Supplementary Information for

Mapping the e-petition ecosystem through Social Media: mobilization in the EU across issues and ideologies

CARLO SANTAGIUSTINA [1,2] PEDRO RAMACIOTTI [3,2,4]

[1] Inria, France

[2] Sciences Po médialab, France

[3] Complex Systems Institute of Paris Ile-de-France, CNRS, France

[4] LPI, CY Cergy Paris University, France

Table of Contents

- Table of Contents p.1
- Supplement Summary p.2
- S.1. Conceptual framework supplement p.3
- S.2. Data supplement p.6
- S.3. ManifestoBerta model and issue inference supplement p.10
- S.4. Human- and LLM-based validation of ManifestoBERTa annotations p.14
- S.5. Consistency of annotations by petition URL p.28
- S.6. Petition platform perplexity formula and computational details p.30
- S.7. Supplementary Results p.31
- References p.35

Author 1: carlo.santagiustina@inria.fr; carlo.santagiustina@sciencespo.fr

Author 2: pedro.ramaciotti-morales@cnrs.fr; pedro.ramaciottimorales@sciencespo.fr

Supplement Summary

This supplement provides additional materials and insights into the methodologies, data, results and limitation of the manuscript: *Mapping the e-petition ecosystem through Social Media: mobilization in the EU across issues and ideologies* (Santagiustina and Ramaciotti, 2026), published on the **Journal of Quantitative Description: Digital Media 6(2026)**. DOI: 10.51685/jqd.2026.001.

This supplement, along with the related code and data, is also available at the following link: https://github.com/carlosantagiustina/mapping_the_epetition_ecosystem_through_social_media

The supplement's contents are organized as follows: The first section (**S.1**), Framework operationalization supplement, provides additional information related to the proposed conceptual framework. The second section (**S.2**), provides additional information related to the data collection and some summary statistics. The third section (**S.3**), provides details about the ManifestoBERTa based automated issue annotation and aggregation process. The fourth section (**S.4**), presents Human- and LLM-based validation exercise and related findings. The fifth section (**S.5**), explores the consistency of annotations by petition URL. The sixth section (**S.6**) presents computational details related to the perplexity scores used in Figure 6. Finally, the last section (**S.7**) presents some additional findings that provide interesting insights, which complement those of the manuscript.

S.1. Framework operationalization supplement

In the following section, we present some additional explanations related to our framework for analyzing social media activity related to e-petitioning.

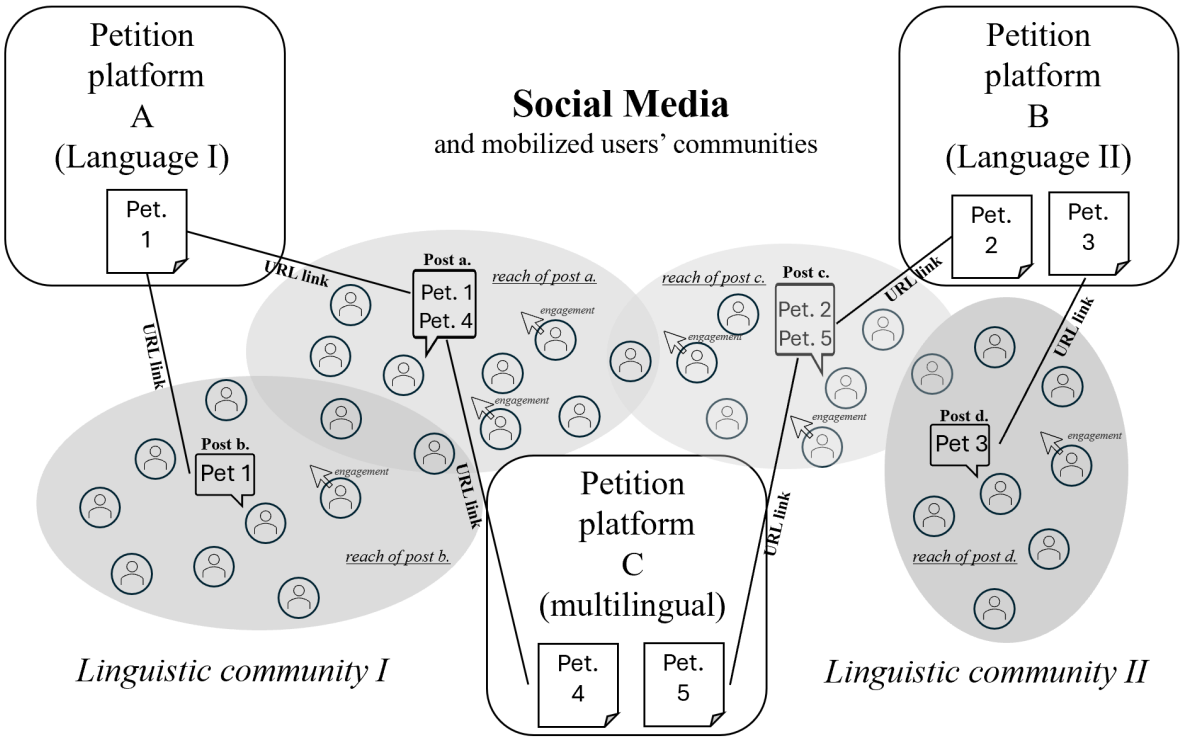

Figure S.1. Visual summary of the Social Media petitioning ecosystem and of the interaction between e-petitions published on multiple petitioning platforms and social media users' online activity.

Figure S.1 represents the social media petitioning ecosystem, across linguistic communities and platforms, highlighting how petitions circulate through online interactions and user engagement (e.g., posts, clicks, views, etc.). It shows that different linguistic communities on social media can engage with petitions from various platforms, some of which operate in a single language, while others are multilingual. Petitioning platforms can be isolated within one linguistic group or serve as bridges that enable petitions to spread across multiple countries and communities. Social media plays a crucial role in this process, as users publish posts containing URLs to e-petition web pages

on petitioning platforms. Depending, among others, on the characteristics of the users posting, their followers, and the algorithmic curation, these posts can have a larger or smaller reach and trigger the engagement of a few or numerous other users.

Different communities might focus on different issues, and petitions may either remain within specific language and ideological groups or expand into broader discussions across multiple linguistic spheres and audience groups. Social media affordances, algorithms, and communities structure this diffusion process, affecting whether petitions will finally influence public discourse at a local, national, or transnational level.

Besides attracting new signatories, social media activity around petitions exerts direct and indirect pressures on political institutions, elected officials and government decision-makers. By mobilizing public opinion on petition-related issues, social media communities, depicted in Figure S.1, can elevate certain issues to prominence in the public debate, much like political manifestos did in the past, making policy responses appear more urgent and appealing to parties, governments, and elected officials. For instance, if a Government Official or a Member of Parliament (MP) who is active on social media notices a surge in engagement from their followers and political audience around a specific petition, they may see an opportunity to align with this momentum. To strengthen their connection with constituents and demonstrate responsiveness, they might endorse the petition, introduce related legislative proposals, or engage in political actions such as posing parliamentary questions, pressuring institutions or bringing the issue to the government. In doing so, social media not only amplifies the petition's reach but also integrates it into broader political and institutional dynamics, reinforcing the interplay between digital activism and formal policymaking. Studying social media activity related to petitions is therefore valuable not only for understanding its effects on the number of signatories but also as a phenomenon in its own right; as the discussions, narratives, and mobilization patterns surrounding petitions on social media can exert an informal yet concrete influence on elected officials and governance bodies, one that is not necessarily mediated by the number of signatures collected.

Studying the social media communities that engage in relation to specific petitioning issues and platforms through their social media activity allows us to characterize the specificities of user activity patterns, map the diverse dynamics of the petitioning ecosystem as a function of language, and understand the implications of online mobilization related to different petitioning platforms, for example, in terms of activity, reach and engagement by issue.

Figure S.2, adapted from Jungherr and Jürgens (2010), summarises the characteristics of four different user mobilization categories as a function of three dimensions: the number of issues discussed by a user; the number of petitions & petition platforms referred to by a user; and the time frame of a user's activity related to petitioning.

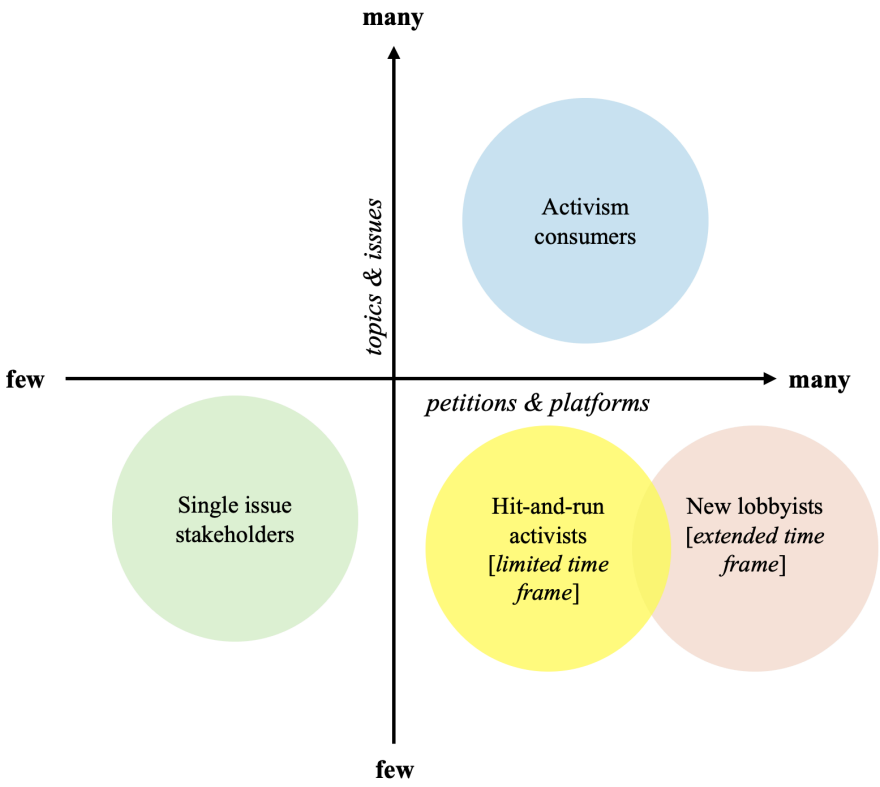

Figure S.2. User mobilization categories. Revisited version of Graph 8 in Jungherr and Jürgens (2010).

Note: Unlike the original figure, the X-axis here captures both the variety of petitions and the variety of petitioning platforms a social media user engages with through their online activity.

S.2. Data supplement

Table S.1 presents the set of language-specific keyword queries used during the exploratory phase (Step I) of data collection. The primary objective of this phase was to gather an initial dataset of petition-related posts across multiple languages. The queries were carefully designed to capture a wide range of variations in the way users express support for petitions in different languages.

For each language, the query includes the word petition together with another keyword related to singing petitions (e.g., sign, signing, signed) or associated with helping or supporting a petitioning cause (e.g., help, aid). These keyword combinations were chosen to maximize recall by covering common phrasings used in petition-related discourse. The Boolean AND operator was used to ensure that retrieved posts contained both the word petition and at least one action-related term, thereby increasing the relevance of the collected exploratory data and reducing false positives.

Table S.1. Exploratory data collection phase I query and N. retrieved posts by year.

Language	Query employed for phase I (i.e., exploratory) data collection
German	(unterschreiben OR unterzeichnen OR unterschriftenaktion OR hilf OR hilfe) AND petition
French	(signez OR signe OR signee OR signer OR signature OR aide OR aidez) AND petition
Dutch	(bevraging OR volksraadpleging OR teken OR teken OR tekenen OR tekenden OR ondertekend OR ondertekenen OR help) AND petitie
Polish	(podpis OR podpiszmy OR podpiszcie OR podpisałem, podpisałeś OR podpisał OR podpisaliśmy OR podpisaliście OR podpisali OR podpiszę OR podpiszesz, podpisze OR podpiszemy OR podpiszecie OR podpiszą OR podpisanie OR podpisanie OR podpisania OR pomóż) AND petycja
Romanian	(semnează OR semnătură OR semnături OR semnat OR semnați OR semnată OR ajuta) AND (petiția OR petiție)
Italian	(firma OR firmato OR firmare OR firma OR aiuta) AND petizione
Spanish	(firma OR firmar OR firme OR firmaron OR firmad OR firmo OR ayuda) and peticion

Table S.3 presents the final query parameters used in our URL-based data collection phase (Step II). This table outlines how we targeted posts containing links to known petitioning platforms, ensuring a more direct mapping of social media content to petition platforms across multiple languages.

Table S.2. Language-specific queries used for data collection, phase II.

Language	**Query**
French	("mesopinions.com" OR "change.org" OR "wemove.eu" OR "petitions.assemblee-nationale.fr" OR "greenvoice.fr" OR "eko.org" OR "damocles.co" OR "petition-patrimoine-danger.fr" OR "avaaz.org" OR "lemouvement.ong" OR "petition-pandemie.com" OR "leslignesbougent.org" OR "soseducation.org" OR "greenpeace.fr" OR "stop-immigration-villages.fr" OR "ifaw.org" OR "non-aux-zfe.fr" OR "non-lectures-drag-queen.fr" OR "petitions.senat.fr" OR "Unepetition.fr" OR "politicall.com" OR "thegoodlobby.fr" OR "redonzeboeren.be" OR "dekamer.mijnopinie.belgium.be/initiatives" OR "petitionenligne.be" OR "parlement-wallonie.be/pwpages?p=petition-detail&id=" OR "europarl.europa.eu/petitions" OR "ciwf.fr" OR "l214.com" OR "allout.org" OR "wesign.it" OR "350.org") lang:fr
German	("change.org" OR "wemove.eu" OR "epetitionen.bundestag.de" OR "campact.de" OR "openpetition.eu" OR "regenwald.org" OR "openpetition.de" OR "finanzwende.de" OR "peta.de" OR "avaaz.org" OR "petitionfuerdemokratie.de" OR "greenpeace.de" OR "citizengo.org" OR "europarl.europa.eu/petitions" OR "innn.it" OR "patriotpetition.org" OR "inkota.de") lang:de
Italian	("change.org" OR "greenpeace.it" OR "wemove.eu" OR "provitaefamiglia.it" OR "chng.it" OR "avaaz.org" OR "buonacausa.org" OR "citizengo.org" OR "thegoodlobby.it" OR "justice-initiative.eu" OR "petizioni.com" OR "europarl.europa.eu/petitions" OR "essereanimali.org" OR "actionnetwork.org" OR "tilt.green" OR "lav.it" OR "allout.org" OR "savethechildren.it" OR "animalequality.it" OR "codepink.org" OR "associazionetbs.org" OR "lidasezolbia.it") lang:it
Dutch	("citizengo.org" OR "petities.nl" OR "amnesty.nl" OR "avaaz.org" OR "greenpeace.nl" OR "stop4deroute.nl" OR "natuurmonumenten.nl" OR "petities.com" OR "redonzeboeren.be" OR "stop4deroute.nl" OR "guerrillagardeners.nl" OR "rightsforum.org" OR "dagelijksestandaard.nl" OR "europarl.europa.eu/at-your-service/nl/be-heard/petitions" OR "change.org" OR "redonzeboeren.be" OR "dekamer.mijnopinie.belgium.be/initiatives" OR "petitionenligne.be" OR "parlement-wallonie.be/pwpages?p=petition-detail&id=" OR "europarl.europa.eu/petitions" OR "degoedezaak.org" OR "dierenrecht.nl" OR "cultuurondervuur.nl") lang:nl
Spanish	("change.org" OR "avaaz.org" OR "peticion.es" OR "thegoodlobby.es" OR "openpetition.eu" OR "europarl.europa.eu/petitions") lang:es
Romanian	("petitieonline.com" OR "declic.ro" OR "change.org" OR "avaaz.org" OR "europarl.europa.eu/petitions") lang:ro
Polish	("change.org" OR "peta.org.uk" OR "petycjeonline.com" OR "avaaz.org" OR "wwf.pl" OR "naszademokracja.pl" OR "polskitarg.pl" OR "europarl.europa.eu/petitions" OR "cenafutra.pl" OR "europabezfuter.pl") lang:pl

Table S.3 provides an overview of the petitioning platforms identified in the dataset collected in Step II, categorized by platform language. For each language, the table lists the domains of petition platforms that appeared in the retrieved data, along with the number of posts referencing those platforms. Additionally, the share of posts for each language is presented as a percentage of the total dataset. The dataset includes posts from petition platforms operating in six individual languages (Dutch, French, German, Italian, Polish, and Romanian) as well as multilingual platforms that operate in at least two of the targeted EU languages. The presence of platforms in the dataset varies significantly across languages, reflecting differences in petitioning practices, platform popularity, and levels of engagement across linguistic communities on X.

Table S.3. Petitioning platforms, number of platforms mentioned in retrieved data and number of posts collected, by language.

Platform Language	Domains of petition platforms in query observed in the collected dataset	N. posts by platform language (% tot.)
Dutch	amnesty.nl; cultuurondervuur.nl; dagelijksestandaard.nl; greenpeace.nl; guerrillagardeners.nl; natuurmonumenten.nl; petities.com; petities.nl; stop4deroute.nl	2 781 (1.11%)
French	350.org; assemblee-nationale.fr; damocles.co; eko.org; greenpeace.fr; greenvoice.fr; l214.com; lemouvement.ong; leslignesbougent.org; mesopinions.com; senat.fr; soseducation.org; wesign.it	129 649 (51.53%)
German	bundestag.de; campact.de; finanzwende.de; greenpeace.de; inkota.de; innn.it; openpetition.de; patriotpetition.org; petitionfuerdemokratie.de; regenwald.org	39 663 (15.76%)
Italian	actionnetwork.org; associazionetbs.org; buonacausa.org; chng.it; lav.it; petizioni.com; provitaefamiglia.it; thegoodlobby.it	1 266 (0.50%)
Polish	naszademokracja.pl; [peta.]org.uk; petycjeonline.com	1 308 (0.52%)
Romanian	petitieonline.com	3 (>0.1%)
Spanish	peticion.es	24 (>0.1%)
Multilingual	allout.org; avaaz.org; change.org; citizengo.org; europa.eu; ifaw.org; petitionenligne.be; wemove.eu	76 924 (30.57%)

Note. The multilingual category includes all platforms for which data was collected in at least two EU languages among the targeted ones listed above.

Table S.4 presents key descriptive statistics from the second phase of data collection (Step II), providing an overview of the retrieved posts, the number of targeted petitioning platforms, user engagement, and language diversity. The table highlights the scale of the dataset, including the number of unique users, user-language pairs, and the average number of posts and mentioned platforms per user, among others.

Table S.4. Summary statistics of data collected from X during the Step II.

Metric	**Value**
Start date - DD/MM/YY	19/02/23
End date - DD/MM/YY	04/11/24
N. retrieved posts matching query	164 370
N. platforms in query (share of targeted, in percent)	53 (63.10 %)
N. unique users	75 076
N. unique user - language pairs	77 006
Avg. N. posts per user	2.19
Avg. N. mentioned platforms per user	1.067

S.3. ManifestoBerta model and issue inference supplement

To systematically analyze the issues discussed in petition-related posts, we employ an automated classification pipeline powered by the ManifestoBERTa sentence model (2023 version). We used the sentence model version (instead of the context model), as we do not have any post-specific textual context text to provide. This automated annotation process is implemented through a Python script that ensures efficient text preprocessing and output generation.

The python script we created to annotate petition-related posts using ManifestoBERTa is designed to handle large-scale data by performing the following tasks:

1. *Loading the model*: The script initializes the ManifestoBERTa model and tokenizer, ensuring they are deployed efficiently on the GPU (if available). In our case we used an NVIDIA RTX A4000 GPU with 16Gb of RAM.
2. Text preprocessing: Raw social media posts are preprocessed and converting text to a suitable tokenized format for the model.
3. *Issue classification*: Tokenized posts fed into the ManifestoBERTa model. The model then assigns a weighted confidence score over a predefined set of issues. To optimize performance, posts are processed in batches, reducing memory usage and ensuring smooth execution for large datasets.
4. *Incremental output writing and memory management*: To efficiently handle large-scale data, the script writes classification results to a CSV file in stages, preventing memory overload. Also, the script systematically clears unused variables and GPU memory after each batch, ensuring optimal resource utilization.

ManifestoBERTa model assigns weighted confidence scores to issues which are included in the Manifesto Project ontology. After classification, we perform an issue aggregation step, detailed

here below, where original issue scores are summed to derive aggregate issue scores (see Table S.5 for a list of issues before and after aggregation).

Let N $(with\ N = 164\,370)$ be the total number of petition related posts in our dataset. Each post obs_i, with $i \in \{1, ..., N\}$, is assigned by ManifestoBERTa a weighted confidence score distribution $P_i = (p_{i,1}, \ldots, p_{i,k})$ over K issues (with $K = 56$), where $p_{i,j}$ represents the model's confidence that issue j is discussed in the i'th post obs_i.

P_i satisfies the following condition:

$$\sum_{j=1}^{K} p_{i,j} = 1, \quad \forall i \in \{1, \ldots, N\}$$

Some ManifestoBERTa issues have suffixes **Positive (Pos.)** and **Negative (Neg.)**, indicating the stance towards an issue. As shown in Table S.5, this is the case for 20 issues. Given that **Positive** and **Negative** stances towards issues are difficult to infer from short social media posts, in our work we remove these suffixes and merge weighted confidence scores of "stanced" issues that have the same base issue label.

Let C be the set of 46 unique base issue labels (i.e., issue names without the **Positive** or **Negative** suffix), listed in the last column of Table S.5. Each base issue $c \in C$ corresponds to the union of its "stanced" versions ($c_{Pos.}$ and $c_{Neg.}$), if any. For such issues, the aggregated score $p_{i,c}$ of issue c in post obs_i is given by:

$$p_{i,c} = p_{i,c_{Pos.}} + p_{i,c_{Neg.}}$$

To infer the most likely base issue c^* for a given post obs_i, we need to identify the base issue with the highest aggregated score for that specific post:

$$c^* = \arg\max_{c \in \mathcal{C}} p_{i,c}$$

Table S.5. ManifestoBERTa issues before and after aggregation. In bold, suffixes of issue labels that were removed during the aggregation phase.

	Issues *before aggregation*	Base issues *after aggregation*
Number of issues	56	46
Issue labels	Agriculture and Farmers**: Positive**; Anti-Growth Economy: **Positive**; Anti-Imperialism; Centralisation; Civic Mindedness: **Positive**; Constitutionalism: **Negative**; Constitutionalism: **Positive**; Controlled Economy; Corporatism/ Mixed Economy; Culture: **Positive**; Democracy; Economic Goals; Economic Growth: **Positive**; Economic Orthodoxy; Economic Planning; Education Expansion; Education Limitation; Environmental Protection: **Positive**; Equality: **Positive**; European Community/Union: **Negative**; European Community/Union: **Positive**; Federalism; Foreign Special Relationships: **Negative**; Foreign Special Relationships: **Positive**; Free Market Economy; Freedom and Human Rights; Governmental and Administrative Efficiency; Incentives; Internationalism: **Negative**; Internationalism: **Positive**; Keynesian Demand Management; Labour Groups: **Negative**; Labour Groups: **Positive**; Law and Order: **Positive**; Market Regulation; Marxist Analysis: Positive; Middle Class and Professional Groups; Military: **Negative**; Military: **Positive**; Multiculturalism: **Negative**; Multiculturalism: **Positive**; National Way of Life: **Negative**; National Way of Life: **Positive**; Nationalisation; Non-economic Demographic Groups; Peace; Political Authority; Political Corruption; Protectionism: **Negative**; Protectionism: **Positive**; Technology and Infrastructure; Traditional Morality: **Negative**; Traditional Morality: **Positive**; Underprivileged Minority Groups; Welfare State Expansion; Welfare State Limitation	Agriculture and Farmers; Anti-Growth Economy; Anti-Imperialism; Centralisation; Civic Mindedness; Constitutionalism; Controlled Economy; Corporatism/ Mixed Economy; Culture; Democracy; Economic Goals; Economic Growth; Economic Orthodoxy; Economic Planning; Education Expansion; Education Limitation; Environmental Protection; Equality; European Community/Union; Federalism; Foreign Special Relationships; Free Market Economy; Freedom and Human Rights; Governmental and Administrative Efficiency; Incentives; Internationalism; Keynesian Demand Management; Labour Groups; Law and Order; Market Regulation; Marxist Analysis; Middle Class and Professional Groups; Military; Multiculturalism; National Way of Life; Nationalisation; Non-economic Demographic Groups; Peace; Political Authority; Political Corruption; Protectionism; Technology and Infrastructure; Traditional Morality; Underprivileged Minority Groups; Welfare State Expansion; Welfare State Limitation

As explained in the ManifestoBERTa 2023 performance report[1], tests done by the authors of the model show that, in general, the model isn't systematically over- or under predicting specific issues. The ManifestoBERTa model appears to be particularly well calibrated for the 1° ranked issue score

[1] https://manifesto-project.wzb.eu/uploads/manifestoberta_sentence_2023_1_1_56topics_perf_report.pdf

ranges above 95%, while it tends to be slightly overconfident in the range [20%, 95%] and underconfident in the [0%, 20%] range.

Even though ManifestoBERTa is often used to label political texts by predicting the most likely issue (i.e., the issue with the highest confidence score for a given observation), the model also provides, for each post, a distribution reflecting (model-inferred) weighted confidence scores of all issues. By this, we mean that the scores produced by the model, which sum to one and lie in the [0, 1] range, retain meaningful information about issues' likelihood and mixing, even if the model is not perfectly calibrated in certain score ranges. We assume that such correlations somehow reflect patterns of issue complementarity (i.e., mixing) or mutual-inhibition.

Hence, by examining the correlation structure among these weighted confidence scores (see Figure 7 in the Manuscript), we can identify which issues are more or less likely to be mixed and therefore appear together.

For instance, if petition-related posts consistently receive a high score for *Environmental Protection* and for *Anti-Growth Economy*, it suggests that these issues are likely appearing together in petition related posts.

S.4. Human- and LLM-based validation of ManifestoBERTa annotations

To validate ManifestoBERTa annotations, we conducted both human- and LLM-based validation tasks using a random sample of 500 observations. Below, we provide the guidelines and instructions used during the human annotations, the LLM prompt and parameters, and an analysis of the alignment results from our two validation exercises.

Human-based annotation guidelines

Each observation was manually annotated by the authors with up to three labels from the Manifesto Project ontology. When multiple issues were present in a post, labels were ranked according to the instructions provided to the Manifesto Project annotators (detailed here: https://manifesto-project.wzb.eu/information/documents?name=handbook_v4).

Since authors do not speak all the languages represented in the dataset, posts in languages we do not fully master were translated into English before annotation. Two types of translations were employed: Google Translate and M2M100 model-based translations (see details here: https://huggingface.co/docs/transformers/en/model_doc/m2m_100). In most cases, these translations were nearly perfectly aligned with each other and with the original content, as verified by comparing translations of languages we master.

In general, we aimed to follow the annotation instructions provided to the Manifesto Project annotators as closely as possible; however, due to differences in data type (posts' text vs. quasi-sentences in manifestos), some deviations and additional rules were necessary to ensure rigorous and comparable annotations of posts about petitions.

Unlike the approach of the Manifesto Project annotators, in situations where multiple issue annotations could apply, we did not consider the context of the post (i.e., the conversation) to choose which label(s)to assign. We made this decision because we did not have access to such

context (i.e., the thread of the post), nor was it provided to the ManifestoBERTa model during annotation.

Similar to the Manifesto Project data, posts about petitions often take the form of "support us for the cause [X] (and [Y]) by doing/mobilizing for [A]". When [X] and [Y] are present and appear to be the primary focus of the message with respect to A, the issues they refer to are ranked as the top two, followed by the mobilizing for [A] issue, as the third-ranked one.

This aligns with the Manifesto annotation guidelines, which specify that goals should take precedence over means (including intermediary goals) when assigning codes. In our case, this also implies that when a post contains only calls for mobilization or support for a petition (e.g., "help your fellow citizens by supporting this petition [petition URL]"), the issue assignment is based solely on the mobilization call. According to the Manifesto Project ontology definitions, if there is no information about the content of the petition but only a call for participation/mobilization/signing the petition this typically falls under the Democracy issue, also depending on the specific content and framing of the call to action or request. We track posts that contain only a generic call to action (participation/mobilization/signing) with an ad hoc dummy variable called "generic".

Human vs ManifestoBERTa annotations alignment

Here follow the summary results and classification performance metrics (Accuracy, Precision, Recall, F1) of the validation exercise comparing ManifestoBERTa annotations with human-based ones, for a sample of 500 randomly selected posts. All posts in the sample, except one that lacked sufficient information to assign an issue label, were manually labelled with at least one issue (the 1° ranked one); 107 were labeled with at least two issues (the 1° and 2° ranked one), and 22 with three issues (the 1°, 2° and 3° ranked one). The top-1 accuracy is 0.78, the top-2 accuracy is 0.89, and the top-3 accuracy is 0.94. This highlights that in about 19 cases out of 20, the three top-ranked issues inferred by ManifestoBERTa contain the "true" (human) label. We recall that, as detailed in the

previous subsection, in cases where only a generic call for participation/mobilization/signing is present in the post, and no detailed information about the petition itself, the observation will typically fall under the Democracy issue, given that the definition of this issue includes the following statement: "*[...] the need for the involvement of all citizens in political decisionmaking; support for either direct or representative democracy [...]"*. This issue somehow also explains the high overall accuracy of the model, which, in our application setting (that of X posts referring to e-petitions), proves to be very good at predicting generic references to the Democracy issue, which, in our sample, is observed much more frequently than other issue classes (see Table S.8).

Across the sample, the mean intersection size between the human top-k issue set ($k \leq 3$) and the model's top-3 issue set is 1.198 labels (i.e., the average hits@3 against a multi-label human annotation of up to three labels is 1.198).

The aforementioned results highlight a substantial overlap between the top-3 issues identified by the model and the manually identified ones, indicating that ManifestoBERTa's non-first-ranked issue scores are informative signals. This is a further reason for considering in our analysis not only the top ranked ManifestoBERTa issue, but the whole distribution of issue weighted confidence scores provided by the model.

Table S.6. Classification performance metrics (N=499).

Accuracy	Avg. Precision	Avg. Recall	Avg. F1	Weighted Precision	Weighted Recall	Weighted F1
0.780	0.553	0.612	0.654	0.798	0.78	0.783

Note. Human 1° issue label by post is considered the ground truth, and ManifestoBERTa's issue label is considered the predicted label.

If we exclude all posts containing only generic calls for participation/mobilization/signing, the classification performance metrics are the following:

Table S.7. Classification performance metrics excluding generic calls for participation/mobilization/signing (N=239), as identified by the human annotator.

Accuracy	Avg. Precision	Avg. Recall	Avg. F1	Weighted Precision	Weighted Recall	Weighted F1
0.628	0.552	0.615	0.640	0.751	0.628	0.664

Note. Human 1° issue label by post is considered the ground truth, and ManifestoBERTa's issue label is considered the predicted label.

While if we look to performance metrics per class (considering only classes with a support of at least 1%, that is 5 observations in the sample) results are the following:

Table S.8. Classification performance metrics per class, considering only issue classes with a support of at least 5 observations (1% of the random sample).

Issue label	Support	Precision	Recall	F1
Democracy	264	0.880	0.920	0.900
Political Authority	46	0.375	0.326	0.349
Environmental Protection	36	0.906	0.806	0.853
Political Corruption	34	0.889	0.706	0.787
Internationalism	23	0.800	0.522	0.632
Welfare State Expansion	23	0.833	0.870	0.851
Culture	13	0.556	0.385	0.455
Freedom and Human Rights	10	0.500	0.800	0.615
Law and Order	10	0.875	0.700	0.778
Civic Mindedness	8	0.714	0.625	0.667
Multiculturalism	8	1.000	1.000	1.000

Note. Human 1° issue label by post is considered the ground truth, and ManifestoBERTa's issue label is considered the predicted label.

LLM-based annotation prompt

We also undertake an issue annotation task using a LLM (ChatGPT 4.1-mini). To achieve this task, we developed a prompt (as simple as possible), instructing the model on how to use the Manifesto Project ontology to identify and label the issues discussed in posts from our random sample of 500 observations.

The system prompt contains the complete Manifesto Project ontology (in a JSON format), including domains, issues (called topics in the prompt), and issue descriptions identical to those provided to human annotators. It also provides a procedure for mapping the content of each post to the ontology entries. The prompt emphasized that the model should produce only the required JSON output format, without any additional commentary or explanation.

To ensure that the LLM's prompt somehow mirrors the human annotation task, we incorporated ranking criteria similar to those used by human annotators, among others to prioritize salient issues (e..g, goal-related). When a post only called for mobilizing, supporting, or signing a petition without specifying the objective or describing the petition itself, the model was instructed to assign the default issue of Democracy and mark the annotation as generic using an ad hoc dummy variable. This approach allowed posts consisting solely of calls for petition mobilization or signing to be handled consistently and in alignment with how they were handled in the human annotation task.

We configured the model parameters to maximize reliability and reproducibility, setting a temperature of 0.00 and a top_p of 0.00 to minimize randomness and enforce deterministic token sampling. Our LLM-based annotation strategy can be considered zero-shot, as no examples are provided. Each post is processed individually by the LLM according to the input format specified at the end of the prompt. We use a separate session for each post annotation, ensuring that annotations of different posts are independent and cannot influence one another.

The system prompt we use is the following:

Table S.9. System prompt used to annotated posts using an LLM

```
You are a multilingual expert political text annotator. Your job is to read a short input text and identify the topics discussed in the text by grounding your annotations and text labels on the provided Ontology, which specifies domains, topics and their descriptions.
Follow these instructions exactly. Do not include any extra commentary, explanations, or text outside the required JSON output, which is specified in the Output Format section.

Ontology:
{
"domains": [
{
"domain": "External Relations",
"topics": [
{
"topic": "Foreign Special Relationships: Positive",
"description": "Favourable mentions of particular countries with which the manifesto country has a special relationship; the need for co-operation with and/or aid to such countries."
},
{
"topic": "Foreign Special Relationships: Negative",
"description": "Negative mentions of particular countries with which the manifesto country has a special relationship."
},
{
"topic": "Anti-Imperialism",
"description": "Negative references to imperial behaviour and/or negative references toone state exerting strong influence
```

(political, military or commercial) over other states. May also include:\n· Negative references to controlling other countries as if they were part of an empire;\n· Favourable references to greater self-government and independence for colonies;\n· Favourable mentions of de-colonisation."
},
{
"topic": "Military: Positive",
"description": "The importance of external security and defence. May include statements concerning:\n· The need to maintain or increase military expenditure;\n· The need to secure adequate manpower in the military;\n· The need to modernise armed forces and improve military strength;\n· The need for rearmament and self-defence;\n· The need to keep military treaty obligations."
},
{
"topic": "Military: Negative",
"description": "Negative references to the military or use of military power to solve conflicts. References to the 'evils of war'. May include references to:\n· Decreasing military expenditures;\n· Disarmament;\n· Reduced or abolished conscription."
},
{
"topic": "Peace",
"description": "Any declaration of belief in peace and peaceful means of solving crises\n– absent reference to the military. May include:\n· Peace as a general goal;\n· Desirability of countries joining in negotiations with hostile countries;\n· Ending wars in order to establish peace."
},
{
"topic": "Internationalism: Positive",
"description": "Need for international co-operation, including co-operation with specific\ncountries other than those idd in 101. May also include references\nto the:\n· Need for aid to developing countries;\n· Need for world planning of resources;\n· Support for global governance;\n· Need for international courts;\n· Support for UN or other international organisations."
},
{
"topic": "European Community/Union: Positive",
"description": "Favourable mentions of European Community/Union in general. May\ninclude the:\n· Desirability of the manifesto country joining (or remaining a\nmember);\n· Desirability of expanding the European Community/Union;\n· Desirability of increasing the ECs/EUs competences;\n· Desirability of expanding the competences of the European Parliament."
},
{
"topic": "Internationalism: Negative",
"description": "Negative references to international co-operation. Favourable mentions\nof national independence and sovereignty with regard to the manifesto\ncountry's foreign policy, isolation and/or unilateralism as opposed to\ninternationalism."
},
{
"topic": "European Community/Union: Negative",
"description": "Negative references to the European Community/Union. May include:\n· Opposition to specific European policies which are preferred by European authorities;\n· Opposition to the net-contribution of the manifesto country to the EU budget."
}
]
},
{
"domain": "Freedom and Democracy",
"topics": [
{
"topic": "Freedom and Human Rights",
"description": "Favourable mentions of importance of personal freedom and civil rights in the manifesto and other countries. May include mentions of:\n· The right to the freedom of speech, press, assembly etc.;\n· Freedom from state coercion in the political and economic spheres;\n· Freedom from bureaucratic control;\n· The idea of individualism."
},
{

"topic": "Democracy",
"description": "Favourable mentions of democracy as the “only game in town”. General support for the manifesto country’s democracy. May also include:\n· Democracy as method or goal in national, international or other organisations (e.g. labour unions, political parties etc.);\n· The need for the involvement of all citizens in political decisionmaking;\n· Support for either direct or representative democracy;\n· Support for parts of democratic regimes (rule of law, division of powers, independence of courts etc.)."
},
{
"topic": "Constitutionalism: Positive",
"description": "Support for maintaining the status quo of the constitution. Support for specific aspects of the manifesto country’s constitution. The use of constitutionalism as an argument for any policy."
},
{
"topic": "Constitutionalism: Negative",
"description": "Opposition to the entirety or specific aspects of the manifesto country’s constitution. Calls for constitutional amendments or changes. May include calls to abolish or rewrite the current constitution."
}
]
},
{
"domain": "Political System",
"topics": [
{
"topic": "Decentralization",
"description": "Support for federalism or decentralisation of political and/or economic power. May include:\n· Favourable mentions of the territorial subsidiary principle;\n· More autonomy for any sub-national level in policy making and/or economics, including municipalities;\n· Support for the continuation and importance of local and regional customs and symbols and/or deference to local expertise;\n· Favourable mentions of special consideration for sub-national areas."
},
{
"topic": "Centralisation",
"description": "General opposition to political decision-making at lower political levels. Support for unitary government and for more centralisation in political and administrative procedures."
},
{
"topic": "Governmental and Administrative Efficiency",
"description": "Need for efficiency and economy in government and administration and/or the general appeal to make the process of government and administration cheaper and more efficient. May include:\n· Restructuring the civil service;\n· Cutting down on the civil service;\n· Improving bureaucratic procedures.\nNote: Specific policy positions overrule this category! If there is no specific policy position, however, this category applies."
},
{
"topic": "Political Corruption",
"description": "Need to eliminate political corruption and associated abuses of political and/or bureaucratic power. Need to abolish clientelist structures and practices."
},
{
"topic": "Political Authority",
"description": "References to the manifesto party’s competence to govern and/or other party’s lack of such competence. Also includes favourable mentions of the desirability of a strong and/or stable government in general."
}
]
},
{
"domain": "Economy",
"topics": [
{
"topic": "Free Market Economy",
"description": "Favourable mentions of the free market and free market capitalism as an economic model. May include favourable references to:\n· Laissez-faire economy;\n· Superiority of individual enterprise over state and control systems;\n· Private property rights;\n· Personal enterprise and initiative;\n· Need for unhampered individual enterprises."
},

{
"topic": "Incentives: Positive",
"description": "Favourable mentions of supply side oriented economic policies (assistance to businesses rather than consumers). May include:\n· Financial and other incentives such as subsidies, tax breaks etc.;\n· Wage and tax policies to induce enterprise;\n· Encouragement to start enterprises."
},
{
"topic": "Market Regulation",
"description": "Support for policies designed to create a fair and open economic market. May include:\n· Calls for increased consumer protection;\n· Increasing economic competition by preventing monopolies and other actions disrupting the functioning of the market;\n· Defence of small businesses against disruptive powers of big businesses;\n· Social market economy."
},
{
"topic": "Economic Planning",
"description": "Favourable mentions of long-standing economic planning by the government. May be:\n· Policy plans, strategies, policy patterns etc.;\n· Of a consultative or indicative nature."
},
{
"topic": "Corporatism/Mixed Economy",
"description": "Favourable mentions of cooperation of government, employers, and trade unions simultaneously. The collaboration of employers and employee organisations in overall economic planning supervised by the state."
},
{
"topic": "Protectionism: Positive",
"description": "Favourable mentions of extending or maintaining the protection of internal markets (by the manifesto or other countries). Measures may include:\n· Tariffs;\n· Quota restrictions;\n· Export subsidies."
},
{
"topic": "Protectionism: Negative",
"description": "Support for the concept of free trade and open markets. Call for abolishing all means of market protection (in the manifesto or any other country)."
},
{
"topic": "Economic Goals",
"description": "Broad and general economic goals that are not mentioned in relation to any other category. General economic statements that fail to include any specific goal.\nNote: Specific policy positions overrule this category! If there is no specific policy position, however, this category applies."
},
{
"topic": "Keynesian Demand Management",
"description": "Favourable mentions of demand side oriented economic policies (assistance to consumers rather than businesses). Particularly includes increase private demand through:\n· Increasing public demand;\n· Increasing social expenditures.\nMay also include:\n· Stabilisation in the face of depression;\n· Government stimulus plans in the face of economic crises."
},
{
"topic": "Economic Growth: Positive",
"description": "The paradigm of economic growth. Includes:\n· General need to encourage or facilitate greater production;\n· Need for the government to take measures to aid economic growth."
},
{
"topic": "Technology and Infrastructure: Positive",
"description": "Importance of modernisation of industry and updated methods of transport and communication. May include:\n· Importance of science and technological developments in industry;\n· Need for training and research within the economy (This does not imply education in general (see category 506);\n· Calls for public spending on infrastructure such as roads and bridges;\n· Support for public spending on technological infrastructure (e.g.: broadband internet, etc.)."
},
{
"topic": "Controlled Economy",
"description": "Support for direct government control of economy. May include, for instance:\n· Control over prices;\n·

Introduction of minimum wages."
},
{
"topic": "Nationalisation",
"description": "Favourable mentions of government ownership of industries, either partial or complete; calls for keeping nationalised industries in state\nhand or nationalising currently private industries. May also include favourable mentions of government ownership of land."
},
{
"topic": "Economic Orthodoxy",
"description": "Need for economically healthy government policy making. May include calls for:\n· Reduction of budget deficits;\n· Retrenchment in crisis;\n· Thrift and savings in the face of economic hardship;\n· Support for traditional economic institutions such as stock market and banking system;\n· Support for strong currency."
},
{
"topic": "Marxist Analysis",
"description": "Positive references to Marxist-Leninist ideology and specific use of Marxist-Leninist terminology by the manifesto party (typically but not necessary by communist parties)."
},
{
"topic": "Anti-Growth Economy: Positive",
"description": "Favourable mentions of anti-growth politics. Rejection of the idea that all growth is good growth. Opposition to growth that causes environmental or societal harm. Call for sustainable economic development."
}
]
},
{
"domain": "Welfare and Quality of Life",
"topics": [
{
"topic": "Environmental Protection",
"description": "General policies in favour of protecting the environment, fighting climate change, and other “green” policies. For instance:\n· General preservation of natural resources;\n· Preservation of countryside, forests, etc.;\n· Protection of national parks;\n· Animal rights.\nMay include a great variance of policies that have the unified goal of environmental protection."
},
{
"topic": "Culture: Positive",
"description": "Need for state funding of cultural and leisure facilities including arts and sport. May include:\n· The need to fund museums, art galleries, libraries etc.;\n· The need to encourage cultural mass media and worthwhile leisure activities, such as public sport clubs."
},
{
"topic": "Equality: Positive",
"description": "Concept of social justice and the need for fair treatment of all people. This may include:\n· Special protection for underprivileged social groups;\n· Removal of class barriers;\n· Need for fair distribution of resources;\n· The end of discrimination (e.g. racial or sexual discrimination)."
},
{
"topic": "Welfare State Expansion",
"description": "Favourable mentions of need to introduce, maintain or expand any public social service or social security scheme. This includes, for example, government funding of: \n· Health care \n· Child care \n· Elder care and pensions \n· Social housing \nNote: This category excludes education."
},
{
"topic": "Welfare State Limitation",
"description": "Limiting state expenditures on social services or social security. Favourable mentions of the social subsidiary principle (i.e. private care before state care);"
},
{
"topic": "Education Expansion",
"description": "Need to expand and/or improve educational provision at all levels. \nNote: This excludes technical training."

```
},
{
"topic": "Education Limitation",
"description": "Limiting state expenditure on education. May include:\n· The introduction or expansion of study fees at all educational levels\n· Increasing the number of private schools."
}
]
},
{
"domain": "Fabric of Society",
"topics": [
{
"topic": "National Way of Life: Positive",
"description": "Favourable mentions of the manifesto country's nation, history, and general appeals. May include: \n· Support for established national ideas;\n· General appeals to pride of citizenship;\n· Appeals to patriotism;\n· Appeals to nationalism;\n· Suspension of some freedoms in order to protect the state against subversion."
},
{
"topic": "National Way of Life: Negative",
"description": "Unfavourable mentions of the manifesto country's nation and history. May include: \n· Opposition to patriotism; \n· Opposition to nationalism; \n· Opposition to the existing national state, national pride, and national ideas."
},
{
"topic": "Traditional Morality: Positive",
"description": "Favourable mentions of traditional and/or religious moral values. May include:\n· Prohibition, censorship and suppression of immorality and unseemly behaviour;\n· Maintenance and stability of the traditional family as a value;\n· Support for the role of religious institutions in state and society."
},
{
"topic": "Traditional Morality: Negative",
"description": "Opposition to traditional and/or religious moral values. May include: \n· Support for divorce, abortion etc.; \n· General support for modern family composition; \n· Calls for the separation of church and state."
},
{
"topic": "Law and Order: Positive",
"description": "Favourable mentions of strict law enforcement, and tougher actions against domestic crime. Only refers to the enforcement of the status quo of the manifesto country's law id. May include: \n· Increasing support and resources for the police; \n· Tougher attitudes in courts; \n· Importance of internal security."
},
{
"topic": "Civic Mindedness: Positive",
"description": "Appeals for national solidarity and the need for society to see itself as united. Calls for solidarity with and help for fellow people, familiar and unfamiliar. May include: \n· Favourable mention of the civil society; \n· Decrying anti-social attitudes in times of crisis; \n· Appeal for public spiritedness; \n· Support for the public interest."
},
{
"topic": "Multiculturalism: Positive",
"description": "Favourable mentions of cultural diversity and cultural plurality within domestic societies. May include the preservation of autonomy of religious, linguistic heritages within the country including special educational provisions."
},
{
"topic": "Multiculturalism: Negative",
"description": "The enforcement or encouragement of cultural integration. Appeals for cultural homogeneity in society."
}
]
},
{
"domain": "Social Groups",
"topics": [
{
"topic": "Labour Groups: Positive",
```

"description": "Favourable references to all labour groups, the working class, and unemployed workers in general. Support for trade unions and calls for the good treatment of all employees, including: \n· More jobs; \n· Good working conditions; \n· Fair wages; \n· Pension provisions etc."
},
{
"topic": "Labour Groups: Negative",
"description": "Negative references to labour groups and trade unions. May focus specifically on the danger of unions 'abusing power'."
},
{
"topic": "Agriculture and Farmers: Positive",
"description": "Specific policies in favour of agriculture and farmers. Includes all types of agriculture and farming practises. Only statements that have agriculture as the key goal should be included in this category."
},
{
"topic": "Middle Class and Professional Groups",
"description": "General favourable references to the middle class. Specifically, statements may include references to: \n· Professional groups, (e.g.: doctors or lawyers); \n· White collar groups, (e.g.: bankers or office employees), \n· Service sector groups (e.g.: IT industry employees); \n· Old and/or new middle class."
},
{
"topic": "Underprivileged Minority Groups",
"description": "Very general favourable references to underprivileged minorities who are defined neither in economic nor in demographic terms (e.g. the handicapped, homosexuals, immigrants, indigenous). Only includes favourable statements that cannot be classified in other topics."
},
{
"topic": "Non-economic Demographic Groups",
"description": "General favourable mentions of demographically defined special interest groups of all kinds. They may include: \n· Women; \n· University students; \n· Old, young, or middle aged people. \nMight include references to assistance to these groups, but only if these do not fall under other topics."
}
]
}
]
}

Procedure (what you must do):
- Extract the "text" field from the provided JSON input.
- Identify and map topics(s) discussed in the text to entries in the provided Ontology (use the ontology topic and domain labels verbatim).
- Select the top 3 topics of the text according to the Topic Ranking Guidelines and Criteria below.
- Produce only the JSON output in the exact Output Format section. No additional keys, comments, or text.

Mapping rules:
- Use the ontology entries verbatim for topic names ("topic") and descriptions ("description") to anchor your topic labels.
- Multiple topics may be present; if this is the case, choose the top 3 topics based on the Topic Ranking Guidelines and Criteria.
- Do not invent topics not present in the Ontology and do not change topic labels.

Topic Ranking Guidelines and Criteria:
- Topical relevance: Assess how central the topic is to the main message or communicative purpose of the text. The topic should represent what the text is primarily about, not just a secondary mention or example.
- Alignment to ontology: Evaluate how closely the text's content matches the description of the topic from the ontology. Prefer precise matches over vague ones. When multiple ontology topics apply, rank higher the most specific topic that best captures the text's theme and intent.
- Topical salience: Consider how prominently the topic appears within the text. Mentions in the main clause, key phrases, or hashtags indicate higher salience than those in peripheral or supporting contexts.
- Topical engagement: Identify signs of engagement with the topic, such as opinions, evaluations, or claims. If the text is expressing a stance or elaborating on a topic it should rank higher.

Remark: If and only if the post provides no information about the subject or issue of the petition, then insert a single top_topic in the output with topic "Democracy", domain "Freedom and Democracy", and set generic to "true".

```
Input format:
{
"id": {string},
"text": {string}
}

Output format:
{
"top_topics": [
{
"rank": {integer},
"topic": {string},
"domain": {string}
},
{
"rank": {integer},
"topic": {string},
"domain": {string}
},
{
"rank": {integer},
"topic": {string},
"domain": {string}
}
],
generic:{binary:true/false}
}
```

LLM vs ManifestoBERTa annotations alignment

Here follow the summary results and classification performance metrics (Accuracy, Precision, Recall, F1) of the validation exercise exploring the alignment of ManifestoBERTa annotations with LLM-based ones, for a sample of 500 randomly selected posts.

All posts in the sample were labelled by the LLM with at least one issue (the 1° ranked one); 117 were labeled with at least two issues (the 1° and 2° ranked one), and 117 with three issues (the 1°, 2° and 3° ranked one).

The top-1 accuracy is 0.686, the top-2 accuracy is 0.802, and the top-3 accuracy is 0.846. This highlights that in about 17 cases out of 20, the three top-ranked issues inferred by ManifestoBERTa contain the “true” (LLM-based) label.

Across the sample, the mean intersection size between the LLM top-k issue set ($k \leq 3$) and ManifestoBERTa's top-3 issue set is 1.04 labels.

The results reported in the tables below indicate that overall agreement between LLM-based annotations and ManifestoBERTa is reasonably strong, as reflected by relatively high weighted precision, recall, and F1 scores. However, weighted metrics are largely driven by high-frequency issue classes. In contrast, performance for rare issue classes is lower and more volatile. For these low-support categories, even a small number of disagreements between LLM and ManifestoBERTa (1° ranked issue) annotations can lead to sharp drops in precision, recall, and F1 scores. This pattern is partly a mechanical consequence of the limited sample size, as several rare classes will appear only once in the evaluation set, such that any mismatch results in a precision (or recall) of zero.

Table S.10. Classification performance metrics (N=500).

Accuracy	Avg. Precision	Avg. Recall	Avg. F1	Weighted Precision	Weighted Recall	Weighted F1
0.686	0.326	0.321	0.514	0.786	0.686	0.716

Note. LLM 1° issue label by post is considered the ground truth, and ManifestoBERTa's issue label is considered the predicted label.

If we exclude all posts containing only generic calls for participation/mobilization/signing, the classification performance metrics are the following:

Table S.11. Classification performance metrics excluding generic calls for participation/mobilization/signing (N=169), as identified by the LLM annotator.

Accuracy	Avg. Precision	Avg. Recall	Avg. F1	Weighted Precision	Weighted Recall	Weighted F1
0.515	0.439	0.342	0.585	0.585	0.515	0.494

Note. LLM 1° issue label by post is considered the ground truth, and ManifestoBERTa's issue label is considered the predicted label.

The per-class performance metrics, shown below, reveal heterogeneity in the alignment between (1° ranked) LLM-based annotations and ManifestoBERTa across issue categories. This variation is strongly associated with class prevalence and conceptual specificity. We remind the reader that these patterns also underscore the limitations of per-class evaluation in small samples. As for classes

with limited support, performance metrics are highly sensitive to few annotation mismatches and should therefore be interpreted cautiously.

Table S.12. Classification performance metrics per class, considering only issue classes with a support of at least 5 observations (1% of the random sample).

Issue label	Support	Precision	Recall	F1
Democracy	339	0.960	0.785	0.864
Environmental Protection	24	0.594	0.792	0.679
Political Corruption	21	0.519	0.667	0.583
Law and Order	19	0.500	0.211	0.296
Political Authority	16	0.175	0.438	0.250
Welfare State Expansion	14	0.417	0.714	0.526
Freedom and Human Rights	12	0.438	0.583	0.500
Anti-Imperialism	8	1.000	0.125	0.222
Internationalism	8	0.400	0.750	0.522

Note. LLM 1° issue label by post is considered the ground truth, and ManifestoBERTa's issue label is considered the predicted label.

S.5. Consistency of annotations by petition URL

To assess whether ManifestoBERTa produces consistent top-ranked issue labels across posts referring to the same petition-related URL, we compute and examine, in Figure 3, the empirical distribution of the share of posts whose 1° ranked issue label matches the modal 1° label for that URL (x-axis), plotted against the proportion of the number of unique URLs (y-axis).

Because multiple posts often link to the same URL, variation in rank one issue assignments within a petition allows us to determine whether the model behaves consistently / systematically across repeated observations of the same underlying latent object (the petition).

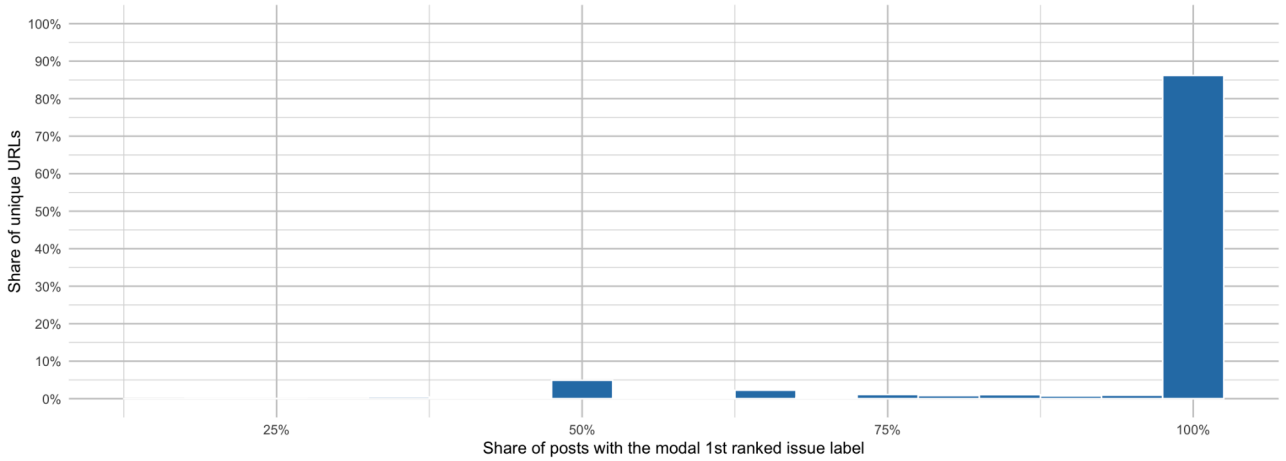

Figure S.3. Histogram of agreement on the top-ranked issue label by URL, considering only URLs with n.posts ≥2.

Note. The x-axis represents the share of posts, by URL, whose 1° ranked label equals that URL's modal rank one label. The y-axis shows the proportion of URLs falling in that label agreement share, with bin widths of 5%. URLs in the 100% on the x-axis indicate perfect agreement of 1° ranked issue labels across all posts referring to the same URL. Lower values reflect dispersion of 1° ranked issues across posts referencing the same URL.

While the model may exhibit miscalibration in specific score intervals, leading to over- or under-estimation of issue propensity scores, the distribution shown in Figure S.3 demonstrates that its top-ranked issue assignments are highly consistent within petitions. The histogram indicates that

approximately 86% of URLs fall into the]95%, 100%] agreement bin, meaning that every post associated with those petitions receives (almost always) the same top-ranked issue label. The remaining URLs are almost entirely concentrated in the 50–75% agreement range. Importantly, no substantial mass appears in the lower bins, and the distribution is sharply skewed towards the higher bin, representing 95%+ agreement. These results highlight the consistency of ManifestoBERTa annotations in our domain of application, providing strong evidence that the model produces outputs that are stable for different posts, and therefore different texts, referring to the same petition-related URL. While it is true that the model may exhibit miscalibration issues for some score intervals, occasionally over- or under-estimating issue scores, the observed consistency supports its reliability for the purposes of this work.

S.6. Petition platform perplexity formula and computational details

To quantify the diversity of petitions referenced through posts, for each petition platform we compute the perplexity of the empirical distribution of petition referencing. Our rationale in choosing this metric stems from its intuitive interpretation: A distribution that has a perplexity value of x is one that has the same entropy as one in which the distribution is uniform, but over a number of options or categories equal to x. The diagonal in Figure 6 (see Manuscript) allows readers to better understand deviations from this uniform distribution.

Let our dataset contain n_j distinct URLs (references to petitions) related to the petition platform j's domain, indexed by $i = (1, \ldots, n_j)$. For each unique URL i, we count the number of posts $c_i^{(j)}$ referring to it, and the total number of posts $C^{(j)}$ referring to the platform j. Where:

$$C^{(j)} = \sum_{i=1}^{n_j} c_i^{(j)}$$

We then compute the relative frequency distribution of unique URLs referring to the petition platform j. Where the relative frequency $x_i^{(j)}$ of the URL i is defined as:

$$x_i^{(j)} = \frac{c_i^{(j)}}{C^{(j)}}, \qquad i = 1, \ldots, n_j$$

Accordingly, the empirical probability distribution of unique URLs referring to platform j is:

$$P_j = \left(x_1^{(j)}, x_2^{(j)}, \ldots, x_{n_j}^{(j)} \right)$$

The Shannon entropy of this distribution is:

$$H(P_j) = -\sum_{i=1}^{n_j} x_i^{(j)} \log_2 x_i^{(j)}$$

The Perplexity of this distribution is:

$$\text{Perplexity}(P_j) = 2^{H(P_j)} = 2^{-\sum_{i=1}^{n_j} x_i^{(j)} \log_2 x_i^{(j)}}$$

For an overview of the perplexity metric, we refer the reader to the work by Ramaciotti et al. (2021).

S.7. Supplementary Results

Unique combinations of users and domains

In Table S.13, we show counts of User-Domain (UD) combinations by language. In the first row of the table, we see the total number of unique UD pairs, i.e., unique combinations of users and domains, while the second row shows the ratio of these unique pairs to the total number of posts for each language. This ratio, referred to as the UDN ratio, indicates how concentrated or dispersed the petition-related posting activity is: low scores, such as those observed for French, point to a scenario where relatively few users and domains account for a large share of the posts, whereas higher scores, such as those seen for multilingual platforms, suggest a broader distribution of posts among more diverse users and domains population.

Table S.13. Number of unique users-domains pairs (UD pairs) and UDN ratio by language.

Platform Language	Dutch	French	German	Italian	Polish	Spanish	Romanian	Multilingual
N. unique User-Domain pairs (UD)	1 371	18 294	17 759	566	637	12	2	41 433
UD pairs / N. posts (UDN ratio)	49.30%	14.11%	44.77%	44.71%	48.70%	50%	66.67%	53.86%

Table S.14 shows the size and share of the user population that has mentioned a specific number of distinct petition platform domains, with columns representing the number of domains a user has engaged on X. This table reveals that around 94% of users concentrate their petition-related posting activity on a single petitioning platform, while only about 6% engage with more than one platform.

Table S.14. Number of domains in query mentioned.

N. domains in query mentioned	1	2	3	4	5	6	7	8
N. users (% tot)	70 720 (94.20%)	3 775 (5,03%)	530 (0.71%)	43 (0.06%)	7 (>0.01%)	0 (>0.01%)	1 (>0.01%)	0 (>0.01%)

Distribution of users by the number of distinct petitions and topics mentioned

Figure S.4, shows the distribution of users by the number of distinct petitions mentioned in their posts and number of issues discussed. We remind the reader that issues have been inferred using ManifestoBerta and that we consider here only the most probable base issue for each post, meaning that users who have posted only once can discuss just one issue. This simplification is necessary to project users' mobilization patterns in these two dimensions in a more straightforward manner.

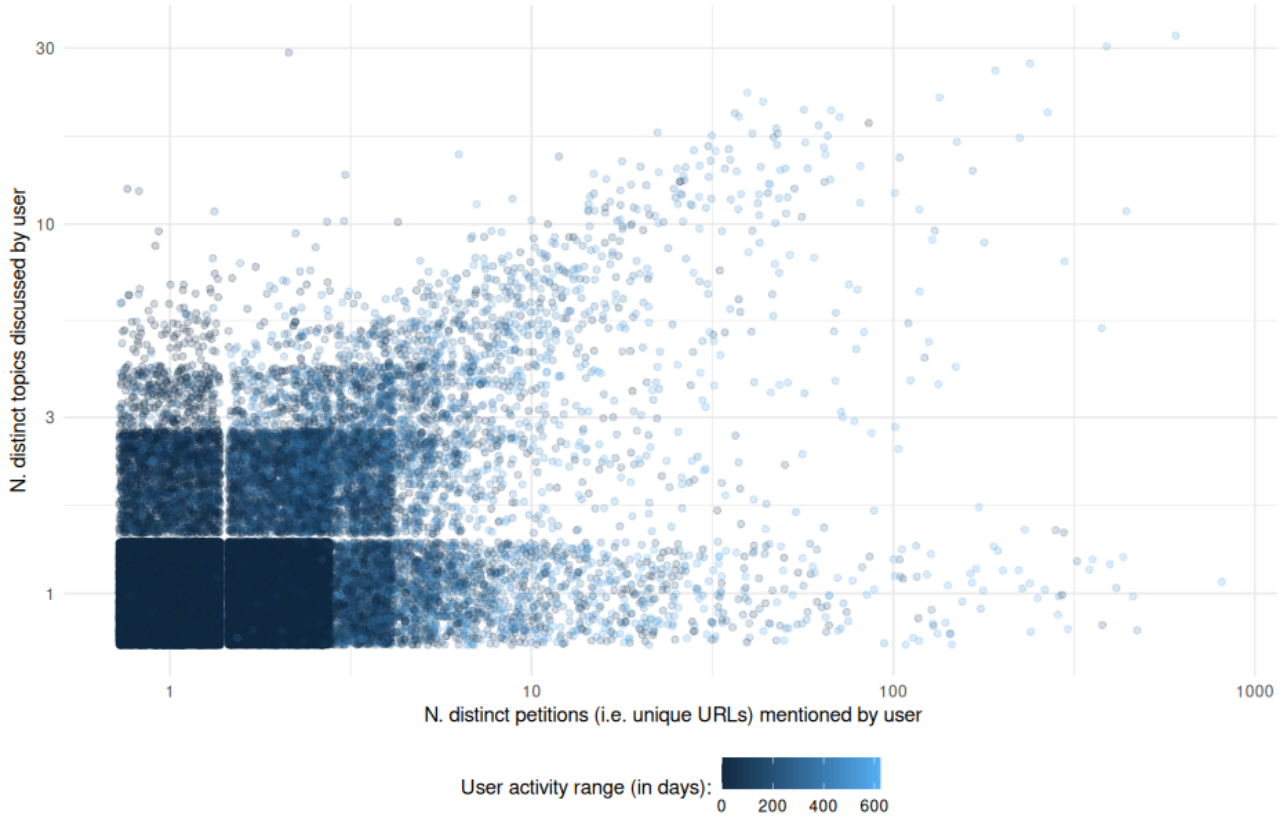

Figure S.4. Jittered scatter plot of the number of distinct issues discussed by a user as a function of the number of distinct petitions mentioned by the user.

Note. The number of distinct issues discussed by a user is measured as the number of distinct top-scoring issues (i.e., the issue with the highest weighted confidence score) in their posts inferred through the ManifestoBERTa model. The

number of distinct petitions mentioned by a user corresponds to the number of unique URLs. The color scale represents the user's activity range in days. Jitter parameters are set to 0.14 for both axes. Axes have been log-transformed to compress extreme values.

The figure shows a clear concentration of users in the lower left area, where most discuss only one or a few distinct issues and mention only one or a few petitions. This suggests that among social media users engaging with petition-related issues on X, the majority participate on a very limited scale, focusing on a small number of issues and petitions. As the number of distinct petitions increases, there is a gradual spread in the number of distinct issues discussed, but the overall distribution remains constrained, indicating that even among users who engage with many petitions, few diversify their discussions significantly. The color scale, representing user activity range in days, shows that users who remain active for longer periods tend to be more distributed across higher values on both axes, meaning they are more likely to engage with a broader range of petitions and issues.

Overall, the figure suggests that activism on X is largely issue-specific and that most users do not engage in a broad or sustained manner. However, those with longer engagement periods tend to participate in discussions that cover a wider range of issues and petitions, though such users remain a minority. Finally, there is a small yet non-negligible fraction of users who engage with a single issue across many distinct petitions over long time frames, aligning with the New Lobbyists user category pattern first described by Jungherr and Jürgens (2010).

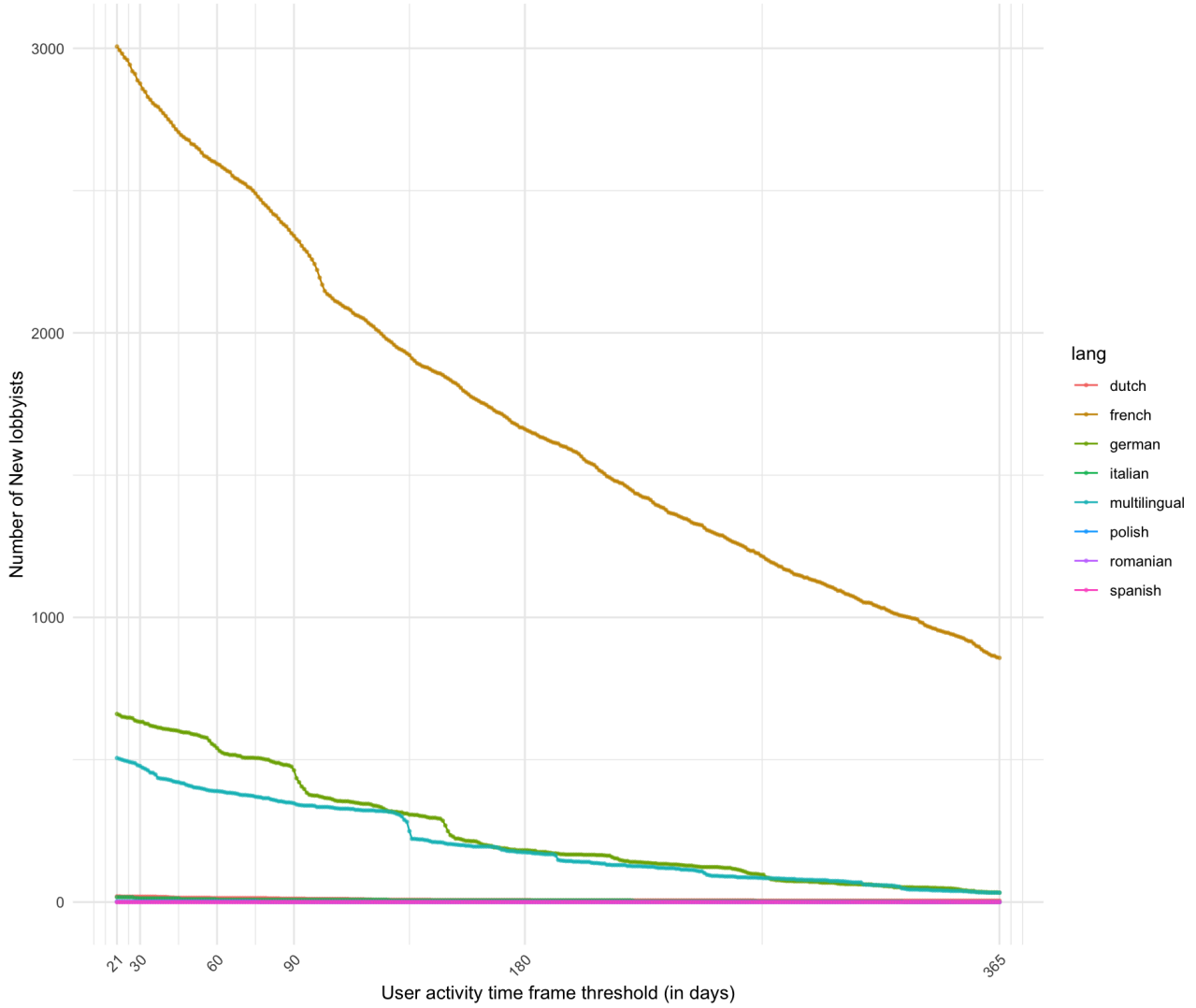

Figure S.5. Number of users classified as new lobbyists for different user-activity threshold values.

Figure S.5 shows that the markedly higher number of users classified as New lobbyists among French-language users is not an artifact of the activity-threshold choice. Across the full range of user-activity thresholds evaluated (from 21 days to 365 days), both the absolute counts (and also the proportion) of French-language users identified as new lobbyists remains consistently and substantially greater than those observed for other languages.

References

Burst, Tobias, Pola Lehmann, Simon Franzmann, Denise Al-Gaddooa, Christoph Ivanusch, Sven Regel, Felicia Riethmüller, Bernhard Weßels, and Lisa Zehnter. ManifestoBERTa. Wissenschaftszentrum Berlin für Sozialforschung (WZB) / Göttingen: Institut für Demokratieforschung (2023). https://doi.org/10.25522/manifesto.manifestoberta.56topics.sentence.2023.1.1

Jungherr, Andreas, and Pascal Jürgens. "The political click: political participation through E‐Petitions in Germany." *Policy & Internet* 2, no. 4 (2010): 131-165.

Ramaciotti Morales, Pedro, Robin Lamarche-Perrin, Raphaël Fournier-S'Niehotta, Rémy Poulain, Lionel Tabourier, and Fabien Tarissan. "Measuring diversity in heterogeneous information networks." *Theoretical computer science 8*59 (2021): 80-115.

Ramaciotti Morales, Pedro, Jimena Royo-Letelier, Jean-Philippe Cointet, and Armin Pournaki. Attitudinally-positioned European sample dataset. HAL working paper (2024) https://hal.science/hal-04807916v1/document