significant public visibility. The selection process entailed verification of each account's authenticity and activity, thereby ensuring the relevance and reliability of the data corpus. In exceptional cases, such as in countries with highly fragmented or pluralistic political systems, parties below the 5% threshold could be included if local experts deemed their participation essential for capturing the full spectrum of political pluralism.

The sampling of media outlets followed a parallel logic. Local partners were instructed to identify the 10 to 20 most influential and widely recognised media organisations within their respective national contexts. This selection was intentionally inclusive of both mainstream (trustworthy) and non-mainstream (potentially untrustworthy or disinformation-prone) outlets, reflecting the project's aim to capture the diversity of information sources shaping public opinion. Only outlets with active X accounts were considered, and the lists were validated using local knowledge, fact-checking resources, and, where available, academic literature on media influence and credibility.

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

The “general public” category is defined as comprising all social media users, age 18 and above, with public X accounts, not included in the politicians or media classifications, thereby representing the organic dimension of political discourse on X.

Data gathering was conducted using XAPI v2 (X Corp, 2024). The general filtering methodology employed was designed to ensure that the collected social media data accurately reflected the relevant political discourse within each participating country, while generating a comprehensive sample. Following XAPI v2 capabilities we have different filters that can be applied during extraction:

- Country Filtering: This is achieved by utilising the country metadata provided by X, however, it must be noted, this information is not massively included by users, therefore geolocation filtering widely reduces the potential sample.
- Language Filtering: X posts metadata also includes a language tag that is automatically generated by X, therefore all posts include it, however not all languages are included din their classification, excluding for example, “Macedonia”.
  Timeframe Filtering: X post metadata includes also a timestamp that can be used to identify the publication date and time.
  Keyword Filtering: XAPI enables to extract post based on keyword search. To ensure relevance, keyword sets targeting politically related topics were curated for each language, translated as needed, and adapted to country-specific contexts when appropriate. Queries were constructed by joining these keywords with "OR" operators, coupled with filters for language and country when metadata allowed. Because X API imposes character limits on queries, these keyword queries were subdivided into smaller segments to comprehensively cover the full set of keywords without losing any data coverage. Duplicate posts were deleted after extraction.
- Account Filtering: XAPI also allows for the extraction of post from specific accounts.

Posts were extracted for the 2022- 2024 period including periods around national and EU elections. For each country and language, further filtering was applied to extract post relevant to the study context:

- posts from the general public were filtered by language, country, and the comprehensive set of politically related keywords provided by the partners in each country;
- the media outlets were defined through specific account lists provided by project partners, filtered by language and the political keywords provided by partners in order to exclude non-political news posted by these outlets;
- politicians' posts were extracted following the predefined accounts list and filtering by language.

Country filtering is only used for general public post extraction, as media outlets and politicians have already been sorted by partners, therefore their contextual relevant for the study country is established.

Additionally, due to data extraction limits, comments on politicians’ posts were sampled only for a focused period, one month before and one month after each national (Table 1) and EU election (09/06/2024) during the studied period, and during two months of a non-electoral period (09/05 – 09/07/2023).

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Table 1 National election dates

Country	National election
Austria	29/09/2024
Bulgaria	27/10/2024
Poland	15/11/2023
Denmark	01/11/2022
North Macedonia	08/05/2024
Bosnia and Herzegovina	02/10/2022

Due to the limitations of the metadata language classification, no general public posts could be extracted from North Macedonia and Bosnia and Herzegovina, as their language were not properly tagged by X, and country metadata was available for only a very small sample of posts, insufficient for the intended research.

The extracted data included the following fields to support detailed analysis: id, created_at (timestamp), author_id, username, language, text content, like_count, reply_count, repost_count, quote_count, country and country_code (when available), is_reply (flag indicating whether the post is a comment), and is_long_post (flag indicating extended posts). The long post flag was also relevant as now X allows posts longer than 280 characters; for technical reasons, the full text is displayed in an additional field. When this happened during extraction, the full text was recorded in the text field for practical reasons in the analysis. These fields ensure a comprehensive picture of the content, author, engagement metrics, and geographic context.

All extracted data were saved as flat CSV files for compatibility and ease of analysis. This methodological approach enabled an in-depth, structured assembly of politically related discourse on social media at multiple levels (general public, political, and media) across the six study countries and over a multi-year period encompassing both national, EU elections and non-electoral periods, providing valuable insights into the dynamic conversations surrounding gender in these regions.

2.2 MANUAL ANNOTATION

As a subsequent stage, manual annotation was conducted specifically for posts identified as potentially political in nature. This procedure involved several intertwined objectives. Initially, human coders systematically revised each post to verify whether it was genuinely political. Posts confirmed as pertaining to politics were then further examined to identify the values and emotions they expressed.

For each positively identified political post, coders were instructed to assign binary codes (1/0) indicating if the post was about politics or not. Additionally, for each political post, annotators evaluated and recorded the salient emotional tone or value orientation as expressed in the text. This was operationalised as a categorical variable, denoting emotions present in the post (fear/anxiety, happiness/enthusiasm, anger, hate and emotional neutrality) and values (benevolence, security, universalism, tradition, self-direction and power), according to a predefined coding scheme. The development of this coding scheme is documented in Annex A, which provides the operational definitions of each category and the rationale for the final category choices and coding decisions. For each of these emotions

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

or values, coders were asked to note 0 or 1, 1 indicating the presence of such emotion or values in the text of the post. The full codebook, covering development details, definitions, and rationale, is provided in Annex A.

Particularly for comments, political classification required a more nuanced analytical approach that performed classification based on both the original post and the comment itself, recognising that comments cannot be fully understood in isolation from their parent content. This dual-context methodology was essential because contextualisation was needed to fully understand the scope, meaning and intention of the comments; a comment that might appear politically neutral when viewed alone could carry significant political implications when considered alongside the original post it responds to. To ensure focused and relevant analysis, only comments responding to posts that had already been tagged as political were provided to the coders for annotation of values and emotions, streamlining the coding workload while maintaining consistency in the dataset. This approach acknowledged that political engagement on social media platforms often occurs through layered interactions, where the full political meaning emerges from the relationship between original content and user responses, enabling coders to capture more accurate assessments of political values and emotional expressions that might otherwise be misclassified when analysed in isolation.

Manual coding was conducted across all relevant datasets, with a stratified sample including the different types of users addressed. For each country and language, a sample of 2000 posts were coded. This sample was distributed to have a representative distribution of the different types of content subject to this study. The distributed was as follows: 1020 posts from which 340 were general public pots, 340 were politicians' posts and 340 were media posts, the remaining 980 were comments. For the countries where general public posts could not be extracted the distribution was 1000 posts, from with 500 were political post and 500 media post and 1000 comments. In both cases comments were selected from the annotated politicians' post.

For reliability and consistency, dual annotation was performed on 20% of the sample by two independent coders. A minimum Krippendorff's alpha of 0.67 (2/3) for each coded category was required as a threshold before proceeding to code the entire dataset. This ensured an acceptable level of inter-coder agreement, thereby reinforcing the validity of the manual labels as ground truth for subsequent predictive modelling tasks.

2.3LLM FINE-TUNNING

2.3.1 DATA PREPARATION

The preprocessing pipeline involved several key steps. First, standardisation of column names across datasets, particularly renaming 'text_reply' columns to 'text' for comments. Second, removal of rows with missing text content using dropna operations. Third, filtering to include only binary annotations (0/1) for classification tasks. Finally, language filtering to focus on the seven target languages, excluding any miscoded entries.

For the political content classification task, the final dataset contained 8,030 entries after preprocessing. For emotion classification, 11,884 entries were retained, and for values classification, 11,909 entries were used after removing rows with missing annotations in the respective categories.

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

2.3.2 MODEL ARCHITECTURE AND CONFIGURATION

The study employed XLM-RoBERTa-base (Conneau et al., 2020) as the foundation model for all classification tasks. XLM-RoBERTa was selected for its demonstrated effectiveness in multilingual natural language processing tasks and its ability to handle the diverse linguistic contexts present in the dataset. This model, developed by Facebook AI Research, has shown superior performance across multiple languages without requiring language-specific fine-tuning, making it particularly well-suited for cross-lingual analysis spanning the six countries in this study (Conneau et al., 2020). The model's robust architecture and pre-training on large-scale multilingual corpora enabled consistent classification performance across the different languages represented in the dataset:

- Political Content Classification: Configured as a binary classification task with two labels: "non_political" (0) and "political" (1). The model architecture included a sequence classification head with 2 output classes.
- Emotion Classification: The model was configured as a multi-label classification problem and 5 output labels corresponding to the five emotional dimensions.
- Values Classification: Similarly structured as a multilabel classification task with six human values categories. The model utilised the same multilabel configuration with 6 output labels for the six value dimensions.

2.3.3. TRAINING CONFIGURATION AND HYPERPARAMETERS

All models employed the XLM-RoBERTa tokenizer (Conneau et al., 2020) with consistent parameters: maximum sequence length of 512 tokens for political classification and 256 tokens for emotion and values classification, padding to maximum length, and truncation for longer sequences.
The sample was also split and stratified into a training dataset and test data, accounting for balance between tags and languages.

- Political Classification: Employed stratified splitting by combining language and political labels to ensure balanced representation across linguistic groups. The data was split 80/20 for training and testing with stratification by language and politics tag keys.
- Multilabel Tasks: Used MultilabelStratifiedKFold (Sechidis et al., 2011) for complex stratification that accounts for both language diversity and label distribution across multiple target variables. For each language group, a 5-fold stratified split was performed, with the first fold used for train/test division.

The training configuration arguments follow the learning rate 1e-5, batch size 32, 3epochs,
The best model selection was based on F1-score for political classification and F1-macro for multilabel tasks.

2.3.4 EVALUATION METRICS

To assess model performance, we employed distinct evaluation protocols tailored to each classification task, ensuring both comprehensive and nuanced analysis.

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

For **political content classification**, we measured accuracy, precision and recall for the positive class, positive-class F1 score and the area under the receiver operating characteristic curve (AUC-ROC). For **emotion** and **human values** classification, both formulated as multilabel tasks, we reported micro- and macro-averaged precision, recall, accuracy, and F1 scores, alongside per-label F1 values and macro-averaged AUC-ROC.

To evaluate cross-linguistic robustness, we computed all aforementioned metrics separately for each of the target languages. This stratified evaluation enables identification of language-dependent performance disparities and informs potential calibration or data augmentation strategies.

Paired bootstrap resampling (1,000 iterations) was employed to derive 95% confidence intervals for primary metrics (F1 and AUC-ROC), enabling statistical comparison between models and across languages.

2.3.5 IMPLEMENTATION

The implementation utilised the HuggingFace Transformers library (Wolf et al., 2020) for model training and evaluation. Custom dataset preparation functions converted pandas DataFrames to HuggingFace Dataset objects with appropriate tokenization. The Trainer API was employed with custom metric computation functions and early stopping callbacks to prevent overfitting. Trained models and tokenizers were saved using HuggingFace's standard persistence methods. Comprehensive evaluation results were exported containing general metrics, language-specific performance, per-label metrics, and summary statistics that can be checked in the Annex B.

2.3.6 MODEL CARDS

Model cards for all trained large language models are provided as complementary materials in the form of Markdown files, offering a more standardised and widely adopted format for documenting AI models than the narrative description in this chapter. While the chapter explains the models' architecture, training data, and evaluation in detail, the accompanying model cards present the same information in a structured template specifically designed for LLMs, including sections on intended use, limitations, ethical considerations, and performance metrics. This dual approach ensures both comprehensive methodological transparency within the text and quick, standardised reference documentation that supports reuse, comparison, and integration of the models in other work packages. Model cards can be found in Annex F.

2.4 AUTOMATIC ANNOTATION

The process begins with loading and preprocessing the data. Once loaded, these individual data frames are concatenated to form a comprehensive dataset. To maintain data integrity and avoid annotation biases, duplicate messages, identified via unique message IDs, are removed prior to annotation.

Following data preparation, the annotation is performed using the transformer-based deep learning models previously fine-tuned for each specific task. Each model is locally stored and loaded for the annotation process. The corresponding tokenizer is also instantiated to ensure

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

that the input text is appropriately tokenized for the model. Importantly, the model is switched to evaluation mode, which guarantees that predictions are generated without altering the model parameters, maintaining the model's stability during inference. Models are implemented sequentially, first the political classification model was implemented and then only those posts classified as political were further categorised using the emotions and the values multi-label classification models.

The core of the annotation methodology employs the Hugging Face pipeline abstraction configured for text classification. This pipeline is set up with key parameters such as truncation, padding, and a maximum token length of 512 to effectively manage input sequences of various sizes. Computation resources are optimised by utilising a GPU when available, thereby accelerating the annotation process on large datasets.

Each message within the dataset is sequentially fed into the classification pipeline. The model outputs predicted labels along with confidence scores for each text entry. These predicted labels are then appended to the dataset, thereby enabling fully automated annotation. This automated process is consistent and reproducible, eliminating the need for manual labelling and supporting scalability in handling diverse and large-scale text corpora.

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

3. DATA COLLECTION AND ANNOTATION RESULTS

Through systematic implementation of the data collection methodology described above, we successfully assembled a comprehensive multilingual corpus comprising 2,169,852 posts and comments from X spanning the period 2022–2024. This substantial dataset was obtained through targeted extraction processes that leveraged API queries for data extraction, both for general public content and from the curated lists of politicians and media outlets provided by local partners across the study regions. The resulting corpus covers seven linguistic and geographical contexts: Austria (AT), Bosnia and Herzegovina (BA), Bulgaria (BG), Denmark (DK), North Macedonia (MK), the Albanian-language MK subset (MK_AL), and Poland (PL), distributed across four distinct source categories: comments, media outlets, politicians, and general public posts. The detailed distribution per country and category of extracted and annotated data (after deleting duplicates) can be consulted in Table 2. Note. It should be noted that due to X metadata limitations regarding language and country tagging, general-public posts could not be reliably extracted for North Macedonia and Bosnia and Herzegovina, though politician and media content from these regions was successfully obtained through the partner-provided account lists.

Country	Comments	Media	Politicians	General public	Total
AT	290014	45049	40318	14261	389642
BA	266010	76600	35588	0	378198
BG	37159	13471	10920	5910	67460
DK	232258	16187	32886	17935	299266
MK	58069	40864	12067	0	111000
MK_AL	0	8969	0	0	8969
PL	572651	96645	215874	30147	915317
Total	1456161	297785	347653	68253	2169852

Table 2 Total number of posts and comments extracted and annotated per country and category

To complement the overall corpus statistics, we also include the distributions, both absolute counts and normalised percentages, in the Annex Table 11 and Table 12. These tables provide granular detail at the intersection of country and category, enabling cross-national and cross-source comparisons that inform subsequent inferential analyses.

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

4.ANALYSIS METHODOLOGY

4.1 EXPLORATORY HYPOTHESES, THEORETICAL ANCHORING AND OPERATIONALISATION

The analytical work presented in this deliverable was mainly conceived as exploratory, aiming to uncover emotional and value-related structures within online political discourse across European contexts. Rather than testing narrowly specified hypotheses, the analyses were guided by three broad, theoretically informed research questions formulated in *D 3.2*. These questions served as an analytical compass, ensuring conceptual coherence with the broader ENCODE framework while allowing data-driven discovery across large-scale multilingual social media corpora. The hypotheses were rooted in the project's theoretical foundations in affective pluralisation and emotional politics, which highlight how emotional communication both reflects and shapes political identity, value expression, and democratic participation. In addition, two major exogenous events-the COVID-19 pandemic and the war in Ukraine-provided a shared emotional backdrop for European public discourse. Both crises profoundly reconfigured the affective and moral registers of online communication, amplifying themes of fear, security, solidarity, and resilience. Within this context, the hypotheses aimed to capture how such large-scale disruptions influenced emotional expression, value orientation, and user engagement in digital democratic spaces.

The table below summarises the three exploratory hypotheses, their theoretical underpinnings, contextual relevance, and operational implementation within the analyses of D3.3.

Hypothesis / Research Question	Theoretical Grounding (from WP2 / D3.2)	Contextual Relevance (Exogenous Events)	Operationalisation in D3.3
H1. What role do emotions play in online discourse in relation to the values and identity of the user?	Based on theories of affective pluralisation and value–identity integration, which posit that political actors and citizens express emotions that align with underlying value commitments and self-concepts.	The war in Ukraine and COVID-19 intensified uncertainty and collective identity negotiation, foregrounding values of security, solidarity, and autonomy.	Analysed through emotion–value correlations (Section 5.1) and OLS regressions by user category (Section 5.2), comparing emotional expression among politicians, media, and the general public to uncover role-specific emotional regimes.
H2. What emotions and values are conveyed in untrustworthy news compared to other news in social media?	Draws on theories of disinformation and affective framing, suggesting that untrustworthy media employ heightened emotional and moral cues to attract attention and reinforce in-group identity.	COVID-19 misinformation and Ukraine-related propaganda provided paradigmatic cases of emotionally charged and value-laden narratives circulating in the digital sphere.	Operationalised through the media clustering and engagement analyses (Sections 4.3 & 5.4), contrasting emotional and value profiles between "emotional" and "neutral" media content, revealing which value frames dominate in affect-driven communication.

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Hypothesis / Research Question	Theoretical Grounding (from WP2 / D3.2)	Contextual Relevance (Exogenous Events)	Operationalisation in D3.3
H3. Which affects and emotions are most triggering in terms of generating responses and reactions in social media?	Anchored in emotional contagion and digital mobilisation theory, emphasising that specific emotions (e.g., enthusiasm, anger) drive online engagement and political participation.	During and after COVID-19 and the Ukraine war, emotional mobilisation became a defining feature of digital interaction, with both positive and negative affects shaping participation.	Tested through OLS regressions of engagement on emotion × category and emotion × period interactions (Sections 4.4 & 5.5), quantifying how different emotions, positive or negative, affect engagement dynamics across actors and electoral contexts.

Taken together, these exploratory hypotheses provided the conceptual structure for analysing how emotions and values intertwine in digital democratic discourse. By situating emotional expression within the broader social and political disruptions of recent years, the analyses in this deliverable illuminate how affective dynamics underpin value communication, identity formation, and engagement online.

4.2 STATISTICAL ANALYSIS FRAMEWORK FOR EMOTION VALUE CORRELATION

Each emotion-value combination was tested for statistical independence using the chi-square statistic. Chi-square tests of independence were conducted to examine associations between emotional categories and Schwartz's basic human values. The analysis employed a binary coding approach for both emotional states and value orientations. All statistical computations were performed using Python 3.12.4 with the statsmodels (Seabold & Perktold, 2010) and pandas (McKinney et al., 2010) libraries. Additionally, scikit-learn (Pedregosa et al., 2011) was employed for machine learning tasks, while seaborn (Waskom, 2021) and matplotlib (Hunter, 2007) were used for data visualisation.

Due to the multiple statistical tests performed across emotion-value pairs, a Bonferroni correction was applied to control for Type I error inflation. This conservative approach adjusts the alpha level by dividing the conventional significance threshold ($\alpha = 0.05$) by the total number of comparisons, ensuring robust statistical inference across the full matrix of associations.

The phi coefficient (φ) was calculated as a measure of association strength for each significant emotion-value pair. The phi coefficient represents the correlation between two binary variables and provides standardised effect size estimates ranging from 0 (no association) to 1 (perfect association). Effect sizes were interpreted using conventional guidelines: $\varphi = 0.1$ (small effect), $\varphi = 0.3$ (medium effect), $\varphi = 0.5$ (large effect).

4.3 USER CATEGORY AND ELECTION PERIOD EMOTIONAL ANALYSIS

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

We investigated systematic differences in emotional expression across user categories (media outlets and politicians) relative to a baseline of general users, by applying regression analysis. For each emotion, we estimated an ordinary least squares (OLS) regression model where the emotion score served as the outcome and text category (with the general public as the reference level) served as the predictor. From these models, we extracted coefficient estimates, confidence intervals, and significance tests for each non-baseline category. Finally, we visualised the estimated effects and their uncertainty to assess how emotional intensity varies by category relative to posts. This approach provided a coherent framework for quantifying and comparing emotion differences across textual sources.

The same analysis is performed for the different electoral periods, taking as baseline the non-electoral period defined before we compare the regression coefficients for both national and EU elections to discern if the only emotional discourse changes tone in the dates in the window of 1 month before and one month after the electoral date.

4.4 MEDIA CLUSTERING

Research indicates that misinformation (“fake news”) systematically deploys more affect-laden language and elicits stronger emotional responses than verified reporting. Comparative text analyses show that fake news uses more subjective, emotional, and clickbait-style wording than real news (Horne & Adalı, 2017), while large-scale diffusion studies find that false stories provoke greater arousal (e.g., fear, disgust, surprise) and, consequently, spread farther and faster on social platforms (Vosoughi, Roy, & Aral, 2018). Building on this evidence, we assess whether more emotionally charged content attracts higher levels of user reaction (likes, replies, reposts, quotes) and whether engagement is patterned by specific value frames (e.g., security, power, universalism).

The analysis of media outlets builds on this, by using a binary clustering approach to categorise media-related posts based on the presence of emotions. The dataset was filtered to include only posts classified under the "media" category. Clusters were defined using the emotional_neutrality binary variable: Cluster 1 ("Neutral") comprised posts with emotional_neutrality = 1, while Cluster 2 ("Emotional") included all remaining posts (emotional_neutrality = 0).

This binary classification was grounded in the distinction between emotionally neutral content and emotionally activated content, providing clear interpretability for media strategy applications while ensuring adequate sample sizes for robust statistical analysis.

Dependent variables included four engagement metrics: like count, reply count, repost count, and quote count, representing different forms of user interaction behaviour. Independent variables encompassed the six values defined for this study.

Cluster differences were examined using means, standard deviations, and medians for all variables. Independent samples t-tests were conducted to assess significant differences between clusters.

4.5 ENGAGEMENT ANALYSIS

The analysis focuses on social media engagement patterns. The dependent variable, engagement, was constructed as the sum of four interaction metrics: like count, reply count, repost count, and quote count. To address the highly skewed distribution typical of social

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

media engagement data, a logarithmic transformation is applied using the natural logarithm of engagement plus one, denoted as log(engagement + 1), which serves as our outcome variable throughout the analysis.

An Ordinary Least Squares (OLS) regression framework is employed to examine the interactive effects of user category and comments or period and emotional tone on social media engagement. The model specification follows a factorial design structure:

$$\log(\text{engagement} + 1) = \alpha + \beta_1(\text{category/period}) + \beta_2(\text{emotion}) + \beta_3(\text{category} \times \text{emotion}) + \varepsilon$$

where category/period and emotion variables are represented through dummy coding, and the interaction terms capture the differential effects of emotional content across different content categories.

Emotional neutrality was established as the baseline for emotion dimensions, using general public posts for the category dimension, and the non-election window for electoral period comparison. This approach was chosen for its theoretical interpretability, as it represents the most neutral possible combination, providing a clear reference point against which all other combinations can be evaluated.

The dummy variable construction omitted the emotional neutrality and general public posts or emotional neutrality and non-election period, ensuring that all estimated coefficients represent deviations from this baseline condition.

Interaction effects between categories and emotions, as well as period and emotions, were operationalised through the multiplication of respective dummy variables. For each non-baseline category "i" and non-baseline emotion "j", we constructed interaction terms as:

$$Interaction\_\{ij\} = Category/Period\_i \times Emotion\_j$$

This generates a comprehensive set of interaction terms that capture how the effect of emotional content varies across different emotions used by particular actors or at particular periods, relative to the baseline condition. Model estimation was conducted using ordinary least squares with heteroscedasticity-robust standard errors.

Given the log-linear specification of our dependent variable, coefficient interpretation follows semi-elasticity principles. For any coefficient β, the percentage change ($\%\Delta$) in engagement is calculated as:

$$\%\Delta = 100 \times (exp(\beta) - 1)$$

This transformation allows for intuitive interpretation of results in terms of percentage changes in engagement levels relative to the baseline condition, facilitating practical understanding of effect magnitudes.

For interaction effects, we computed combined effects by summing the relevant main effects and interaction coefficients before applying the exponential transformation. This approach provides the total effect of each combination relative to the baseline, accounting for both individual main effects and their interactive components.

The analytical approach distinguishes between two types of effects:

- Main category/emotion/period effects: the impact of category/emotion/period when the baseline category for the other variables applies.

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

- Interaction effects: the additional impact arising from specific combinations beyond what would be predicted by additive main effects.

This decomposition enables identification of synergistic relationships where certain pairings produce engagement effects that exceed the sum of their individual components, as well as antagonistic relationships where such combinations underperform relative to additive expectations.

The resulting coefficient matrix provides a comprehensive mapping of engagement effects across all combinations, with each cell representing the percentage change in engagement relative to the baseline condition.

4.6 COUNTRY COMPARISON

The multinational structure of the ENCODE corpus enables a systematic comparison of emotional and value-laden political communication across six European contexts, Austria (AT), Bosnia and Herzegovina (BA), Bulgaria (BG), Denmark (DK), North Macedonia (MK, including MK-AL of posts in Albanian language), and Poland (PL). Task 2.3 country profiles are considered in this analysis as the source of country-specific information for the comparisons. Although the overall architecture of online political communication exhibits shared features (e.g., anger-dominated comments, emotionally restrained media actors), the *intensity*, *direction*, and *value alignment* of these emotional expressions differ substantially from one country to another. This section synthesises those differences, drawing directly on **Annex C Table 11 and Table 12** (cross-national emotion/value distributions), the chi-square heatmaps for emotion-value associations (**Figure 1 and Figure 2**), and the user-category emotion differences (**Figure 3**).

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

5. RESULTS

5.1 CHI-SQUARE ANALYSIS OF EMOTION-VALUE ASSOCIATIONS

The logistic regression analysis attempting to predict emotional categories based on value predictors encountered convergence failure, preventing the estimation of reliable model parameters. This analytical limitation necessitated reliance on the chi-square independence tests as the primary inferential framework for examining emotion-value associations. The comprehensive chi-square analysis revealed systematic patterns of association between emotions and values in political discourse. Out of 30 tested combinations (Figure 1), 27 associations achieved statistical significance even after stringent Bonferroni correction (adjusted α = 0.00167), demonstrating robust relationships between emotional expressions and value orientations. Detailed results can be consulted in Annex ANNEX D – Analysis results .

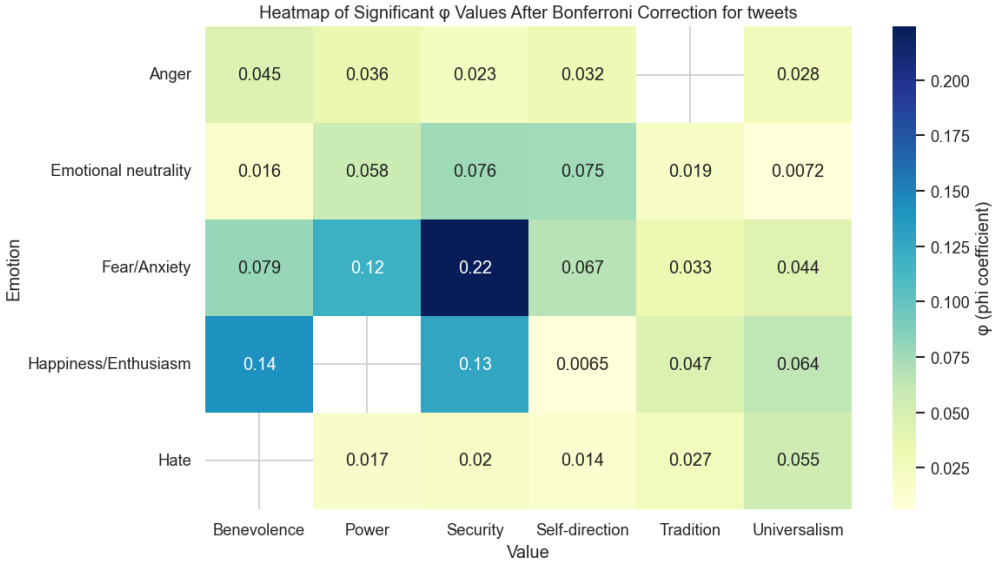


Figure 1 Heatmap of Significant φ Values After Bonferroni Correction for posts

We could not find a strong association between any of the emotion-values pairs analysed. The most pronounced associations emerged between fear/anxiety and security (φ = 0.224), representing the strongest relationship in the dataset. This substantial effect indicates that fear-based emotional expressions co-occur with security-oriented value statements in approximately 22% of cases beyond what would be expected by chance.

Happiness/enthusiasm demonstrated strong associations with benevolence (φ = 0.142) and security (φ = 0.128), suggesting that positive emotional expressions frequently align with prosocial and stability-oriented values. These medium-to-small effect sizes indicate slightly meaningful practical significance beyond statistical significance.

Fear/anxiety also showed notable associations with power (φ = 0.119), indicating that anxiety-based discourse often co-occurs with power-related value expressions.

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Additionally, three emotion-value combinations failed to achieve statistical significance after multiple testing correction: happiness/enthusiasm with power (p = 0.3394, φ = 0.001), anger with tradition (p = 0.5397, φ = 0.001). and hate with benevolence (p = 0.3580, φ = 0.001).

These findings reveal systematic emotional-evaluative structures underlying political communication, with security-fear and happiness-benevolence representing the most robust association patterns. The absence of happiness-power connections suggests distinct affective pathways for different value orientations, supporting theoretical models of emotion-cognition integration in political contexts.

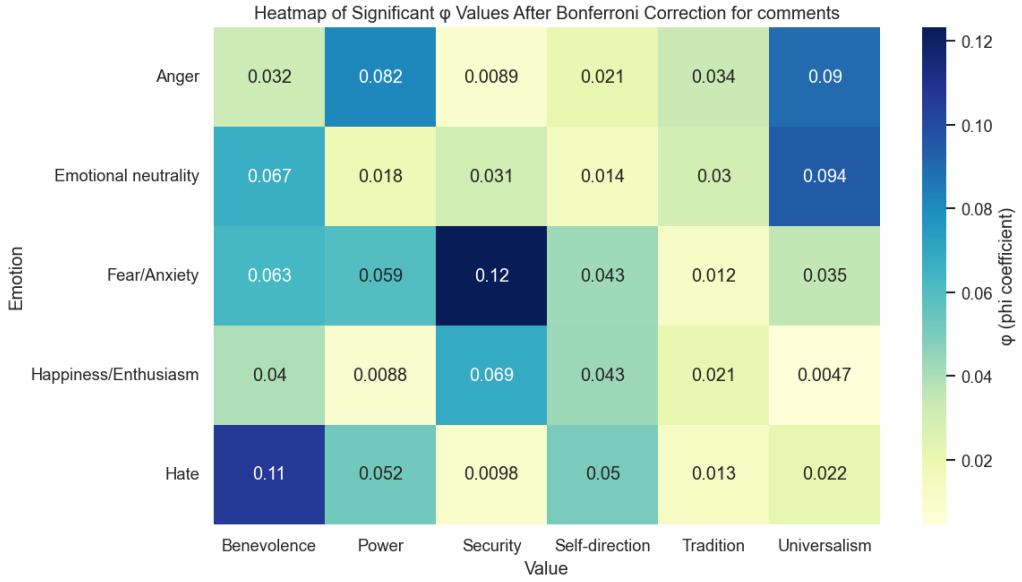


Figure 2 Heatmap of Significant φ Values After Bonferroni Correction for comments

To further understand the reaction this communication generates, we applied the same analysis to comments on posts by politicians to explore differences in emotion and value correlations (Figure 2). Chi-square tests show that emotions and value frames in comments are not independent: all 30 emotion×value pairs remain significant after Bonferroni correction (all $p < .001$). Given the large sample, we emphasise effect sizes (φ≈.005–.123), which are uniformly small yet indicate systematic dependencies. These results warrant interpretation in terms of patterned co-occurrence rather than large practical effects. Detailed results can be consulted in Annex Table 14 .

In this case the strongest associations (nonetheless relatively small) were fear/anxiety with security (φ = 0.123), hate with benevolence (φ = 0.106), emotional neutrality with universalism (φ = 0.094), and anger with universalism (φ = 0.090) and power (φ = 0.082). Happiness/enthusiasm related most to security (φ = 0.069) and self-direction (φ = 0.043); fear/anxiety also paired with benevolence (φ = 0.063) and power (φ = 0.059). In contrast, happiness/enthusiasm showed negligible links to universalism (φ = 0.005) and power (φ = 0.009), indicating that celebratory discourse rarely leans on those frames.

However, small φ values mean effects are modest despite very small *p*-values; we therefore avoid over-stating practical significance. Additionally, χ^2 is non-directional: it signals dependence, but not which categories are over- or under-represented.

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

5.2 USER CATEGORIES' AND COMMENTS EMOTIONAL EXPRESSION

The coefficient estimates presented in Figure 3 quantify the extent to which emotional expression in politicians, media outlets and comments to politicians posts diverges from that of the general public political debate (posts). Across all emotions, the 95% confidence intervals exclude zero, indicating statistically significant deviations from the baseline (Annex D Table 15).

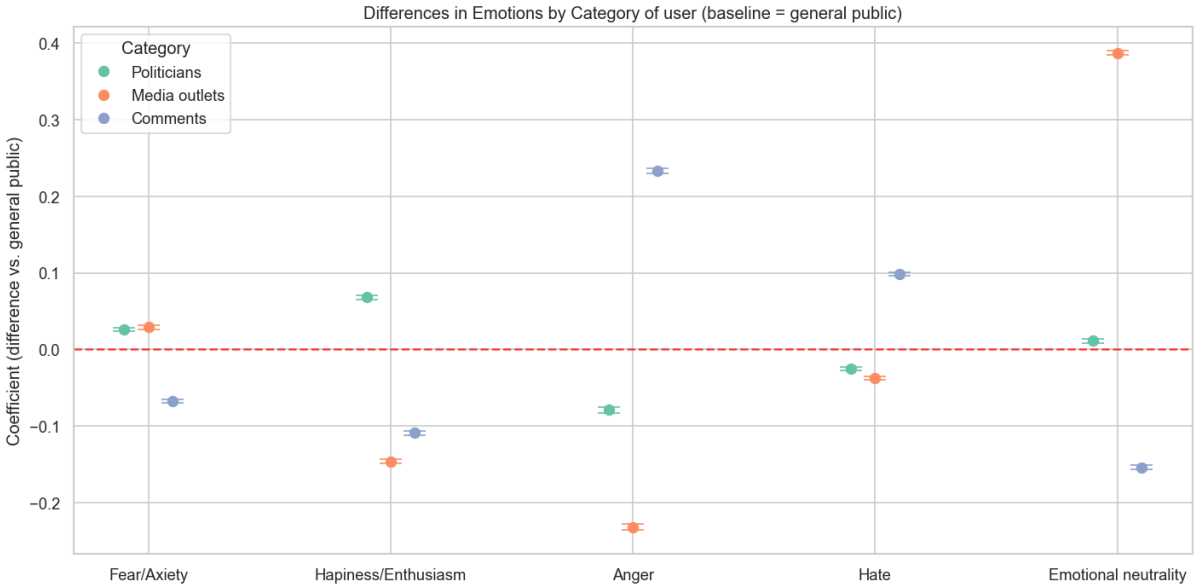


Figure 3: Coefficient plot showing differences in emotional expression for politicians, media outlets and comments to politicians posts compared to posts (general public)

Politicians' communications exhibit a bimodal affective signature characterised by simultaneous amplification of certain emotions and attenuation of others. Notably, happiness and enthusiasm also increase (coef = 0.068; CI [0.065, 0.071]), reflecting strategic positive framing. Concurrently, fear and anxiety are also more present than in the general public discourse (coef = 0.025; CI [0.022, 0.028]), indicating a propensity to emphasise threat-related content. In contrast, expressions of anger (coef = –0.079; CI [–0.083, –0.076]) and hate (coef = –0.025; CI [–0.026, –0.024]) are muted, suggesting a calibrated avoidance of overt hostility. Emotional neutrality shows a very small over the baseline value (coef = 0.011; CI [0.007, 0.015]), indicative of occasional neutral or informational tone.

Media discourse manifests a markedly neutralised affective profile. Emotional neutrality is pronounced being the most significant difference from the baseline of users (coef = 0.385; CI [0.381, 0.389]), reflecting a preference for detached or objective tone. They also show slightly greater degree of fear and anxiety (coef = 0.028; CI [0.025, 0.031]), consistent with sensationalist or risk-focused reporting. However, positive affect is substantially suppressed, with happiness and enthusiasm sharply lower (coef = –0.146; CI [–0.150, –0.143]). Expressions of anger (coef = –0.230; CI [–0.233, –0.226]) and hate (coef = –0.037; CI [–0.038, –0.036]) are also attenuated relative to posts, reinforcing a restrained emotional register.

Emotional expression within comments responding to politicians' posts reveals distinctive affective patterns when compared to general-public posts. Specifically, the regression

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

analysis indicates a statistically significant elevation in expressions of anger (coef = 0.233, CI [0.230, 0.237]) and hate (coef = 0.098, CI [0.096, 0.101]), suggesting that comments are more likely to convey hostile or adversarial sentiments. Conversely, emotional neutrality is markedly reduced (coef = −0.154, CI [−0.157, −0.152]), indicating a diminished presence of impartial or informational tone. Expressions of fear and anxiety (coef = −0.068, CI [−0.070, −0.066]) as well as happiness and enthusiasm (coef = −0.109, CI [−0.112, −0.107]) are also significantly lower in comments, highlighting a tendency to suppress both threat-related and celebratory affect. These findings, derived from ordinary least squares (OLS) models with posts serving as the reference category, collectively underscore the emotionally charged nature of comment-based discourse in political contexts.

Taken together, the results delineate three distinct affective regimes within the platform's political communication when benchmarked against general-public posts on political topics. Politicians deploy a bimodal emotional strategy, simultaneously amplifying positive affect (happiness/enthusiasm) and elevating threat salience (fear/anxiety), while dampening antagonistic expressions (anger and hate) and keeping neutrality near baseline, consistent with strategic impression management that seeks to mobilise support without incurring reputational costs associated with overt hostility. Media outlets exhibit a markedly neutralised register, with a large increase in emotional neutrality and a suppression of both positive and negative affect, alongside a modest rise in fear/anxiety, a pattern congruent with journalistic conventions of detached reporting tempered by risk-oriented coverage. By contrast, comments directed at politicians concentrate antagonistic affect (higher anger and hate) and de-emphasise both positivity and neutrality, while also lowering fear/anxiety, indicating a reactive, confrontational mode of engagement that privileges adversarial signalling over informational or celebratory content. From an interpretive standpoint, platform-level affect should not be treated as a unitary construct: aggregate estimates will mask these systematic, role-contingent asymmetries.

5.3 ELECTORAL PERIODS EMOTIONAL EXPRESSION

The coefficient estimates presented in Figure 4 quantify the extent to which emotional expression in posts during electoral periods diverges from that of non-electoral periods. Identically, the same relation is described in Figure 5 but in this case for comments to politicians posts. Across all emotions, the 95% confidence intervals exclude zero, indicating statistically significant deviations from the baseline (Annex D Table 16 & Table 17)

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

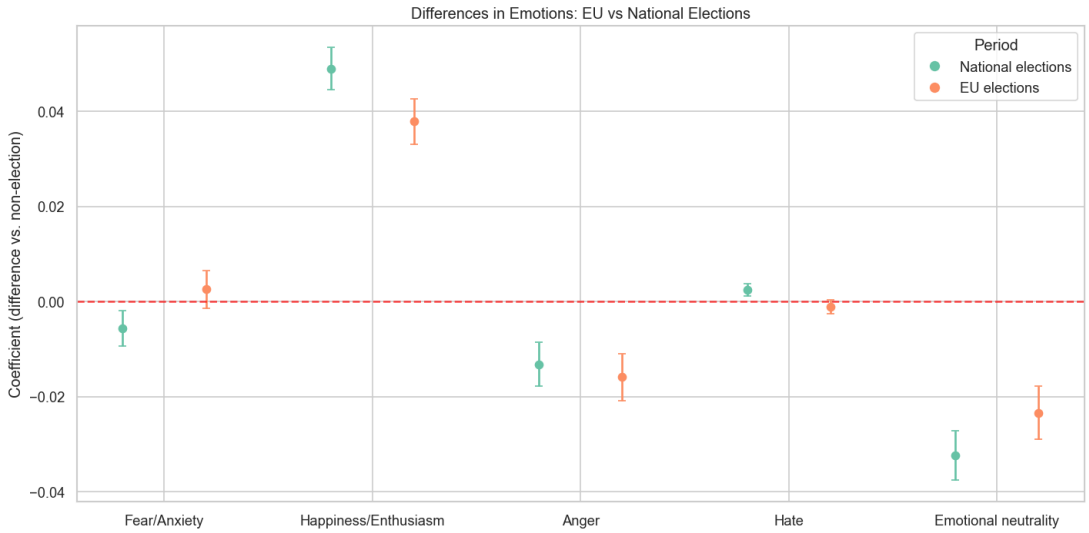


Figure 4 Coefficient plot showing differences in emotional expression in posts for EU & national election period (1 month prior and after election date) compared to posts in a non-electoral period.

During elections, the public timeline shifts toward a mobilising, upbeat register (Figure 4). Happiness/enthusiasm increases significantly during both national elections (coef = 0.049, CI [0.045, 0.053]) and EU elections (coef = 0.038, CI [0.033, 0.043]). Anger becomes less salient across both national (coef = −0.0132, CI [−0.0178, −0.0086]) and EU elections (coef = −0.0159, CI [−0.0208, −0.0110]), hinting at a reorientation from protest to message amplification. Fear/anxiety does not dominate, around national contests it edges down (coef = −0.0057, CI [−0.0094, −0.0020]), and around EU contests it stays broadly stable (coef = 0.0026, CI [−0.0014, 0.0066]). Emotional neutrality declines during both national (coef = −0.0324, CI [−0.0376, −0.0272]) and EU elections (coef = −0.0234, CI [−0.0290, −0.0178]). Signals of hate show only marginal movement, both at national elections (coef = 0.0024, CI [0.0011, 0.0037]) and at EU elections (coef = −0.0011, CI [−0.0025, 0.00027]). Overall, the non-electoral baseline gives way to more positive, less neutral discourse when the campaign clock is running.

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Figure 5 Coefficient plot showing differences in emotional expression in comments to politicians' posts for EU & national election period (1 month prior and after election date) compared to comments to politicians' posts in a non-electoral period.

Reply spaces behave differently, with confrontation intensifying during election windows. Anger rises substantially at both national (coef = 0.0143, CI [0.0120, 0.0165]) and EU elections (coef = 0.0198, CI [0.0176, 0.0220]), revealing an oppositional mode of engagement. Emotional neutrality contracts during national (coef = −0.0113, CI [−0.0125, −0.0100]) and EU elections (coef = −0.0088, CI [−0.0101, −0.0076]). Notably, hate does not surge but rather declines modestly in both national (coef = −0.0242, CI [−0.0259, −0.0226]) and EU elections (coef = −0.0099, CI [−0.0115, −0.0082]), potentially reflecting moderation practices or self-selection away from extreme language. Alternatively, it is also plausible that hate constitutes such an intense and internally driven emotion that its expression remains largely invariant to the electoral calendar. The emotional mix differs by election type: happiness/enthusiasm increases during national elections (coef = 0.0273, CI [0.0260, 0.0287]) but decreases slightly at EU elections (coef = −0.0032, CI [−0.0045, −0.0018]), while fear/anxiety falls during national elections (coef = −0.0061, CI [−0.0072, −0.0050]) but slightly increases in EU elections (coef = 0.0021, CI [0.0010, 0.0032]), this may be due to EU elections foregrounding more transnational issues that are closely tied to global problems likely to trigger fear. In short, replies concentrate the heat of electoral politics, more pushback, less neutrality, while showing only muted changes in overt hostility.

Electoral cycles leave distinct imprints on online political discourse, but these effects diverge sharply between the general public timeline and the reply spaces attached to political actors. Whereas public-facing posts during campaign periods adopt a more mobilising and affectively positive tone, with anger receding, neutrality contracting, and anxiety only modestly engaged, comment sections take on a different tenor. Replies to politicians intensify contestation: anger becomes more pronounced, neutral registers diminish, and communicative focus shifts towards antagonistic interaction. At the same time, signs of hate speech do not structurally escalate; if anything, they decline slightly, potentially reflecting moderation pressures or self-censorship. These contrasting dynamics underscore a dual conservationism of electoral communication: public timelines serve as arenas of enthusiasm and amplification, while reply threads crystallise electoral confrontation, with emotional trajectories shaped by both national and European contexts.

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

5.4 MEDIA CLUSTERING ANALYSIS

The clustering analysis of emotional and neutral posts by media outlets on X, describing how this emotional change affects their diffusion and value association.

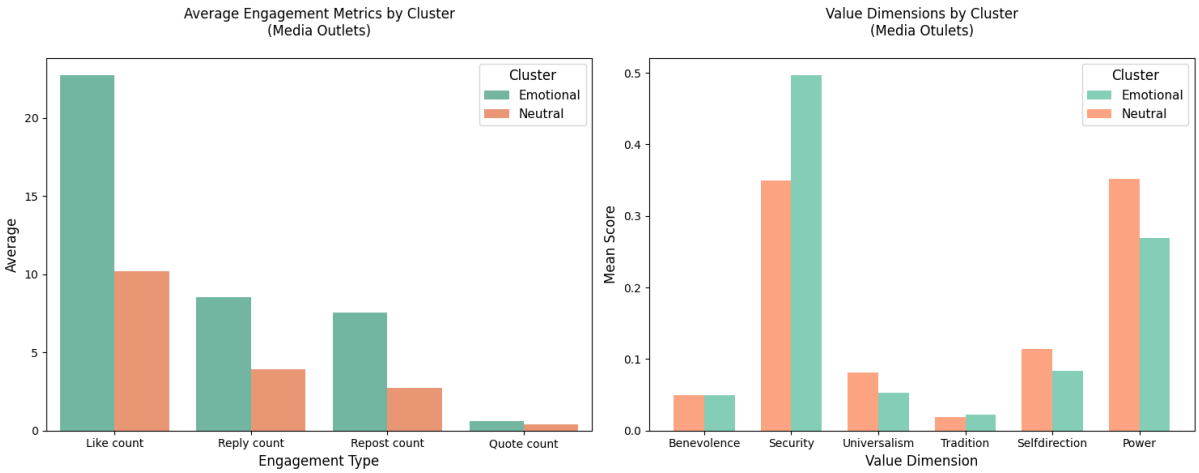


Figure 6: Engagement and values of media posts with and without emotional charge.

Emotional content significantly outperformed neutral content across all engagement metrics ($p < 0.001$ for all comparisons). Emotional posts generated substantially higher average engagement: likes ($M = 22.72$ vs. 10.22), replies ($M = 8.52$ vs. 3.95), reposts ($M = 7.53$ vs. 2.76), and quotes ($M = 0.61$ vs. 0.42). Total engagement per post was 127% higher for emotional content ($M = 39.38$) compared to neutral content ($M = 17.36$), representing a large practical effect. Despite comprising only 37.3% of posts, emotional content accounted for 57.4% of total engagement activity.

Significant differences emerged across all value dimensions ($p < 0.001$). Security values were more the only ones that were more prominent in emotional posts ($M = 0.496$ vs. 0.350). Conversely, neutral content showed higher universalism ($M = 0.081$ vs. 0.053), power orientation ($M = 0.269$ vs. 0.352) and self-direction values ($M = 0.115$ vs. 0.084). Tradition and benevolence values showed minimal difference between clusters.

The results show that emotionally activated media content consistently generates superior engagement across all metrics, with emotional posts achieving approximately double the engagement rates of neutral content. The value dimension analysis reveals that emotional content appeals to security motivations, while neutral content aligns with universalism, power and self-direction values.

5.5 ENGAGEMENT ANALYSIS

5.5.1 EMOTIONS AND USER CATEGORY/COMMENTS

The regression analysis yielded a substantial explanatory power with an R-squared value of 0.584, indicating that the category-emotion interaction model accounts for approximately 58.4% of the variation in log-transformed engagement levels. This represents a strong model

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

fit for social media engagement data, which is typically characterised by high heterogeneity. The F-statistic of 163.1 ($p < 0.001$) confirms the overall statistical significance of the model, while the large sample size of 2,169,852 observations provides robust statistical power for detecting meaningful effects. Detailed results are included in Annex D Table 18, Table 19 and Table 20.

The analysis reveals pronounced heterogeneity in engagement patterns across content categories relative to the baseline "general public posts" category. Politicians' content demonstrates the most substantial main effect, with a coefficient of 2.737 ($p < 0.001$), translating to a remarkable 1,444% increase in engagement compared to baseline content when emotional tone remains neutral. This finding underscores the inherently high-engagement nature of political discourse on social media platforms.

Conversely, comment content exhibits a significant negative main effect ($\beta = -0.838$, $p = 0.001$), corresponding to a 56.7% reduction in engagement relative to baseline posts. This suggests that comment-based content generates substantially lower engagement when emotional neutrality is maintained. Media content shows a moderate negative but non-significant effect ($\beta = -0.267$, $p = 0.484$), indicating a 23.4% decrease in engagement that does not reach statistical significance.

The emotional tone analysis reveals happiness/enthusiasm as the most potent emotional driver, with a coefficient of 0.848 ($p = 0.001$), representing a 133.4% increase in engagement when applied to neutral post content. This finding aligns with established literature on positive emotion contagion in digital environments.

Fear/anxiety demonstrates a moderate positive but non-significant effect ($\beta = 0.366$, $p = 0.140$), suggesting a 44.2% engagement increase that fails to reach statistical significance. Similarly, anger ($\beta = 0.224$, $p = 0.367$) and hate ($\beta = 0.105$, $p = 0.680$) show modest positive effects of 25.1% and 11.1% respectively, though neither achieves statistical significance as main effects.

The interaction terms reveal complex patterns that significantly modify the main effects across different category-emotion combinations. Several notable interaction patterns emerge from the analysis, as presented in the interaction matrix (Figure 7).

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

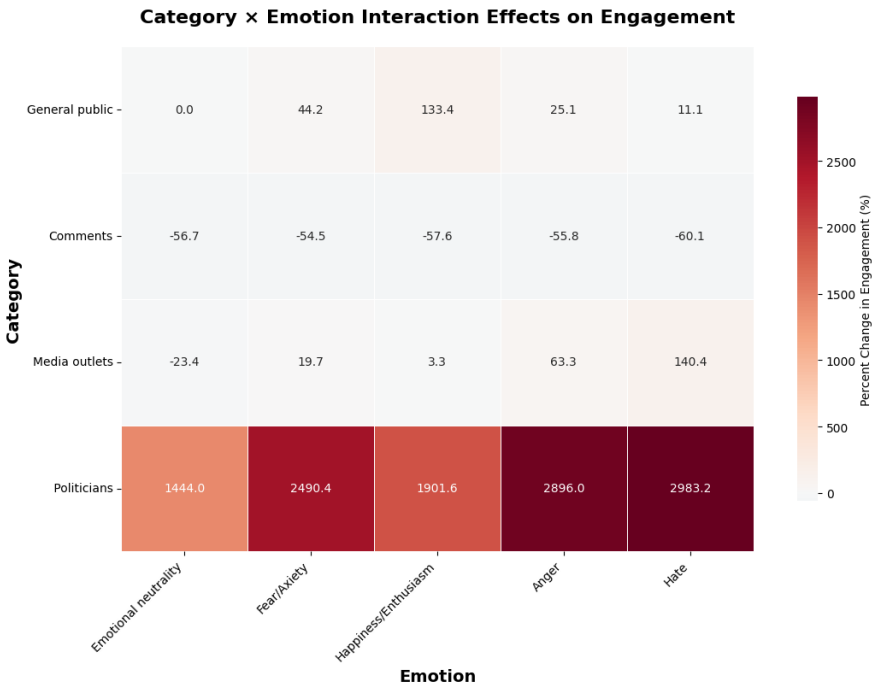


Figure 7 Interaction matrix for emotion x category in relation to engagement

Political content interactions consistently amplify emotional effects, with all political-emotion combinations producing substantial engagement increases. The pol × hate interaction ($\beta = 0.587$, $p = 0.037$) and pol × anger interaction ($\beta = 0.439$, $p = 0.093$) suggest that negative emotions are particularly potent when combined with political content, yielding total effects of 2,983% and 2,896% respectively.

Comment content interactions demonstrate a contrasting pattern, with the happiness/enthusiasm interaction producing a significant negative coefficient ($\beta = -0.868$, $p < 0.001$). This interaction effect partially offsets the positive main effect of happiness, resulting in comment content with a happy emotional tone still underperforming relative to baseline (-57.6% total effect).

Media content interactions exhibit the most variable pattern, with a particularly strong media × hate interaction ($\beta = 1.039$, $p = 0.020$) that transforms media content from a negative main effect into a substantial positive combined effect of 140.4%.

The significance patterns reveal that main category effects demonstrate greater statistical reliability than main emotion effects, with political content showing robust significance and comment content achieving borderline significance. Among interaction effects, the most reliable findings involve combinations with political content and negative emotional tones, as well as the suppressive effect of happiness on comment engagement.

The generally weak significance of main emotion effects, contrasted with stronger interaction effects, suggests that emotional impact is highly context-dependent and cannot be adequately understood without considering content category. This finding emphasises the importance of the interaction framework in understanding social media engagement dynamics.

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

The results demonstrate that content category serves as a primary determinant of engagement potential, with emotional tone acting as a significant modifier of these baseline effects. The exceptional performance of political content across all emotional conditions, combined with the amplifying effects of negative emotions, suggests that political discourse occupies a unique position in social media engagement ecosystems. Meanwhile, the consistent underperformance of comment content indicates structural barriers to engagement that transcend emotional framing strategies.

5.1.2 EMOTIONS AND ELECTORAL PERIODS

The electoral period-emotion interaction model demonstrates considerably lower explanatory power compared to the category-based analysis, with an R-squared of 0.041, indicating that temporal electoral dynamics and emotional content account for approximately 4.1% of the variation in engagement levels. While this represents a modest effect size, the F-statistic of 3,866 ($p < 0.001$) confirms the statistical significance of the temporal-emotional framework. The substantial sample size of 1,419,944 observations provides sufficient statistical power to detect meaningful effects despite the relatively small effect magnitudes. Detailed results can be consulted in the Annex D Table 21, Table 22, and Table 23

The reduced explanatory power suggests that electoral timing, while statistically significant, represents a secondary factor in engagement determination compared to the type of users and the category of post or comment, highlighting the primacy of authorship and communication means over temporal context in driving social media interactions.

5.6 COUNTRY COMPARISONS

Emotional Expression Across Countries: Cross-National Variation in Hostility and Reactivity
Across all countries, comments on politicians' posts emerge as the emotional "hot zone" of the political communication ecosystem. This pattern is visualised indirectly in Figure 3, where comments are associated with large positive coefficients for *anger* (0.233) and *hate* (0.098) relative to general public posts, while happiness and fear are sharply reduced. When this overarching pattern is disaggregated by country, however, substantial national differences emerge.

Based on Table 12, we observe that Bosnia and Herzegovina (BA) exhibits the most extreme emotional negativity: 89.34% of comments express anger, and 99.27% express hate. This near-ubiquity of hostile affect aligns with broader research on contentious digital discourse in ethnically polarised political systems. Denmark (DK), despite having a more stable political environment, shows similarly elevated antagonism (87.85% anger; 89.33% hate). This suggests that robust democratic norms do not necessarily dampen reactive negativity in reply channels, possibly due to Denmark's high social media penetration. Austria (AT) presents a slightly milder but still intense pattern (86.60% anger; 94.95% hate), consistent with the country's polarised debates on migration and governance. Poland (PL) shows comparatively lower levels, though still high in absolute terms (anger ~66%, hate ~76%), mirroring the country's highly polarised but structurally less fragmented digital public sphere.
Bulgaria (BG) sits between AT and PL levels (anger 79.12%, hate 96.81%), reflecting a communication environment marked by frustration with corruption and governance. Taken together, while all countries show heightened antagonism in comments, the *degree*

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

of this negativity differs sharply. These variations reflect national political tensions, media trust levels, and the role of digital platforms in public debate.

Politicians: Divergent Emotional Styles

Figure 3 shows the general tendency of politicians to amplify *happiness/enthusiasm* (+0.068) and *fear/anxiety* (+0.025) relative to the general public, while reducing anger and hate. This creates a characteristic "bimodal profile" of reassurance and mobilisation.

Country-level distributions (Annex Table 12) show striking variation:
AT and PL stand out for comparatively high *positive emotionality* in political communication. Austrian politicians express happiness/enthusiasm in 28.49% of posts, Polish politicians in 26.37%, both above the cross-country median. This aligns with political strategies centred on hopeful narratives and emotional mobilisation during polarised electoral cycles. BG also exhibits relatively high happiness (39.25%) but couples it with comparatively high fear/anxiety (28.59%), creating a "dual emotional register" blending mobilisation with threat-based appeals. BA displays lower positive affect (18.9%) coupled with elevated fear/anxiety (16.59%), consistent with political communication styles that emphasise insecurity, identity, and risk. DK and MK show more restrained emotionality, consistent with institutional norms of moderation (DK) and the presence of hybrid political communication cultures (MK). These national divergences demonstrate that politicians' strategic use of emotion is embedded in the political and cultural context: some emphasise positive mobilisation (AT, PL), others threat or vigilance (BA, BG), and still others maintain emotional restraint (DK).

Media: National Variants of Emotional Neutrality

Media outlets across all countries show strong emotional neutrality, as reflected in Figure 3, where media display a large positive coefficient for emotional neutrality (+0.385) and reduced levels of all explicit emotions. Still, national deviations are notable:

BA and AT media show relatively high fear/anxiety (BA: 47.57%, AT: 25.55%), suggesting that even ostensibly neutral outlets may adopt more threat-oriented framing in environments with heightened political stress or sensationalist media competition. PL and BG media display moderate emotional expression but place stronger emphasis on *values* rather than emotions (see Section 2). DK media is the most neutral, consistent with a strong regulatory and public-service journalism tradition that discourages overt affective signalling.

Value Expression: Cross-Country Divergence in Normative Framing

Security as a Shared but Unevenly Emphasised Value: Across countries, **security** appears prominently in emotionally charged communication, aligning with the strongest emotion–value association found in Figure 1 and Figure 2 (fear/anxiety - security).

However, different nations exhibit different *intensity levels*: BA, AT, and DK show particularly high security emphasis in comments (BA: 64.48%, AT: 58.71%, DK: similar high levels), consistent with either political fragmentation (BA) or debates on migration and welfare (AT, DK). PL and BG show strong security emphasis in political and media content, reflecting geopolitical pressures (Ukraine war proximity for PL) and institutional instability (BG).
These patterns align with the chi-square analyses in Figure 1, where the strongest effect (φ = 0.224) links fear/anxiety to security and appears consistently across countries.

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Universalism: A Divisive and Polarising Value: National differences are most evident in the value of *universalism*: in BA and AT comments, universalism appears prominently (AT: 58.49%, BA: 60.28%) *despite* high levels of anger and hate. This indicates that universalist values (equality, tolerance, human rights) become active battlegrounds in highly polarised reply spaces. DK, BG, and PL, especially in media and politicians' posts, display *lower universalism*, suggesting these debates are less central or expressed differently. This disparity mirrors emotional contestation: where universalism is salient, it often coexists with antagonistic emotions, reflecting contentious identity politics.

Tradition and Power: Country-Specific Identity Signifiers: Values such as *tradition* and *power* show the strongest national signature:
PT exhibits exceptionally high references to *tradition* across comments and political communication, consistent with the country's strong cultural conservatism. AT and PT have elevated *power* values, which may reflect political debates around sovereignty, governance, and institutional control. BG shows a relatively balanced value distribution, whereas DK shows lower overall value signalling, consistent with its emotionally neutral profile.

Linking National Patterns to Statistical Figures: Emotion–Value Associations: The heatmaps reveal that the strongest emotion–value ties-fear-security, happiness-benevolence- are robust across countries. The national comparison suggests:

Countries with higher fear/anxiety levels (BA, BG) also show stronger security framing in both comments and political content. Countries where happiness is used politically (AT, PL) display more benevolence and universalism in political discourse. Thus, the heatmap patterns scale differently across national contexts, intensifying or softening depending on local communication cultures.

User Category Differences when combined with Table 12: Countries with the most negative comment cultures (BA, AT, DK) also show the strongest alignment with Figure 3 "anger-heavy, happiness-reduced" comment profile. Countries where politicians rely more on positive mobilisation (AT, PL) align with Figure 3 positive coefficients for happiness (+0.068). Thus, the country-level data illustrate how the general-category patterns play out with varying intensity across nations.

Conclusion: A Fragmented European Emotional Landscape
The cross-country comparison reveals that while all six national communication ecosystems share certain structural patterns, particularly the negativity of comments and the neutrality of media, the emotional and value dynamics differ sharply across national contexts. These differences reflect:

- political polarisation (BA, PL),
- media system characteristics (DK, AT),
- institutional trust levels (BG, MK),
- and geopolitical pressures (PL, BG, AT).

Crucially, these national variations *magnify or attenuate* the broader patterns visualised in the figures: fear-security linkages (Figure 1), category-based emotional signatures (Figure 3), and cross-category emotional hierarchies. Together, they indicate that European political emotions online are shaped not only by platform dynamics but also by deeply rooted national political cultures, turning social media into a multilayered emotional topography rather than a uniform communication space.

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

CONCLUSIONS

Drawing on a multilingual corpus of 2,169,852 posts and comments across Austria, Bosnia and Herzegovina, Bulgaria, Denmark, North Macedonia, and Poland, this deliverable demonstrates that the emotional architecture of online political communication is systematic yet context-dependent, with modest but robust emotion–value alignments and pronounced role- and period-specific asymmetries.

Chi-square tests show that most emotion–value pairs are statistically dependent, albeit with small effect sizes; the most substantive pattern links fear/anxiety to security, while happiness/enthusiasm co-occurs more often with benevolence and, to a lesser degree, security, indicating that stability- and prosocial frames scaffold both threat- and celebration-oriented discourse. Against this background, affect is strongly stratified by actor: politicians combine mobilisation with reassurance by elevating happiness/enthusiasm and fear/anxiety while dampening anger and hate; media adopt a markedly neutral register with a slight tilt toward fear/anxiety consistent with risk-focused reporting; and comments to politicians concentrate antagonistic affect (higher anger and hate) while de-emphasising both positivity and neutrality, underscoring the reactive and confrontational nature of reply spaces. Electoral timing further reconfigures these baselines: public timelines around both national and EU elections become more positive and less neutral (with anger receding), whereas replies intensify contestation (anger rises, neutrality falls) without a structural surge in hate, plausibly reflecting moderation or self-selection effects.

Engagement dynamics reveal that content category is the primary driver (with posts by politicians showing the largest amplification), while emotional tone acts as a modifier: happiness/enthusiasm is the most consistent main-effect enhancer, and negative emotions gain traction chiefly through interactions with politicians and, at times, media content; in line with this, emotional media posts attract substantially higher activity than neutral ones.

Taken together, these results caution against treating “the online public” as effectively homogeneous: aggregate estimates will obscure systematic actor- and period-specific regimes of expression and diffusion. Substantively, they suggest that constructive, value-grounded positive framing can increase attention without resorting to polarising rhetoric, while the antagonistic affordances of reply spaces call for targeted, proportionate interventions to mitigate hostility and support deliberation. Practically, the evidence developed here provides the empirical foundation for D3.4 (Catalogue of Best Practices) and informs the co-creation of future emotional narratives (WP6) and policy-oriented foresight (WP7), with the overarching aim of strengthening democratic resilience through emotionally intelligent communication. The data annotated also serves to generate the synthetic post from WP4 biometric research, and the insights and further work on specific topics will be done within WP5 to triangulate research outcomes from the survey studies.

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

REFERENCES

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised Cross-lingual Representation Learning at Scale. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8440–8451. https://doi.org/10.18653/v1/2020.acl-main.747

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., & others. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Horne, B. D., & Adali, S. (2017). *This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News* (No. arXiv:1703.09398). arXiv. https://doi.org/10.48550/arXiv.1703.09398

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95.

McKinney, W., & others. (2010). Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51–56) [Computer software]. SciPy.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Reitz, K. (2022). *Requests: HTTP for Humans* (Version 2.28.1) [Computer software]. https://requests.readthedocs.io/

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. *Proceedings of the 9th Python in Science Conference*, *57*.

Sechidis, K., Tsoumakas, G., & Vlahavas, I. (2011). On the Stratification of Multi-label Data. In D. Gunopulos, T. Hofmann, D. Malerba, & M. Vazirgiannis (Eds.), *Machine Learning and Knowledge Discovery in Databases* (Vol. 6913, pp. 145–158). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-23808-6_10

Van Rossum, G., & Foundation, P. S. (n.d.). *Python Programming Language* (Version 3.x) [Computer software]. https://www.python.org/

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., … Rush, A. M. (2020). Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural*

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Language Processing: System Demonstrations, 38–45. https://www.aclweb.org/anthology/2020.emnlp-demos.6

X Corp. (2024). *X API v2*. https://docs.x.com/x-api

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

ANNEXES

ANNEX A – CODEBOOK

INTRODUCTION

Unit of Analysis

The unit of analysis for this annotation framework includes:

- Posts from X (formerly Twitter)
- Comments associated with the X posts

All material should be treated equally as a potential site for annotation.

Each individual post, transcribed segment, or comment is a separate unit to be coded.

Scope of Annotation

For each unit of analysis, annotators must code both values and emotions:

- Assess and code for the presence of values and emotions conveyed, whether by the original poster, the author of a comment, or a subject described within the text.
- Code both explicit and implicit emotions:
  - Explicit emotions are those directly stated or named (e.g., "I am scared", "She was thrilled").
  - Implicit emotions are those indicated through context, behaviour, narrative structure, tone descriptions, or indirectly referenced feelings (e.g., "My hands shook as I read the message" – implying fear or anxiety).
- Code all values mentioned in the text, regardless of whether the poster agrees, disagrees, or is neutral about them.
  - This includes values the poster supports, criticises, discusses neutrally, or attributes to others.
  - If multiple values appear in the same text, code each one as present.
  - Do not try to judge the poster's personal stance; simply mark every value that is referenced or discussed in any way.

Columns

Each column can be given any of the allowed values regardless of the answers in the other columns.

IMPORTANT: No cells can be left blank. All the cells of the codebook must be filled with an allowed value.

POLITICS

Column	Annotation:	Description
On topic	YES: 1 NO: 0	Posts related to politics (allocation of resources and power contestation values, norms, or ideals with societal impact) are on topic within the ENCODE definition. Some guiding questions to identify these posts can be:

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

		• Does it relate to the activity of political institutions (decision-making, lawmaking, judiciary, wars, transnational institutions)? • Does it relate to legal rights, norms, or responsibilities of citizens and/or non-citizens? • Does it relate to values that regulate collective life (e.g., peace, justice, equality, welfare) or private life of importance to the political (e.g., distribution of responsibilities in the family, gendered norms, ideas relating to identity)? • Does it relate to dealing with social or global challenges using regulations (e.g., climate change, global political order, so-called gender or identity politics)?

VALUES

Values Framework for Social Media Analysis

We will manually annotate 6 of the values from the Basic Human Values theory of Shalom Schwartz[1]. This selection was based on desk research which found that Schwarz model is the most widely used and relevant for understanding values, especially in relation to politics. In addition, it has frequently been used for social media analysis, including manual annotation, which is an important requirement for this work. Furthermore, research conducted in WP2 confirms this and proposes this as the main model to use for the ENCODE project.

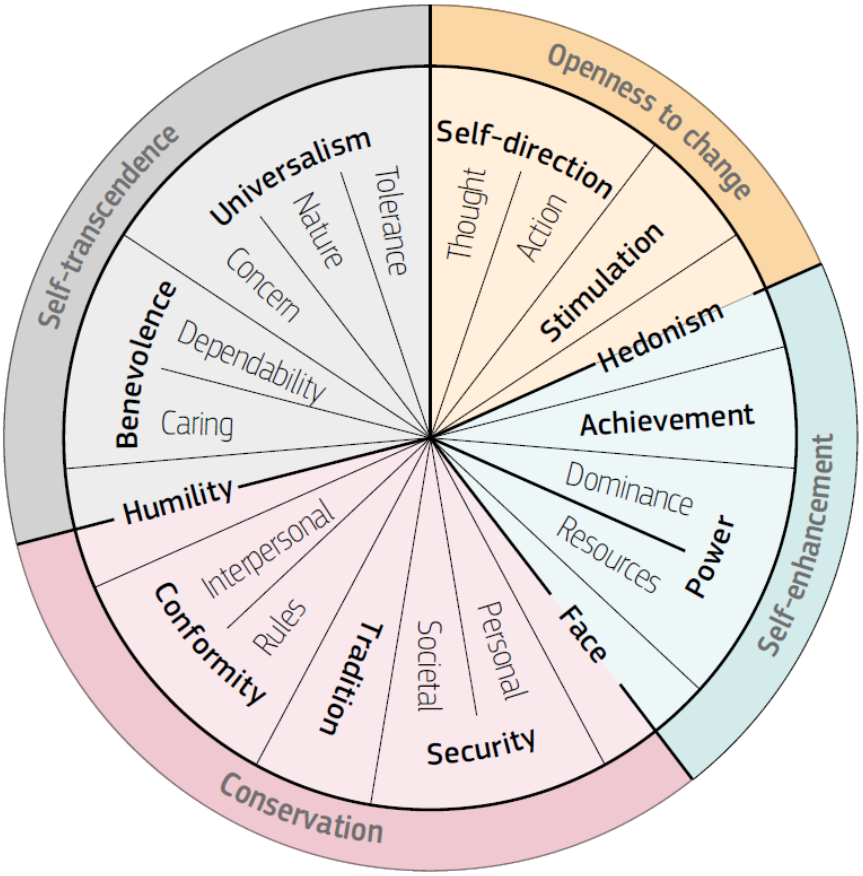
The selection of the 6 values has been made based on the following criteria:

1. Parsimony and distinction: To ease manual annotation, to ensure a high-quality and consistent annotated corpus, we want to limit the number on values as much as possible and make sure that they are as different as possible.
2. Coverage: Ensure coverage of the four higher order values of Schwartz model: Openness to change, Conservation, Self-enhancement and Self-transcendence.
3. Relevance: We want to include the most relevant and frequently expressed emotions in the EU

Criteria 1 and 2 goes hand in hand since parsimony is easiest achieved by removing values that are similar to one another. Moreover, similar values will be captured under the same higher-order value as displayed in the figure below.

[1] chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://scholarworks.gvsu.edu/cgi/viewcontent.cgi?article=1116&context=orpc

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).



In line with this reasoning, we also discard values that cover several higher-order values such as hedonism. To ensure relevance, we rely on the report from the European Commission's Joint Research Centre[2]. From large-scale survey, it has been identified that the following values are the most common in the EU:

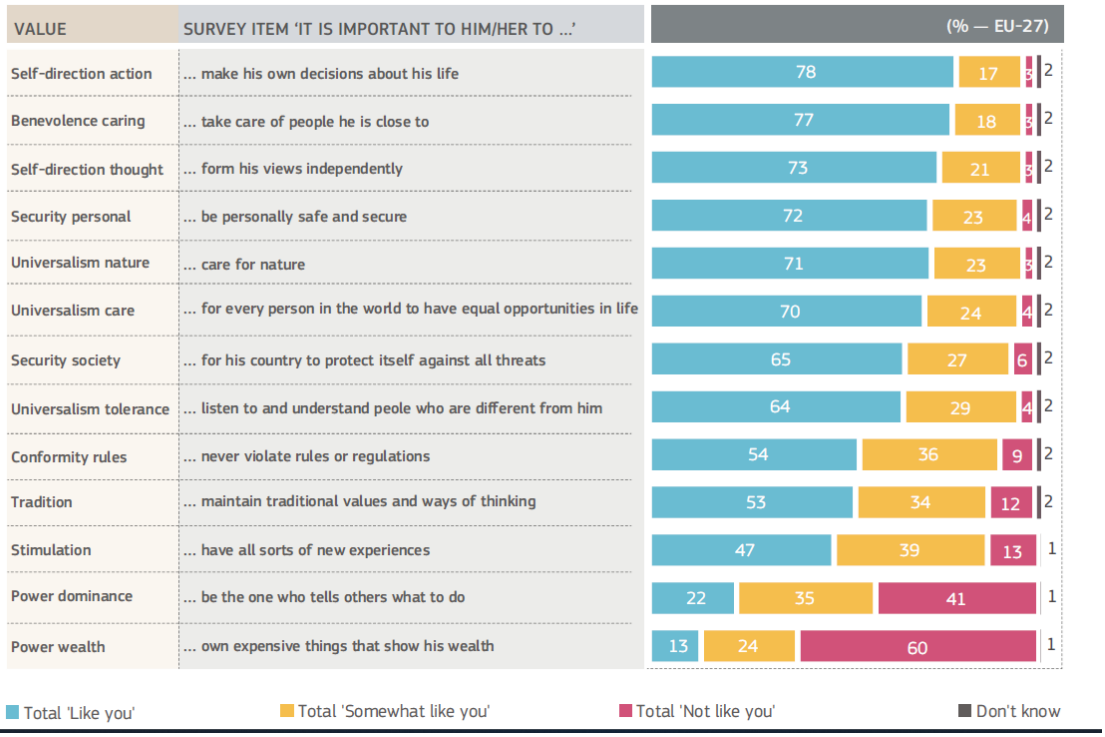
- Self-direction – Openness to change
- Benevolence – Self transcendence
- Security - Conservation
- Universalism - Self transcendence
- Conformity/tradition – Conservation

The Figure below shows the results of the survey.

[2] JRC Publications Repository - Values and Identities - a policymaker's guide

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Figure 6: Personal values priority in the EU.



Therefore, and since these values fulfil our other two criteria: They cover 3 out of 4 of the higher-order values and they are distinct from each other and limit the number of values, we select these values (except conformity) for the framework of analysis and the manual annotation. During the methodology design, it was hard to make a clear operational definition to be able to differentiate this nuance between them in a simple social media post without contextual information. We select tradition over conformity (for conservation) since we believe that tradition will be easier to manually annotate and since it is more expected to be more frequent in political discussions.
Finally, we include power to cover the last higher-order value of Self-enhancement.
Consequently, our final selection is:

- Self-direction – Openness to change
- Benevolence – Self transcendence
- Security - Conservation
- Universalism - Self transcendence
- Tradition – Conservation
- Power – Self-enhancement

Benevolence
Defining goal: preserving and enhancing the welfare of those with whom one is in frequent personal contact (the 'in-group'). Benevolence values derive from the basic requirement for smooth group functioning (cf. Kluckhohn, 1951) and from the organismic need for affiliation (cf. Maslow, 1965). Most critical are relations within the family and other primary groups. Benevolence values emphasize voluntary concern for others' welfare.
(helpful, honest, forgiving, responsible, loyal, true friendship, mature love)
[sense of belonging, meaning in life, a spiritual life].

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Security
Defining goal: safety, harmony, and stability of society, of relationships, and of self. Security values derive from basic individual and group requirements (cf. Kluckhohn, 1951; Maslow, 1965). Some security values serve primarily individual interests (e.g., clean), others wider group interests (e.g., national security). Even the latter, however, express, to a significant degree, the goal of security for self or those with whom one identifies.
(social order, family security, national security, clean, reciprocation of favours)
[healthy, moderate, sense of belonging]

Universalism
Defining goal: understanding, appreciation, tolerance, and protection for the welfare of *all* people and for nature. This contrasts with the in-group focus of benevolence values. Universalism values derive from survival needs of individuals and groups. But people do not recognize these needs until they encounter others beyond the extended primary group and until they become aware of the scarcity of natural resources. People may then realize that failure to accept others who are different and treat them justly will lead to life-threatening strife. They may also realize that failure to protect the natural environment will lead to the destruction of the resources on which life depends. Universalism combines two subtypes of concern—for the welfare of those in the larger society and world and for nature.
(broadminded, social justice, equality, world at peace, world of beauty, unity with nature, wisdom, protecting the environment)
[inner harmony, a spiritual life]

Tradition
Defining goal: respect, commitment, and acceptance of the customs and ideas that one's culture or religion provides. Groups everywhere develop practices, symbols, ideas, and beliefs that represent their shared experience and fate. These become sanctioned as valued group customs and traditions. They symbolize the group's solidarity, express its unique worth, and contribute to its survival (Durkheim, 1912/1954; Parsons, 1951). They often take the form of religious rites, beliefs, and norms of behaviour.
(respect for tradition, humble, devout, accepting my portion in life)
[moderate, spiritual life]

Self-Direction
Defining goal: independent thought and action--choosing, creating, exploring. Self-direction derives from organismic needs for control and mastery (e.g., Bandura, 1977; Deci, 1975) and interactional requirements of autonomy and independence (e.g., Kluckhohn, 1951; Kohn & Schooler, 1983).
(creativity, freedom, choosing own goals, curious, independent)
[self-respect, intelligent, privacy]
Power
Defining goal: social status and prestige, control or dominance over people and resources. The functioning of social institutions apparently requires some degree of status differentiation (Parsons, 1951). A dominance/submission dimension emerges in most empirical analyses of interpersonal relations both within and across cultures (Lonner, 1980). To justify this fact of social life and to motivate group members to accept it, groups must treat power as a value. Power values may also be transformations of individual needs for dominance and control. Value analysts have mentioned power values as well (e.g., Allport, 1961).
(authority, wealth, social power)

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

[preserving my public image, social recognition]

Categories & Operational Definitions

Column	Annotation:	Description
Benevolence	YES: 1 NO: 0	Preserving and enhancing the welfare of those with whom one is in frequent personal contact (the 'in-group'). Detection Cues: • Lexical: "helpful", "honest", "forgiving", "responsible", "loyal", "friendship", "love", "caring", "support" • Contextual: Posts about helping family/friends, community support, acts of kindness Some guiding questions to identify these posts: • Does the post describe actions taken to help others in one's community? • Are there expressions of loyalty or responsibility toward close connections? • Does the content emphasise interpersonal care or nurturing relationships?
Security	YES: 1 NO: 0	Safety, harmony, and stability of society, of relationships, and self. Detection Cues: • Lexical: "safety", "defence", "protection", "stability", "order", "security", "peace", "threat", "danger" • Contextual: Content about national security, social stability, protection policies Some guiding questions to identify these posts: • Does the post express concerns about safety or stability? • Are there references to defensive measures or protection of society? • Does the content emphasise maintaining order or preventing threats?
Universalism	YES: 1 NO: 0	Understanding, appreciation, tolerance, and protection for the welfare of all people and for nature. Detection Cues: • Lexical: "equality", "justice", "rights", "environment", "nature", "peace", "fairness", "tolerance", "protection" • Contextual: Content about equal rights, environmental protection, global concerns Some guiding questions to identify these posts: • Does the post advocate for the welfare of people beyond one's immediate group? • Are there expressions of concern for environmental issues or nature? • Does the content promote tolerance, equality, or justice for all?

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Tradition	YES: 1 NO: 0	Respect, commitment, and acceptance of the customs and ideas that one's culture or religion provides. Detection Cues: • Lexical: "tradition", "respect", "faith", "heritage", "values", "culture", "religion", "customs", "history" • Contextual: References to religious practices, cultural norms, historical continuity Some guiding questions to identify these posts: • Does the post emphasise adherence to established cultural or religious practices? • Are there references to respecting traditional institutions or authorities? • Does the content promote values associated with cultural or religious heritage?
Self-Direction	YES: 1 NO: 0	Independent thought and action-choosing, creating, exploring. Detection Cues: • Lexical: "freedom", "choice", "independence", "decision", "explore", "create", "individual", "autonomy", "initiative" • Contextual: Content about making personal choices, individual rights, creative expression Some guiding questions to identify these posts: • Does the post emphasise personal choice or individual decision-making? • Are there expressions valuing independence or freedom of action? • Does the content promote self-reliance or taking initiative?
Power	YES: 1 NO: 0	Social status and prestige, control or dominance over people and resources. Detection Cues: • Lexical: "authority", "wealth", "control", "influence", "status", "dominance", "leadership", "power", "strength" • Contextual: Content about social hierarchy, wealth accumulation, leadership positions Some guiding questions to identify these posts: • Does the post emphasise social status or prestige? • Are there expressions valuing control over resources or influence over others? • Does the content focus on leadership authority or social dominance?

EMOTIONS

Emotion Framework for Social Media Analysis

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Based on collaborative categorisation efforts and Ekman's foundational model, we propose **four core emotion categories and one subcategory** to optimise for clear differentiation in social media contexts. This framework balances theoretical coherence with practical detection needs, ensuring direct applicability to downstream project components like sentiment tracking and behavioural prediction.
The process followed the literature research done in D3.1 and the methodology defined in D3.2. Based on this work, we started by defining key constraints and requirements in the definition of emotions.

- **Limited number of emotions.** The coding process requires a limited number of emotions to facilitate the manual coding and enable the successful training of the LLM model.
- **Common and easy to code on social media.** The defined emotions need to appear numerous times within the analysis cohort in order to be able to train the LLM model. Therefore, rare or hard-to-identify emotions are discarded. Furthermore, they need to be clear and non-ambiguous as the manual coding needs to be consistent.
- **Aligned with the ENODE project.** The work in WP3 will be using other work packages, and we also base our definitions on the work done by WP2.

Taking this into consideration, the initial proposal to start the discussion was the Ekman model. Paul Ekman's Basic Emotion Theory posits that emotions comprise discrete, universal categories recognisable across cultures through distinctive facial expressions. Ekman originally identified six basic emotions: anger, disgust, fear, joy, sadness, and surprise, proposing these as fundamental, biologically, based affective states with evolutionary significance. This categorical approach assumes emotions function as independent constructs with distinct neurophysiological pathways, facial expressions, and subjective experiences.
This emotion set is widely used, abides by the limitations detailed above and will also be employed by WP4. However, as it is based on facial expression, some emotions may not be so easy to identify on social media, and given the focus on democracy and emotions, we perform a workshop with all the partners to define which emotions we should employ for the analyses in WP3.
Additionally, it was considered the inclusion of an emotionally neutral state, which can also be understood as calm or rationality. Emotional neutrality represents a vital classification category in sentiment analysis frameworks, as evidenced by its implementation in state-of-the-art NLP systems that classify text as "positive, negative, or neutral"[3]. Even though we are moving from sentiment analysis to emotional analysis, it is still relevant to include such categories as studies of emotional communication patterns among different demographics, where certain populations demonstrate deliberate emotional restraint in digital contexts[4]. This suggests neutrality often represents an intentional communication strategy rather than merely the absence of emotion.

The Selected Emotions

[3] Nakib, A.M., Khan, P., Ullah, M.M., Kawser,, M.L., Jayed, A.K., & Zim, S.K. (2024). Harnessing Advanced NLP Techniques for Automated Personality Analysis and Future Behavior Prediction from Social Media Posts. *Middle East Research Journal of Engineering and Technology*.
[4] Đumić, T., & Veljković, B. (2024). THE ROLE OF SOCIAL MEDIA IN EMOTIONAL COMMUNICATION AMONG INDIVIDUALS IN THEIR THIRD AGE. *Teorija in praksa*.

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

The framework strategically prioritises emotions most prevalent and impactful in political discourse:

1. Fear/Anxiety: This unified category captures reactions to both immediate political threats and uncertainties about future political developments. It was selected because threat perception is a fundamental driver of political behaviour, particularly around elections and policy debates. The decision to combine fear and anxiety reflects their shared functional properties in threat processing, acknowledging that in political discourse, these states often blend together when citizens express concerns about societal issues.

2. Happiness/Enthusiasm: This positive emotional spectrum was included to capture emotional reactions to political victories, policy successes, and expressions of support for political candidates or movements. The intensity spectrum from contentment to enthusiasm allows researchers to distinguish between mild satisfaction and mobilising excitement in political contexts.

3. Anger: Selected for its prevalence in political discourse, anger captures responses to perceived injustice, corruption, or policy failures. This emotion is particularly relevant for understanding political mobilisation and engagement, as anger often motivates political action more effectively than sadness or other negative emotions.

4. Hate (subcategory of Anger): This intensified form of anger was specifically distinguished to capture the more enduring, group-targeted hostility that characterises polarised political environments. By separating hate as a subcategory rather than a distinct emotion, the framework acknowledges the psychological relationship between situational anger and more persistent ideological rejection.

5. Emotional Neutrality: Though not part of Ekman's original model, neutrality was included to capture deliberate affective restraint in political communication. This category is essential for identifying factual reporting, strategic ambiguity, and attempts at objective political analysis amidst emotionally charged discourse.

This framework balances theoretical foundations with practical application needs, optimising for clear differentiation in social media political discourse while maintaining sufficient nuance to capture the emotional dynamics of democratic engagement.

Categories & Operational Definitions

Column	Annotation:	Description
Fear/ anxiety	YES: 1 NO: 0	A threat-response emotional state characterised by anticipation of potential harm or danger, manifesting as either acute reactions to immediate threats (fear) or persistent concern about future uncertainties (anxiety). While theoretically distinguishable, anxiety is an affect and fear is an emotion, as per ENCODE's approach. This unified approach recognises the shared functional properties of fear and anxiety in threat processing while acknowledging their phenotypic variations, providing greater methodological consistency in contexts where precise differentiation presents significant measurement challenges. Detection Cues:

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

		• Lexical: "scared", "worried", "nervous", "panic", "threat" • Contextual: Future-oriented statements, hypothetical scenarios ("What if..."), risk assessments Some guiding questions to identify these posts can be: • Does the post express concern about potential negative outcomes or dangers? • Does the content contain explicit fear/anxiety vocabulary (e.g., afraid, worried, scared)? • Does the author describe physical symptoms associated with fear/anxiety (e.g., heart racing, can't sleep)? • Is there evidence of catastrophic thinking or worst-case scenario planning? • Does the post contain questions about uncertain futures or hypothetical negative scenarios?
Happiness/ enthusiasm	YES: 1 NO: 0	Happiness is a positive emotional state characterised by feelings of joy, contentment, and satisfaction. Enthusiasm is an intense form of happiness, often accompanied by excitement and eagerness towards a particular activity or goal. Detection Cues: • Lexical: "excited", "love", "awesome", emojis (😄, 🎉), superlatives • Contextual: Achievement announcements, celebratory posts, positive reviews Some guiding questions to identify these posts can be: • Does the post contain explicit positive emotion vocabulary (e.g., happy, thrilled, delighted)? • Are positive emojis or celebratory punctuation (e.g., exclamation points) present? • Does the post celebrate political successes or victories? • Are there expressions of enthusiasm for political candidates or policies? • Does the content express anticipation for enjoyable future events?
Anger	YES: 1 NO: 0	Anger is an emotional response to perceived threats, injustice, or frustration expressed in social media posts through language indicating displeasure, blame attribution, or perceived unfairness. Code a post as expressing anger when it contains explicit or implicit expressions of discontent directed at specific actions, decisions, or situations, ranging from mild irritation to intense outrage. Detection Cues: • Lexical: "angry", "mad", "furious", "irritated", "frustrated". Intensifiers: "so", "very", "extremely", "completely". Punctuation patterns: Multiple exclamation points, ALL CAPS segments. • Contextual: Blame attribution: Identifying specific responsible parties. Narration of triggering events: "When X happened, I got angry". References to unfair

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

		treatment: "This shouldn't be allowed". Expression of negative consequences: "This made me waste my time" Some guiding questions to identify these posts can be: • Does the post contain explicit anger vocabulary or intensifiers? • Is there blame attribution directed at specific individuals, organisations, or actions? • Does the content describe situations perceived as unfair or unjust? • Does the author express frustration about negative consequences they experienced? • Are there signs of frustration or outrage about political processes?
Hate	YES: 1 NO: 0	Hate is an intensified, enduring form of anger characterised by persistent negative evaluation of individuals or groups, going beyond reaction to specific incidents to express fundamental rejection of the target's perceived nature or essence. Code a post as expressing hate when it contains dehumanising language, expressions of contempt, or desires for harm directed at entire categories of people rather than specific actions. Detection Cues: • Lexical: "hate", "despise", "loathe", "can't stand"; dehumanising language, comparing targets to animals, objects, or diseases; generalising terms: "all of them", "these people", "they always" • Contextual: Group-based attributions - directing hostility at categories of people; describing targets as inherently flawed; eliminationist rhetoric - Wishing for removal, silencing, or harm Some guiding questions to identify these posts can be: • Does the post express hate towards political groups or ideologies? • Does the post contain explicit hate vocabulary directed at groups rather than situations? • Is dehumanising language or metaphor used to describe the targets? • Does the content generalise negative characteristics to entire categories of people? • Is there expression of desire for the targets to be harmed, silenced, or eliminated? • Does the author present the targets as inherently and unchangeably flawed?
Emotional neutrality	YES: 1 NO: 0	A communication state characterised by the absence of distinct positive or negative emotional markers, typified by objective presentation of information, factual reporting, or deliberate affective restraint in social media content. Detection Cues: • Lexical: Factual reporting, passive voice, hedges ("maybe", "possibly"), domain-specific vocabulary

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

		without emotional loading, absence of intensifiers, superlatives, or emotion-laden adjectives. • Contextual: News sharing without commentary, procedural descriptions Some guiding questions to identify these posts can be: • Does the post primarily present factual information without evaluative language? • Is the content written in an objective tone without emotional qualifiers? • Is there use of hedging or neutral reporting language? • Is the post free from intensifiers, superlatives, and emotion-signalling punctuation? • Does the content maintain a balanced perspective without expressing personal approval or disapproval?

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

ANNEX B – LMM TRAINING METRICS

POLITICS BINARY CLASSIFICATION

Table 3 General metric LLM politics binary classification

Metric	Value
Accuracy	0.907846
F1	0.939788
Precision	0.931452
Recall	0.948276
F1 Macro	0.871751
AUC-ROC	0.933384

Table 4 Per-language metric LLM politics binary classification

Language	Accuracy	F1	Precision	Recall	AUC_ROC
SQ	0.8775	0.903353	0.898039	0.90873	0.931896
DE	0.959799	0.978495	0.968085	0.98913	0.936232
DA	0.890547	0.938202	0.922652	0.954286	0.838242
PL	0.955882	0.976	0.963158	0.989189	0.887909
BG	0.911765	0.932836	0.919118	0.94697	0.951284
MK	0.88	0.92053	0.932886	0.908497	0.917258
SR	0.909091	0.935252	0.921986	0.948905	0.939691

EMOTIONS MULTI-LABEL CLASSIFICATION

Table 5 General metric LLM emotions multi-label classification

Metric	Value
F1 Macro	0.971777
F1 Micro	0.978355
Precision Macro	0.971292
Recall Macro	0.97242
AUC-ROC Macro	0.997817

Table 6 Per language metric LLM emotions multi-label classification

Language	F1_Macro	F1_Micro	Precision_Macro	Recall_Macro	AUC_ROC_Macro
PL	0.967117	0.972047	0.962457	0.972162	0.997909
BG	0.98204	0.985137	0.987599	0.976666	0.998352
SR	0.966082	0.975904	0.971685	0.961127	0.996696
DA	0.964928	0.981331	0.972601	0.957785	0.99874
SQ	0.935966	0.956946	0.970156	0.90734	0.973718
DE	0.969008	0.9812	0.956245	0.982621	0.999206
MK	0.973021	0.982433	0.963315	0.983921	0.999661

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Table 7 Per emotion metrics LLM emotions multi-label classification

Emotion	F1	Precision	Recall	AUC-ROC
Fear or anxiety	0.959091	0.947605	0.970859	0.996129
Happiness or enthusiasm	0.979857	0.986937	0.972877	0.998072
Anger	0.984385	0.996524	0.972539	0.999307
Hate	0.956628	0.951988	0.961312	0.996953
Emotional neutrality	0.978927	0.973404	0.984513	0.998623

VALUES MULTI-LABEL CLASSIFICATION

Table 8 General metric LLM values multi-label classification

Metric	Value
F1 Macro	0.991557
F1 Micro	0.991846
Precision Macro	0.992853
Recall Macro	0.9903
AUC-ROC Macro	0.998512

Table 9 Per language metric LLM values multi-label classification

Language	F1_Macro	F1_Micro	Precision_Macro	Recall_Macro	AUC_ROC_Macro
SQ	0.914849	0.93095	0.944422	0.889598	0.970454
DA	0.988545	0.99447	0.992596	0.984698	0.998485
MK	0.997377	0.998174	0.99628	0.998521	1
DE	0.992393	0.989899	0.985397	0.999648	0.999989
PL	0.991658	0.99548	0.996226	0.987321	0.999415
SR	0.996052	0.997295	0.997401	0.994757	0.999778
BG	0.997071	0.997042	0.99892	0.995277	0.999993

Table 10 Per value metric LLM values multi-label classification

Values	F1	Precision	Recall	AUC-ROC
Benevolence	0.991615	0.996737	0.986545	0.997336
Security	0.990358	0.982855	0.997976	0.999265
Universalism	0.990038	0.991939	0.988145	0.999133
Tradition	0.99159	0.996326	0.9869	0.997331
Self-Direction	0.991935	0.993757	0.990121	0.998989
Power	0.993805	0.995501	0.992115	0.999017

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

ANNEX C – ANNOTATED SAMPLE DISTRIBUTION

Table 11 Distribution of the annotated dataset per country and emotion/value - absolute figures

country_dataset	category	fear_anxiety	happiness_enthusiasm	anger	hate	emotional_neutrality	benevolence	security	universalism	tradition	self direction	power
AT	comments	7540	24862	220607	28926	8079	43816	37681	15599	11932	12488	168498
AT	media	3944	1972	6141	267	32725	3409	12747	3779	569	2566	21979
AT	pol	3153	11604	19071	687	5803	3726	11146	5973	948	2879	15646
AT	posts	801	2298	8935	583	1644	1243	2608	1317	282	1129	7682
BA	comments	11333	42190	178571	13934	19982	18919	85411	4509	11793	93128	52250
BA	media	15041	7229	10526	43	43761	3596	36009	2078	806	15110	19001
BA	pol	5245	11518	10771	59	7995	5370	11038	893	1154	6597	10536
BG	comments	1145	1938	22513	5061	6502	2913	5129	3822	1831	3956	19508
BG	media	1233	527	1582	27	10102	773	2971	1662	226	2805	5034
BG	pol	1025	1786	3551	64	4494	1502	2383	1445	738	1669	3183
BG	posts	182	299	810	76	4543	589	731	763	165	734	2928
DK	comments	26611	25401	107010	32263	40973	4565	44941	104848	43839	17555	16510
DK	media	1127	416	636	351	13657	329	6599	6253	665	643	1698
DK	pol	5212	9798	8797	2255	6824	1324	10602	16713	1700	949	1598
DK	posts	2642	4322	5362	1248	4361	410	6619	8159	1053	579	1115
MK	comments	1405	5483	38810	4886	7485	5369	7614	3771	14816	4029	22470
MK	media	3726	1959	6143	44	28992	3087	10054	3545	2298	2400	19480
MK	pol	567	5910	2270	68	3252	2773	2922	1646	1159	590	2977

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

MK_AL	media	1554	590	665	19	6141	0	0	0	0	0	0
PL	comments	34705	49816	326097	119814	42219	94856	184578	44782	24291	42078	182066
PL	media	19082	6833	18512	790	51428	3585	52060	3786	1599	7195	28420
PL	pol	37002	56788	60321	2760	59003	9410	113806	13173	5600	8334	65551
PL	posts	4887	7552	10865	989	5854	1012	13684	1931	494	1859	11167

Table 12 Distribution of the annotated dataset per country and emotion/value - percentages (%)

country_dataset	category	fear_anxiety	happiness_enthusiasm	anger	hate	emotional_neutrality	benevolence	security	universalism	tradition	self direction	power
AT	comments	48.84	61.03	86.60	94.95	16.74	83.95	58.71	58.49	86.90	65.51	78.81
AT	media	25.55	4.84	2.41	0.88	67.82	6.53	19.86	14.17	4.14	13.46	10.28
AT	pol	20.42	28.49	7.49	2.26	12.03	7.14	17.37	22.40	6.90	15.10	7.32
AT	posts	5.19	5.64	3.51	1.91	3.41	2.38	4.06	4.94	2.05	5.92	3.59
BA	comments	35.84	69.24	89.34	99.27	27.85	67.85	64.48	60.28	85.75	81.10	63.89
BA	media	47.57	11.86	5.27	0.31	61.00	12.90	27.19	27.78	5.86	13.16	23.23
BA	pol	16.59	18.90	5.39	0.42	11.14	19.26	8.33	11.94	8.39	5.74	12.88
BG	comments	31.94	42.59	79.12	96.81	25.36	50.42	45.74	49.69	61.86	43.17	63.64
BG	media	34.39	11.58	5.56	0.52	39.40	13.38	26.49	21.61	7.64	30.61	16.42
BG	pol	28.59	39.25	12.48	1.22	17.53	26.00	21.25	18.79	24.93	18.21	10.38
BG	posts	5.08	6.57	2.85	1.45	17.72	10.20	6.52	9.92	5.57	8.01	9.55
DK	comments	74.77	63.60	87.85	89.33	62.25	68.87	65.36	77.11	92.77	88.99	78.92
DK	media	3.17	1.04	0.52	0.97	20.75	4.96	9.60	4.60	1.41	3.26	8.12
DK	pol	14.64	24.53	7.22	6.24	10.37	19.98	15.42	12.29	3.60	4.81	7.64

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

DK	posts	7.42	10.82	4.40	3.46	6.63	6.19	9.63	6.00	2.23	2.94	5.33
MK	comments	24.66	41.07	82.18	97.76	18.84	47.81	36.98	42.08	81.08	57.40	50.01
MK	media	65.39	14.67	13.01	0.88	72.97	27.49	48.83	39.56	12.58	34.19	43.36
MK	pol	9.95	44.26	4.81	1.36	8.19	24.69	14.19	18.37	6.34	8.41	6.63
MK_AL	media	100.00	100.00	100.00	100.00	100.00						
PL	comments	36.27	41.17	78.43	96.35	26.64	87.13	50.69	70.33	75.95	70.76	63.39
PL	media	19.94	5.65	4.45	0.64	32.45	3.29	14.30	5.95	5.00	12.10	9.90
PL	pol	38.67	46.94	14.51	2.22	37.22	8.64	31.25	20.69	17.51	14.01	22.82
PL	posts	5.11	6.24	2.61	0.80	3.69	0.93	3.76	3.03	1.54	3.13	3.89

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

ANNEX D – ANALYSIS RESULTS

Table 13 Chi-Square Test Results for posts

Emotion	Value	Chi2	p-value	φ (phi coefficient)
fear_anxiety	benevolence	4415.073	0.0000	0.079
fear_anxiety	security	35437.466	0.0000	0.224
fear_anxiety	universalism	1369.721	0.0000	0.044
fear_anxiety	tradition	745.650	0.0000	0.033
fear_anxiety	self direction	3188.781	0.0000	0.067
fear_anxiety	power	9952.834	0.0000	0.119
happiness_enthusiasm	benevolence	14273.818	0.0000	0.142
happiness_enthusiasm	security	11528.201	0.0000	0.128
happiness_enthusiasm	universalism	2863.204	0.0000	0.064
happiness_enthusiasm	tradition	1536.246	0.0000	0.047
happiness_enthusiasm	self direction	29.880	0.0000	0.007
happiness_enthusiasm	power	0.913	0.3394	0.001
anger	benevolence	1431.423	0.0000	0.045
anger	security	366.110	0.0000	0.023
anger	universalism	564.294	0.0000	0.028
anger	tradition	0.376	0.5397	0.001
anger	self direction	707.211	0.0000	0.032
anger	power	926.421	0.0000	0.036
hate	benevolence	0.845	0.3580	0.001
hate	security	281.746	0.0000	0.020
hate	universalism	2154.535	0.0000	0.055
hate	tradition	504.175	0.0000	0.027
hate	self direction	141.516	0.0000	0.014
hate	power	200.705	0.0000	0.017
emotional neutrality	benevolence	180.259	0.0000	0.016
emotional_neutrality	security	4119.865	0.0000	0.076
emotional_neutrality	universalism	36.270	0.0000	0.007
emotional_neutrality	tradition	262.465	0.0000	0.019
emotional_neutrality	self direction	3960.632	0.0000	0.075
emotional_neutrality	power	2333.878	0.0000	0.058

Table 14 Chi-Square Test Results for comments

Emotion	Value	Chi²	p-value	Phi
fear_anxiety	benevolence	5770.26	0.0	0.063
fear_anxiety	security	22112.421	0.0	0.123
fear_anxiety	universalism	1819.195	0.0	0.035
fear_anxiety	tradition	210.201	0.0	0.012
fear_anxiety	self direction	2707.183	0.0	0.043
fear_anxiety	power	5135.794	0.0	0.059
happiness_enthusiasm	benevolence	2301.471	0.0	0.04
happiness_enthusiasm	security	6862.067	0.0	0.069
happiness_enthusiasm	universalism	32.219	0.0	0.005
happiness_enthusiasm	tradition	630.115	0.0	0.021
happiness_enthusiasm	self direction	2745.267	0.0	0.043

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

happiness_enthusiasm	power	111.995	0.0	0.009
anger	benevolence	1505.483	0.0	0.032
anger	security	114.887	0.0	0.009
anger	universalism	11741.389	0.0	0.09
anger	tradition	1668.946	0.0	0.034
anger	self direction	645.881	0.0	0.021
anger	power	9676.183	0.0	0.082
hate	benevolence	16446.59	0.0	0.106
hate	security	140.131	0.0	0.01
hate	universalism	719.916	0.0	0.022
hate	tradition	255.169	0.0	0.013
hate	self direction	3681.066	0.0	0.05
hate	power	3999.707	0.0	0.052
emotional_neutrality	benevolence	6545.169	0.0	0.067
emotional_neutrality	security	1367.695	0.0	0.031
emotional_neutrality	universalism	12887.533	0.0	0.094
emotional_neutrality	tradition	1289.591	0.0	0.03
emotional_neutrality	self direction	299.676	0.0	0.014
emotional_neutrality	power	473.124	0.0	0.018

Table 15 Category regression results per emotion

emotion	term	coef	ci_low	ci_high	pval
fear_anxiety	C(category, Treatment(reference="posts"))[T.comments]	-0.067892514	-0.07003184	-0.065753187	0
fear_anxiety	C(category, Treatment(reference="posts"))[T.media]	0.028777467	0.026459315	0.031095619	9.6074E-131
fear_anxiety	C(category, Treatment(reference="posts"))[T.pol]	0.025448757	0.023161817	0.027735698	1.9051E-105
happiness_enthusiasm	C(category, Treatment(reference="posts"))[T.comments]	-0.109222286	-0.11174337	-0.106701202	0
happiness_enthusiasm	C(category, Treatment(reference="posts"))[T.media]	-0.146449187	-0.149181008	-0.143717366	0
happiness_enthusiasm	C(category, Treatment(reference="posts"))[T.pol]	0.068155938	0.065460898	0.070850978	0
anger	C(category, Treatment(reference="posts"))[T.comments]	0.233148502	0.229567846	0.236729158	0
anger	C(category, Treatment(reference="posts"))[T.media]	-0.232079372	-0.235959334	-0.22819941	0
anger	C(category, Treatment(reference="posts"))[T.pol]	-0.079130041	-0.082957763	-0.075302319	0
hate	C(category, Treatment(reference="posts"))[T.comments]	0.098271105	0.096022787	0.100519424	0
hate	C(category, Treatment(reference="posts"))[T.media]	-0.037255495	-0.03969175	-0.03481924	2.4954E-197
hate	C(category, Treatment(reference="posts"))[T.pol]	-0.025479559	-0.027883012	-0.023076106	6.98856E-96
emotional_neutrality	C(category, Treatment(reference="posts"))[T.comments]	-0.154304808	-0.156971901	-0.151637715	0
emotional_neutrality	C(category, Treatment(reference="posts"))[T.media]	0.387006586	0.384116552	0.389896621	0
emotional_neutrality	C(category, Treatment(reference="posts"))[T.pol]	0.011004905	0.008153782	0.013856028	3.87509E-14

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Table 16 Electoral period regression results per emotion for posts

emotion	period	coef	ci_low	ci_high	pval
fear_anxiety	C(period)[T.national_election]	-0.00573	-0.00945	-0.00201	0.002529
fear_anxiety	C(period)[T.eu_election]	0.002566	-0.00143	0.006564	0.208348
happiness_enthusiasm	C(period)[T.national_election]	0.048963	0.044528	0.053399	1.1E-103
happiness_enthusiasm	C(period)[T.eu_election]	0.037875	0.033107	0.042643	1.28E-54
anger	C(period)[T.national_election]	-0.01322	-0.01779	-0.00864	1.47E-08
anger	C(period)[T.eu_election]	-0.01591	-0.02082	-0.01099	2.25E-10
hate	C(period)[T.national_election]	0.00238	0.001081	0.003678	0.000329
hate	C(period)[T.eu_election]	-0.00113	-0.00252	0.00027	0.113991
emotional_neutrality	C(period)[T.national_election]	-0.0324	-0.03757	-0.02722	1.43E-34
emotional_neutrality	C(period)[T.eu_election]	-0.02341	-0.02897	-0.01784	1.67E-16

Table 17 Electoral period regression results per emotion for comments

emotion	period	coef	ci_low	ci_high	pval
fear_anxiety	C(period)[T.national_election]	-0.00609	-0.00717	-0.00501	3.16E-28
fear_anxiety	C(period)[T.eu_election]	0.002102	0.001023	0.00318	0.000133
happiness_enthusiasm	C(period)[T.national_election]	0.027303	0.025954	0.028652	0
happiness_enthusiasm	C(period)[T.eu_election]	-0.00318	-0.00452	-0.00184	3.51E-06
anger	C(period)[T.national_election]	0.014265	0.012028	0.016502	7.61E-36
anger	C(period)[T.eu_election]	0.019795	0.017569	0.022022	5.24E-68
hate	C(period)[T.national_election]	-0.02423	-0.02588	-0.02257	5.2E-181
hate	C(period)[T.eu_election]	-0.00989	-0.01154	-0.00825	5.39E-32
emotional_neutrality	C(period)[T.national_election]	-0.01125	-0.01253	-0.00997	1.67E-66
emotional_neutrality	C(period)[T.eu_election]	-0.00883	-0.0101	-0.00755	5.48E-42

Table 18 OLS Regression Results Model Summary and Regression Coefficients for interacted emotion x category

Statistic	Value
Dep. Variable	engagement_log
R-squared	0.584
Adj. R-squared	0.584
Method	Least Squares
F-statistic	163.1
Prob (F-statistic)	0.00
Date	Fri, 29 Aug 2025
Time	10:14:24
No. Observations	2,169,852
Df Residuals	2,169,832
Df Model	19
Covariance Type	cluster

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Variable	Coefficient	Std. Error	z-value	P>|z|	[0.025, 0.975]
const	1.5097	0.249	6.068	0.000	[1.022, 1.997]
category_comments	-0.8379	0.249	-3.364	0.001	[-1.326, -0.350]
category_media	-0.2667	0.381	-0.700	0.484	[-1.014, 0.480]
category_pol	2.7369	0.289	9.474	0.000	[2.171, 3.303]
fear_anxiety	0.3658	0.248	1.477	0.140	[-0.120, 0.851]
happiness_enthusiasm	0.8477	0.248	3.416	0.001	[0.361, 1.334]
anger	0.2241	0.249	0.902	0.367	[-0.263, 0.711]
hate	0.1049	0.254	0.413	0.680	[-0.393, 0.603]
category_comments_x_fear_anxiety	-0.3154	0.248	-1.272	0.203	[-0.801, 0.170]
category_comments_x_happiness_enthusiasm	-0.8677	0.249	-3.489	0.000	[-1.355, -0.380]
category_comments_x_anger	-0.2026	0.249	-0.814	0.415	[-0.690, 0.285]
category_comments_x_hate	-0.1859	0.254	-0.731	0.465	[-0.684, 0.312]
category_media_x_fear_anxiety	0.0803	0.381	0.211	0.833	[-0.666, 0.826]
category_media_x_happiness_enthusiasm	-0.5483	0.347	-1.580	0.114	[-1.228, 0.132]
category_media_x_anger	0.5328	0.449	1.186	0.235	[-0.347, 1.413]
category_media_x_hate	1.0391	0.446	2.332	0.020	[0.166, 1.912]
category_pol_x_fear_anxiety	0.1517	0.262	0.580	0.562	[-0.361, 0.665]
category_pol_x_happiness_enthusiasm	-0.5881	0.267	-2.206	0.027	[-1.111, -0.066]
category_pol_x_anger	0.4388	0.262	1.677	0.093	[-0.074, 0.952]
category_pol_x_hate	0.5868	0.282	2.081	0.037	[0.034, 1.139]

Diagnostic	Value
Omnibus	348001.780
Prob(Omnibus)	0.000
Jarque-Bera (JB)	803883.602
Skew	0.929
Prob(JB)	0.000
Kurtosis	5.332
Durbin-Watson	1.121
Cond. No.	101

Notes:
[1] Standard Errors are robust to cluster correlation (cluster)

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Table 19 Main Effects vs Baseline (posts + emotional_neutrality)

Variable	Beta	% Effect	CI Low (%)	CI High (%)
const	1.510	352.538	177.887	636.955
category_comments	-0.838	-56.736	-73.445	-29.513
category_media	-0.267	-23.406	-63.716	61.684
category_pol	2.737	1443.950	776.437	2619.857
fear_anxiety	0.366	44.169	-11.266	134.234
happiness_enthusiasm	0.848	133.419	43.512	279.651
anger	0.224	25.119	-23.134	103.664
hate	0.105	11.056	-32.483	82.672

Table 20 All Category × Emotion Combinations vs Baseline

Category	Emotion	Beta	% Effect	Combination Type
comments	anger	-0.816	-55.796	full_interaction
comments	emotional_neutrality	-0.838	-56.736	category_main
comments	fear_anxiety	-0.787	-54.498	full_interaction
comments	happiness_enthusiasm	-0.858	-57.592	full_interaction
comments	hate	-0.919	-60.102	full_interaction
media	anger	0.490	63.270	full_interaction
media	emotional_neutrality	-0.267	-23.406	category_main
media	fear_anxiety	0.179	19.653	full_interaction
media	happiness_enthusiasm	0.033	3.324	full_interaction
media	hate	0.877	140.437	full_interaction
pol	anger	3.400	2895.954	full_interaction
pol	emotional_neutrality	2.737	1443.950	category_main
pol	fear_anxiety	3.254	2490.419	full_interaction
pol	happiness_enthusiasm	2.997	1901.571	full_interaction
pol	hate	3.429	2983.250	full_interaction
posts	anger	0.224	25.119	emotion_main
posts	emotional_neutrality	0.000	0.000	baseline
posts	fear_anxiety	0.366	44.169	emotion_main
posts	happiness_enthusiasm	0.848	133.419	emotion_main
posts	hate	0.105	11.056	emotion_main

Table 21 OLS Regression Results Model Summary and Regression Coefficients for interacted emotion x period

Dependent Variable	engagement_log
R-squared	0.041
Adjusted R-squared	0.041
No. Observations	1,419,944
Df Residuals	1,419,929
Df Model	14

Variable	Coefficient	Std. Error	z	P>|z|	[0.025, 0.975]
const	1.3531	0.007	186.865	0.000	[1.339, 1.367]

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

period_national_election	0.0650	0.010	6.489	0.000	[0.045, 0.085]
period_eu_election	-0.0577	0.010	-5.870	0.000	[-0.077, -0.038]
fear_anxiety	0.1449	0.013	10.812	0.000	[0.119, 0.171]
happiness_enthusiasm	0.3197	0.012	26.067	0.000	[0.296, 0.344]
anger	-0.4572	0.008	-58.295	0.000	[-0.473, -0.442]
hate	-0.7406	0.008	-88.777	0.000	[-0.757, -0.724]
period_national_election_x_fear_anxiety	-0.1033	0.018	-5.667	0.000	[-0.139, -0.068]
period_national_election_x_happiness_enthusiasm	-0.2773	0.016	-17.391	0.000	[-0.309, -0.246]
period_national_election_x_anger	-0.0779	0.011	-7.263	0.000	[-0.099, -0.057]
period_national_election_x_hate	-0.0389	0.011	-3.387	0.001	[-0.061, -0.016]
period_eu_election_x_fear_anxiety	-0.0655	0.018	-3.716	0.000	[-0.100, -0.031]
period_eu_election_x_happiness_enthusiasm	-0.0454	0.016	-2.782	0.005	[-0.077, -0.013]
period_eu_election_x_anger	0.0351	0.011	3.333	0.001	[0.014, 0.056]
period_eu_election_x_hate	0.0914	0.011	8.134	0.000	[0.069, 0.113]

Log-Likelihood	-2.5378e+06
AIC	5.076e+06
BIC	5.076e+06
Omnibus	554298.203
Durbin-Watson	0.788

Standard Errors are heteroscedasticity robust (HC1).

Table 22 Main Effects vs Non-Election + Emotional Neutrality Baseline

Variable	Beta	Pct Effect %	CI Low %	CI High %	P-value
const	1.353	286.940	281.487	292.471	0.0
period_national_election	0.065	6.712	4.639	8.827	0.0
period_eu_election	-0.058	-5.608	-7.410	-3.771	0.0
fear_anxiety	0.145	15.598	12.600	18.676	0.0
happiness_enthusiasm	0.320	37.673	34.403	41.022	0.0
anger	-0.457	-36.696	-37.661	-35.715	0.0
hate	-0.741	-52.319	-53.092	-51.532	0.0

Table 23 All Period × Emotion Combinations vs Baseline

Period	Emotion	Beta	Pct Effect %	Combination Type
eu_election	anger	-0.480	-38.110	full_interaction
eu_election	emotional_neutrality	-0.058	-5.608	period_main
eu_election	fear_anxiety	0.022	2.202	full_interaction
eu_election	happiness_enthusiasm	0.217	24.180	full_interaction
eu_election	hate	-0.707	-50.683	full_interaction

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

national_election	anger	-0.470	-37.510	full_interaction
national_election	emotional_neutrality	0.065	6.712	period_main
national_election	fear_anxiety	0.107	11.247	full_interaction
national_election	happiness_enthusiasm	0.107	11.334	full_interaction
national_election	hate	-0.715	-51.061	full_interaction
non_election	anger	-0.457	-36.696	emotion_main
non_election	emotional_neutrality	0.000	0.000	baseline
non_election	fear_anxiety	0.145	15.598	emotion_main
non_election	happiness_enthusiasm	0.320	37.673	emotion_main
non_election	hate	-0.741	-52.319	emotion_main

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

ANNEX E – ETHICAL REVIEW

The project has received a Research Ethics Screening Confirmation from the Institutional Review Board (IRB) of the Faculty of Social Sciences at the University of Vienna, based on the application submitted for the research project "ENCODE WP3 Analysing Social Media Communication." This confirmation classifies the project as presenting minimal ethical risk, according to the information provided by the applicant and the mitigation strategies described in the project documentation.

It is important to note that this confirmation does not constitute formal ethics approval by the University of Vienna Ethics Committee. The IRB confirmed that formal approval by the ethics committee is not required for this type of research.

The initial application referenced TikTok as it was the first access application submitted by the consortium; however, such data was not finally included due to the delay in the access approval. X data is the subject of analysis of this deliverable. The methodology, as described in the appendix of the IRB approval, involves the automated and manual analysis of publicly available social media content, with strict anonymisation of user data and adherence to ethical and data protection guidelines. The change of social media platform does not affect the methodology or data use in a way that would introduce new ethical concerns, as the same principles and safeguards apply.

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Confirmation-1280-Jörg Matthes
Department of Communication

universität wien
Faculty of Social Sciences

Research Ethics Screening Confirmation

The Institutional Review Board of the Faculty of Social Sciences (IRB) at the University of Vienna confirms that the Research Ethics Screening, filled out by Jörg Matthes for the research project "ENCODE WP3 Analysing Social Media Communication", has classified the project in the category of minimal ethical risk. This self-assessment is based on the information provided by the applicant as enclosed in the appendix. The IRB checked the coherence of the information on potential ethical problems or challenges with the overall project description provided by the applicant and concludes that the applicant has proposed sufficient mitigation strategies for ethical challenges in the project to realise it on the researcher`s own responsibility.

This confirmation does not constitute formal ethics approval by the University of Vienna Ethics Committee, which is not required for this kind of research according to the Statutes of the University of Vienna and the Austrian 2002 Universities Act. If any doubts or ethical challenges occur during the further planning and/or implementation of the research project, or if circumstances change that could raise ethical concerns, please contact the IRB.

- [X] I confirm that I filled out the Research Ethics Screening to the best of my knowledge and that I am aware that this confirmation is only valid together with the information provided as enclosed in the appendix.
- [X] I confirm that I will take all measures necessary to identify and address ethical issues in all stages of my research project.
- [X] I am aware that I have to re-run the Research Ethics Screening if there are changes regarding the research design or the use of data that could raise ethical issues.

Researcher

On behalf of the Institutional Review Board of the Faculty of Social Sciences

25.11.2024 16:02 1

Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

ANNEX F – MODEL CARDS

Model Card — Human Values Classification

(Multilabel)
Model name: ENCODE - Human Values Classifier (XLM-RoBERTa-base)
Document version: v2.0 (22/10/2025)
Owners / Authors: Rodrigo Ortega Izquierdo, Frans Folkvord (Lead Partner: PBY) within the ENCODE consortium.
1. Overview & Objective
Multilabel classifier for six **Schwartz** values: **benevolence**, **security**, **universalism**, **tradition**, **self-direction**, **power**. Applied **only** to texts previously labelled as political.
2. Intended Use (and Out-of-Scope)
Intended: emotion-value correlation studies, actor comparisons, period analyses; inputs to D3.4 (Catalogue of Best Practices), WP6 and WP7.
Not intended: profiling individuals, normative decisions without human review, inferences outside the political domain.
3. Data & Labelling
Size after preprocessing (values): **11,909** entries.
Codebook with operational definitions and detection cues for each value (Annex A).
4. Pipeline & Architecture
Base: XLM-RoBERTa-base (multilabel head with 6 outputs).
Tokenizer: max **256 tokens**.
Split: **MultilabelStratifiedKFold (5 folds)** by language & label distribution; first fold used for train/test.
Hyperparameters: lr=1e-5, batch_size=32, epochs=3; selection by **F1-macro**.
Implementation: HuggingFace Trainer with custom metrics and early stopping.
5. Evaluation (Test)
Global:
F1-macro **0.991557**
F1-micro **0.991846**
Precision-macro **0.992853**
Recall-macro **0.990300**
AUC-ROC-macro **0.998512**
Per value:
Benevolence: F1 **0.991615**, Precision **0.996737**, Recall **0.986545**, AUC **0.997336**
Security: F1 **0.990358**, Precision **0.982855**, Recall **0.997976**, AUC **0.999265**
Universalism: F1 **0.990038**, Precision **0.991939**, Recall **0.988145**, AUC **0.999133**
Tradition: F1 **0.991590**, Precision **0.996326**, Recall **0.986900**, AUC **0.997331**
Self-direction: F1 **0.991935**, Precision **0.993757**, Recall **0.990121**, AUC **0.998989**
Power: F1 **0.993805**, Precision **0.995501**, Recall **0.992115**, AUC **0.999017**
By language (examples):
BG F1-macro **0.997071**, SR **0.996052**, MK **0.997377**, DE **0.992393**, PL **0.991658**, DA **0.988545**, SQ **0.914849**.
6. Analytical Context (for Interpreting Outputs)
Chi-square emotion-value matrix shows robust associations (e.g., **fear/anxiety ↔ security** **φ=0.224**; **happiness/enthusiasm ↔ benevolence** **φ=0.142** in posts), typically small-medium effects; useful for interpreting model outputs.
7. Risks & Limitations

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Semantic overlap between cues (e.g., security vs. power); use thresholds and human review in sensitive cases.
Comment context: methodology requires reading comments together with their parent post.
8. I/O
Input: political text in {DE, DA, PL, BG, MK, SR, SQ}, ≤256 tokens.
Output: vector of six probabilities (pϵ[0,1]) and active labels (configurable threshold).
9. Governance & Rights
Tech stack: Python 3.12.4; pandas/NumPy; HuggingFace Transformers; scikit-learn; statsmodels; seaborn/matplotlib.
Data governance: flat CSV; duplicate removal by message ID; preserved fields (id, timestamp, author_id/username, language, text, engagement metrics, country/country_code when available, reply/long-post flags).
Dissemination & rights: deliverable marked PU–Public; consortium ownership and restrictions on copying/distribution without prior agreement.
10. Source
ENCODE Deliverable: *D3.3 Sentiment analysis*.
Usage
Multilabel inference for Schwartz values with configurable thresholds and batched processing.

```
from transformers import AutoTokenizer, AutoModelForSequenceClassification,
TextClassificationPipeline
import torch, pandas as pd
model_dir = "models/encode_values_xlmr" # <- change to your path
THRESHOLD = 0.50
labels = [
"benevolence",
"security",
"universalism",
"tradition",
"self_direction",
"power",
]
device = 0 if torch.cuda.is_available() else -1
# Load
tokenizer = AutoTokenizer.from_pretrained(model_dir)
model = AutoModelForSequenceClassification.from_pretrained(model_dir)
clf = TextClassificationPipeline(model=model, tokenizer=tokenizer,
device=device, return_all_scores=True, top_k=None)
# Inference
texts = [
"We must protect our borders and keep families safe.",
"Everyone deserves equal rights and a clean environment.",
]
raw = clf(texts, batch_size=32, truncation=True, padding=True)
# Threshold into multilabel sets
active = []
for row in raw:
scores = {d['label']: d['score'] for d in row}
on = [lab for lab in labels if scores.get(lab, 0.0) >= THRESHOLD or
scores.get(f"LABEL_{labels.index(lab)}", 0.0) >= THRESHOLD]
```


This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

active.append(on)
for t, labs in zip(texts, active):
print("TEXT:", t)
print("ACTIVE VALUES:", labs)
Chaining pattern
1. Run **Political** → keep only label=='political'.
2. Run **Emotions** and **Values** in parallel on the filtered set.
3. Persist probabilities along with labels for downstream statistical analysis (e.g., emotion–value chi-square, engagement models).

Model Card — Political Content Detection

(Binary)
Model name: ENCODE - Political Content Classifier (XLM-RoBERTa-base)
Document version: v2.0 (22/10/2025)
Owners / Authors: Rodrigo Ortega Izquierdo, Frans Folkvord (Lead Partner: PBY) within the ENCODE consortium.
1. Overview & Objective
Binary classifier that labels X/Twitter texts as political (1) or non_political (0). It serves as the **first-stage filter**: only texts labelled political proceed to the emotion and values models.
2. Intended Use (and Out-of-Scope)
Intended: academic research and analysis of political discourse on social media within ENCODE (WP3).
Not intended: automated moderation or sanctioning, or rights-impacting decisions without human review.
3. Data & Labelling
Source & period: X API v2, 2022-2024, six countries (AT, BA, BG, DK, MK, PL) and four categories (general public, politicians, media, comments to politicians). **Total corpus**: 2,169,852 posts/comments.
Sampling & ground truth: stratified manual annotation; 20% double-coded; minimum **Krippendorff's α = 0.67** before scaling.
Size after preprocessing (politics task): **8,030** entries.
Languages covered: DE, DA, PL, BG, MK, SR, SQ.
4. Preprocessing & Pipeline
Column harmonization, removal of empty texts/duplicates, strict 0/1 label filtering, language filtering to the seven target languages.
5. Architecture & Training
Base: XLM-RoBERTa-base with a 2-class sequence classification head.
Tokenizer: max **512 tokens**; padding & truncation.
Split: 80/20, stratified by language and politics tag.
Hyperparameters: lr=1e-5, batch_size=32, epochs=3; best model selected by **positive-class F1**.
Implementation: HuggingFace Transformers (Trainer + callbacks).
6. Evaluation (Test)
Overall:
Accuracy **0.907846**
F1 **0.939788**
Precision **0.931452**
Recall **0.948276**
F1-macro **0.871751**
AUC-ROC **0.933384**
By language (examples):

Funded by the European Union
This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

DE F1 **0.978495**, PL **0.976000**, DA **0.938202**, BG **0.932836**, MK **0.920530**, SR **0.935252**, SQ **0.903353**.
7. Risks & Potential Biases
Uneven coverage due to X metadata (e.g., limited general-public data for MK and BA).
Domain shift risk around events/elections/platform changes.
8. Limitations
Context is crucial for comments (classification may require considering the parent post).
9. I/O
Input: text in {DE, DA, PL, BG, MK, SR, SQ}, ≤512 tokens.
Output: label ∈ {0,1}, score ∈ [0,1].
10. Governance & Rights
Tech stack: Python 3.12.4; pandas/NumPy; HuggingFace Transformers; scikit-learn; statsmodels; seaborn/matplotlib for visualization.
Data governance: flat CSV; duplicate removal by message ID; preserved fields (id, timestamp, author_id/username, language, text, engagement metrics, country/country_code when available, reply/long-post flags).
Dissemination & rights: deliverable marked **PU–Public**; consortium ownership and restrictions on copying/distribution without prior agreement.
11. Source
ENCODE Deliverable: *D3.3 Sentiment analysis*.
Usage
Below is a minimal example for **inference** with the political content classifier using HuggingFace Transformers. Replace model_dir with your local checkpoint path.

```
from transformers import AutoTokenizer, AutoModelForSequenceClassification,
TextClassificationPipeline
import torch, pandas as pd
# 1) Load local fine-tuned model
model_dir = "models/encode_politics_xlmr" # <- change to your path
device = 0 if torch.cuda.is_available() else -1
tokenizer = AutoTokenizer.from_pretrained(model_dir)
model = AutoModelForSequenceClassification.from_pretrained(model_dir)
clf = TextClassificationPipeline(model=model, tokenizer=tokenizer,
device=device, return_all_scores=True, top_k=None)
# 2) Single example
text = "Parliament passed a new climate law today."
result = clf(text)[0] # list of dicts: [{label: ..., score: ...}, ...]
# Labels may be saved as 'political'/'non_political' or 'LABEL_1'/'LABEL_0'.
score_map = {d['label']: d['score'] for d in result}
political_score = score_map.get('political', score_map.get('LABEL_1'))
non_political_score = score_map.get('non_political', score_map.get('LABEL_0'))
label = 'political' if political_score >= non_political_score else 'non_political'
print(label, political_score)
# 3) Batch inference (e.g., a pandas Series)
df = pd.DataFrame({"text": [
"President addressed the nation about inflation.",
"Had a great coffee this morning!",
]})
outputs = clf(df["text"].tolist(), batch_size=32, truncation=True)
# Convert to label + score
preds = []
for out in outputs:
```


This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

m = {d['label']: d['score'] for d in out}
pol = m.get('political', m.get('LABEL_1'))
non = m.get('non_political', m.get('LABEL_0'))
preds.append({"label": 'political' if pol >= non else 'non_political',
"score": max(pol, non)})
df = df.join(pd.DataFrame(preds))
print(df)
Chaining tip: run this model first and only forward texts with label=='political' to the emotion/values models.
Model Card – Emotion Classification
(Multilabel)
Model name: ENCODE - Emotion Classifier (XLM-RoBERTa-base)
Document version: v2.0 (22/10/2025)
Owners / Authors: Rodrigo Ortega Izquierdo, Frans Folkvord (Lead Partner: PBY) within the ENCODE consortium.
1. Overview & Objective
Multilabel classifier for five dimensions: **fear/anxiety**, **happiness/enthusiasm**, **anger**, **hate** (sub-category of anger), and **emotional neutrality**. Applied **only** to texts previously labelled as political.
2. Intended Use (and Out-of-Scope)
Intended: comparative analysis by actor (public, politicians, media, comments) and electoral period; emotion-value correlations; engagement modelling.
Not intended: clinical diagnostics, legal determination of hate speech, identity inference.
3. Data & Labelling
Size after preprocessing (emotions): **11,884** entries.
Codebook with operational definitions & cues (lexical/contextual) in Annex A of the deliverable.
Reliability: 20% double-coding; $\alpha \geq 0.67$ threshold prior to full annotation.
4. Pipeline & Architecture
Base: XLM-RoBERTa-base (multilabel head with 5 outputs, sigmoid).
Tokenizer: max **256 tokens**.
Split: **MultilabelStratifiedKFold (5 folds)** by language & label distribution; first fold used for train/test.
Hyperparameters: lr=1e-5, batch_size=32, epochs=3; selection by **F1-macro**.
Implementation: HuggingFace Trainer with custom metrics and early stopping.
5. Evaluation (Test)
Global:
F1-macro **0.971777**
F1-micro **0.978355**
Precision-macro **0.971292**
Recall-macro **0.972420**
AUC-ROC-macro **0.997817**
Per emotion:
Fear/Anxiety: F1 **0.959091**, Precision **0.947605**, Recall **0.970859**, AUC **0.996129**
Happiness/Enthusiasm: F1 **0.979857**, Precision **0.986937**, Recall **0.972877**, AUC **0.998072**
Anger: F1 **0.984385**, Precision **0.996524**, Recall **0.972539**, AUC **0.999307**
Hate: F1 **0.956628**, Precision **0.951988**, Recall **0.961312**, AUC **0.996953**
Emotional neutrality: F1 **0.978927**, Precision **0.973404**, Recall **0.984513**, AUC **0.998623**
By language (examples):
BG F1-macro **0.982040**, MK **0.973021**, DE **0.969008**, PL **0.967117**, DA **0.964928**, SR **0.966082**, SQ **0.935966**.

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

6. Analytical Context (for Interpreting Outputs)
Systematic differences by **actor** and **electoral period**: politicians ↑ happiness & fear; comments ↑ anger & hate; posts in election windows become more positive while comments
grow more confrontational. Use these patterns for interpretation–not as training labels.
7. Risks & Limitations
Cultural/idiomatic sensitivity (irony/sarcasm may degrade accuracy).
"Hate" class is an analytical subcategory of anger; it is **not** a legal determination of hate speech.
8. I/O
Input: political text in {DE, DA, PL, BG, MK, SR, SQ}, ≤256 tokens.
Output: vector of five probabilities ($p \in [0,1]$) and active labels (configurable threshold).
9. Governance & Rights
Tech stack: Python 3.12.4; pandas/NumPy; HuggingFace Transformers; scikit-learn; statsmodels; seaborn/matplotlib.
Data governance: flat CSV; duplicate removal by message ID; preserved fields (id, timestamp, author_id/username, language, text, engagement metrics, country/country_code when available, reply/long-post flags).
Dissemination & rights: deliverable marked **PU–Public**; consortium ownership and restrictions on copying/distribution without prior agreement.
10. Source
ENCODE Deliverable: *D3.3 Sentiment analysis*.
Usage
Example of **multilabel** inference with the emotion classifier (sigmoid outputs). Adjust the THRESHOLD per language if needed.

```
from transformers import AutoTokenizer, AutoModelForSequenceClassification,
TextClassificationPipeline
import torch, pandas as pd
model_dir = "models/encode_emotions_xlmr" # <- change to your path
THRESHOLD = 0.50 # consider label active if score >= THRESHOLD
labels = [
"fear_anxiety",
"happiness_enthusiasm",
"anger",
"hate",
"emotional_neutrality",
]
device = 0 if torch.cuda.is_available() else -1
# Load
tokenizer = AutoTokenizer.from_pretrained(model_dir)
model = AutoModelForSequenceClassification.from_pretrained(model_dir)
clf = TextClassificationPipeline(model=model, tokenizer=tokenizer,
device=device, return_all_scores=True, top_k=None)
# Inference
texts = [
"We should be worried about the security situation at the border.",
"So happy with the election results!",
]
raw = clf(texts, batch_size=32, truncation=True, padding=True)
# Convert to multilabel set by threshold
active = []
```


Funded by the European Union

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

```
for row in raw: # row: list of {label, score}
scores = {d['label']: d['score'] for d in row}
on = [lab for lab in labels if scores.get(lab, 0.0) >= THRESHOLD or
scores.get(f"LABEL_{labels.index(lab)}", 0.0) >= THRESHOLD]
active.append(on)
for t, labs in zip(texts, active):
print("TEXT:", t)
print("ACTIVE EMOTIONS:", labs)
```

Notes

Use **the political classifier first**, then apply this model only to entries labelled political.
For production, consider **class-specific thresholds** (per language) computed on a dev set.

```
# Example: per-label thresholds
TH = {"fear_anxiety": 0.48, "happiness_enthusiasm": 0.55, "anger": 0.52, "hate": 0.50,
"emotional_neutrality": 0.53}
```

Model Card — Emotion Classification

(Multilabel)

Model name: ENCODE - Emotion Classifier (XLM-RoBERTa-base)
Document version: v2.0 (22/10/2025)
Owners / Authors: Rodrigo Ortega Izquierdo, Frans Folkvord (Lead Partner: PBY) within the ENCODE consortium.

1. Overview & Objective

Multilabel classifier for five dimensions: **fear/anxiety**, **happiness/enthusiasm**, **anger**, **hate** (sub-category of anger), and **emotional neutrality**. Applied **only** to texts previously labelled as political.

2. Intended Use (and Out-of-Scope)

Intended: comparative analysis by actor (public, politicians, media, comments) and electoral period; emotion-value correlations; engagement modelling.
Not intended: clinical diagnostics, legal determination of hate speech, identity inference.

3. Data & Labelling

Size after preprocessing (emotions): **11,884** entries.
Codebook with operational definitions & cues (lexical/contextual) in Annex A of the deliverable.
Reliability: 20% double-coding; $\alpha \geq 0.67$ threshold prior to full annotation.

4. Pipeline & Architecture

Base: XLM-RoBERTa-base (multilabel head with 5 outputs, sigmoid).
Tokenizer: max **256 tokens**.
Split: **MultilabelStratifiedKFold (5 folds)** by language & label distribution; first fold used for train/test.
Hyperparameters: lr=1e-5, batch_size=32, epochs=3; selection by **F1-macro**.
Implementation: HuggingFace Trainer with custom metrics and early stopping.

5. Evaluation (Test)

Global:
F1-macro **0.971777**
F1-micro **0.978355**
Precision-macro **0.971292**
Recall-macro **0.972420**
AUC-ROC-macro **0.997817**
Per emotion:
Fear/Anxiety: F1 **0.959091**, Precision **0.947605**, Recall **0.970859**, AUC **0.996129**
Happiness/Enthusiasm: F1 **0.979857**, Precision **0.986937**, Recall **0.972877**, AUC **0.998072**

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

Anger: F1 **0.984385**, Precision **0.996524**, Recall **0.972539**, AUC **0.999307**
Hate: F1 **0.956628**, Precision **0.951988**, Recall **0.961312**, AUC **0.996953**
Emotional neutrality: F1 **0.978927**, Precision **0.973404**, Recall **0.984513**, AUC **0.998623**
By language (examples):
BG F1-macro **0.982040**, MK **0.973021**, DE **0.969008**, PL **0.967117**, DA **0.964928**, SR **0.966082**, SQ **0.935966**.
6. Analytical Context (for Interpreting Outputs)
Systematic differences by **actor** and **electoral period**: politicians ↑ happiness & fear; comments ↑ anger & hate; posts in election windows become more positive while comments
grow more confrontational. Use these patterns for interpretation–not as training labels.
7. Risks & Limitations
Cultural/idiomatic sensitivity (irony/sarcasm may degrade accuracy).
"Hate" class is an analytical subcategory of anger; it is **not** a legal determination of hate speech.
8. I/O
Input: political text in {DE, DA, PL, BG, MK, SR, SQ}, ≤256 tokens.
Output: vector of five probabilities (p∈[0,1]) and active labels (configurable threshold).
9. Governance & Rights
Tech stack: Python 3.12.4; pandas/NumPy; HuggingFace Transformers; scikit-learn; statsmodels; seaborn/matplotlib.
Data governance: flat CSV; duplicate removal by message ID; preserved fields (id, timestamp, author_id/username, language, text, engagement metrics, country/country_code when available, reply/long-post flags).
Dissemination & rights: deliverable marked **PU–Public**; consortium ownership and restrictions on copying/distribution without prior agreement.
10. Source
ENCODE Deliverable: *D3.3 Sentiment analysis*.
Usage
Example of **multilabel** inference with the emotion classifier (sigmoid outputs). Adjust the THRESHOLD per language if needed.

```
from transformers import AutoTokenizer, AutoModelForSequenceClassification,
TextClassificationPipeline
import torch, pandas as pd
model_dir = "models/encode_emotions_xlmr" # <- change to your path
THRESHOLD = 0.50 # consider label active if score >= THRESHOLD
labels = [
"fear_anxiety",
"happiness_enthusiasm",
"anger",
"hate",
"emotional_neutrality",
]
device = 0 if torch.cuda.is_available() else -1
# Load
tokenizer = AutoTokenizer.from_pretrained(model_dir)
model = AutoModelForSequenceClassification.from_pretrained(model_dir)
clf = TextClassificationPipeline(model=model, tokenizer=tokenizer,
device=device, return_all_scores=True, top_k=None)
# Inference
texts = [
```


This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).

```
"We should be worried about the security situation at the border.",
"So happy with the election results!",
]
raw = clf(texts, batch_size=32, truncation=True, padding=True)
# Convert to multilabel set by threshold
active = []
for row in raw: # row: list of {label, score}
scores = {d['label']: d['score'] for d in row}
on = [lab for lab in labels if scores.get(lab, 0.0) >= THRESHOLD or
scores.get(f"LABEL_{labels.index(lab)}", 0.0) >= THRESHOLD]
active.append(on)
for t, labs in zip(texts, active):
print("TEXT:", t)
print("ACTIVE EMOTIONS:", labs)
```

Notes

Use the political classifier first, then apply this model only to entries labelled political.
For production, consider class-specific thresholds (per language) computed on a dev set.

This project has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101132698 ENCODE).