Towards More Transparent Online Campaigning: Detecting Political Campaign Content in Election-related Social Media Posts

Abdullah Alabdullah[1], Conor Gaughan[2], Thomas Flavel[3], Shubhanjay Varma[2], Rachel Gibson[2], Marta Cantijoch Cunill[2], Alexandru Cernat[2] and **Riza Batista-Navarro[4]**
[1]School of Informatics, University of Edinburgh, United Kingdom
[2]School of Social Sciences, University of Manchester, United Kingdom
[3]School of Arts, Languages and Cultures, University of Manchester, United Kingdom
[4]Department of Computer Science, University of Manchester, United Kingdom
Correspondence: riza.batista@manchester.ac.uk

Abstract

A large part of political campaigns during elections is now being conducted online, with political actors leveraging their networks on social media platforms. To maintain transparency in political communications, regulations applicable to online campaigning have been put in place in many democracies. While it should be straightforward for voters to determine who produced and funded online advertisements comprising paid political campaigns, it is much more challenging to detect if organic content, i.e., social media posts, pertains to political campaigning, due to possibly subtle yet suggestive language that can be used by certain actors. In this paper, we investigate the feasibility of automatically detecting whether a given tweet posted by a political actor pertains to political campaigning, and if yes, whether it was conveyed in a direct or indirect (subtle) manner. After establishing an annotation scheme for the task of detecting political campaign content in tweets, we fine-tuned three encoder models (BERT, BERTweet and PoliBERTweet) for the same task and evaluated their performance. Our results show that fine-tuning BERTweet leads to the best macro-averaged F1-score (0.776), although all models consistently struggle to detect indirect campaigning.

1 Introduction

Our society is now entering a "fourth era" of political campaigning, defined as a data-driven, digital-first approach using hyper-personalised micro-targeting and networked communication via social media (Magin et al., 2017; Römmele and Gibson, 2020; Strömbäck, 2008). One platform which has been at the forefront of online political campaigning since 2010 is Twitter (now X), enabling political actors to directly communicate with the electorate (Vergeer, 2015). Across numerous national elections, major party candidates have strategically employed Twitter to disseminate political messages and influence public opinion (Jungherr, 2016).

On the one hand, these political messages can be overtly campaigning (e.g. *"Get out and vote Democrat on November 3rd!"* or *"If he's elected, Corbyn will destroy this country"*). On the other hand, many of them can be more subtle, not directly telling the reader to vote one way or the other but indirectly suggesting it via the language that they employ (e.g. *"Climate change is the most pertinent threat to our survival"* or *"We need an economy that works for the many, not just a wealthy few"*).

Automated detection of online political campaigning is crucial not only for academic research, but also for effective regulation. Electoral law in many countries now requires complete transparency around online campaigning by political parties and politicians, including registry requirements and the use of digital imprints. Examples of these include the 2022 UK Elections Act 2022, the 2025 EU Political Advertising Regulation, Section 325 of the Canada Elections Act, and U.S. Federal Election Commission (FEC) disclaimer rules governing paid online political advertisements on social media and digital platforms. These are applied most strictly to paid campaign advertising but can also apply to many other types of digital material including organic content: tweets and other types of social media posts. Identification of text-based online campaigning is not necessarily straightforward, especially when the messaging is implicit rather than explicit.

With the overarching aim of boosting transparency and trust in political messaging on social media, our work seeks to enable the detection of political campaign content in online posts at scale, by developing new natural language processing (NLP) models for automatically classifying text according to whether it pertains to political campaigning or not. Our contributions include:

- A conceptual framework for capturing political campaign content in election-related

Proceedings of the 7th Workshop on Natural Language Processing and Computational Social Science (NLP+CSS), pages 249–270

tweets, underpinned by an annotation decision tree and guidelines; we report the results of applying this framework on the annotation of tweets by humans and a large language model;

- The development of three baseline transformer-based encoder models fine-tuned for the task of detecting political campaign content; the performance and comparison of these models were systematically evaluated and compared, leading to the selection of the best-performing baseline model which was then applied at scale to US 2020 election tweets.

2 Online Political Campaign Content

Given that this work evaluates models on a US dataset, it would be intuitive to derive our definition of online political campaign content from US online campaigning laws. However, despite similar transparency principles around paid online advertising being conveyed through the Federal Election Commission (FEC) and Federal Election Campaign Act (FECA) disclaimer rules (Fowler et al., 2020), the US framework remains less centralised and less comprehensive than the UK or EU regimes.

Therefore, we construct our framework around statutory guidance laid out by the UK Electoral Commission in accordance with Section 54 of the UK Elections Act 2022 (Electoral Commission, 2026). The benefit of using this framework is that it not only applies to paid digital material, but also to "organic" material that has not been paid for but is political and published by a relevant entity such as a registered party or candidate.

As such, we adopt the following working definition for *online political campaign content* – any digital material posted by political parties or candidates which can be reasonably regarded to influence the public to give support to or withhold support from:

(1) one or more political parties
(2) a candidate or future candidate
(3) an elected office holder
(4) political parties, candidates, future candidates or elected office-holders that are linked by their support for, or opposition to, particular policies, or by holding particular opinions
(5) other categories of candidates, future candidates or elected office-holders that are not based on policies or opinions, e.g., candidates who went to a state school, or Members of Parliament (MPs) who grew up in their constituency.

Here, we distinguish between two forms of campaigning: direct and indirect. Where a party or candidate encourages or discourages support for another political party, candidate or elected office holder (points 1-3 above) by directly mentioning them, we refer to this as **direct campaigning**. Where a party or candidate encourages or discourages support for another political party, candidate or elected office holder through reference to linked policies, opinions or characteristics (points 4-5), we refer to this as **indirect campaigning**. Where content cannot be reasonably regarded to influence the public to support or withhold support from another political party, candidate or elected office holder, irrespective of whether they are directly named or not, we refer to this as **non-campaigning**.

Distinguishing between these three categories is important for flagging digital content posted by political actors that are attempting, either directly or indirectly, to influence the public. Political actors have both a moral and, in many countries, legal obligation to be transparent about campaigning, and this might not always be obvious to voters.

3 Related Work

Computational models for enhancing transparency and accountability in online political communication have been proposed in the past, although majority were concerned with paid advertisements. Sosnovik et al. (2023) collected political advertisements from Facebook during the French election period in 2022 and categorised them according to policy categories using a classification model, while Yoshikawa and Roesner (2025) manually analysed political advertisements shown on news and media websites during the 2024 US elections. To the best of our knowledge, the only work that set out to address the task of automatically detecting political campaign content in organic content (i.e., posts) in social media platforms is that of Achmann-Denkler et al. (2024), which analysed Instagram captions and stories posted during the German elections in 2021, according to whether they contain a "Call to Action" (CTA) – a message that mobilises readers to take specific actions.

4 Task Formulation

In this work, we address the automated detection of political campaign content in tweets authored by political figures during elections. We cast this problem as a text classification task. That is, given a tweet that is known to have come from a political actor, an annotator should label it with one of the three classes in our annotation scheme, namely, `non-campaigning`, `direct_campaigning`, and `indirect_campaigning`, depending on whether it contains campaign content, and if yes, the type of campaigning used.

The next section describes how we constructed a dataset consisting of 27,620 English-language tweets posted by 75 political figures relevant to the 2020 US presidential election. This dataset supports us in our objective to build robust classifiers for campaign content detection, and to systematically assess how domain adaptation, first to Twitter language and then to political and electoral discourse, affects the detection of political campaigning content.

5 Dataset Construction

In this section, we present the steps that we carried out in order to develop a reliable dataset for training and evaluating our text classification models for detecting political campaign content.

5.1 Collection of Social Media Posts

This study uses an in-house dataset of social media posts collected during the 2020 US presidential election campaign period, that was constructed as part of the Digital Campaigning and Electoral Democracy (DiCED) project.[1] It consists of 196,012 tweets from 75 verified political figures, including presidential and vice-presidential candidates and Democratic and Republican senators. For a full list of all political figures involved, see Appendix F.

This dataset was chosen for its focus on official actors, which reduces ambiguity in communicative intent. It thus captures campaign-related discourse more clearly than other existing public datasets (e.g., #Election2020; Chen et al., 2021), which primarily reflect general political discussion rather than campaign communication.

To capture peak campaign activity, we carried out data filtering and included only tweets posted within six weeks before and in the week of the election day (the 3rd of November 2020). After removing predominantly Spanish tweets, our final dataset consists of 27,620 tweets. We consider this dataset to be sufficient in terms of size, given that it is larger than those utilised in prior studies focused on fine-tuning BERT-based encoder models for political tweet classification, such as the work by Grimminger and Klinger, 2021 (3,000 tweets) and Baran et al., 2022 (6,112 tweets). We split the dataset into training (21,984; ~80%), development (2,745; ~10%), and test (2,750; ~10%) sets.

5.2 Annotation Protocol Development

The central challenge of this study is to operationalise the concept of political campaigning that we defined in Section 2 into a multi-level labelling scheme that can be applied consistently by both human annotators and state-of-the-art large language models (LLMs) to detect campaigning intent in tweets. This task is non-trivial, as political tweets are typically short, noisy and often express persuasive intent implicitly rather than through explicit messaging (Vijayaraghavan et al., 2021). To address this, we employ a structured annotation protocol underpinned by a decision tree (DT) and detailed annotation guidelines (see Appendix A).

Our hierarchical decision tree, depicted in Figure 1, decomposes the annotation task into a sequence of binary decisions reflecting increasing levels of interpretive judgment. The DT represents a logical process that a human annotator follows to classify tweets into their respective campaigning classes, rather than a learned DT. Tweets are first evaluated for direct references to political entities (e.g. parties or candidates). If such a reference exists, the annotator determines whether the tweet encourages or discourages support for that entity, resulting in a classification of `direct_campaigning` (`dir_camp`) or `non-campaigning` (`non-camp`). If no political entity is referenced, the tree assesses whether the tweet refers to a related characteristic, such as a policy, ideology, opinion or attribute of a political figure. Tweets that reference such characteristics and express evaluative or persuasive sentiment are classified as `indirect_campaigning` (`ind_camp`). This structure ensures that labels are mutually exclusive and exhaustive while maintaining a clear distinction between direct and indirect campaigning by separating the assessment of entity reference and the discussion of linked characteristics.

[1] https://sites.manchester.ac.uk/diced/

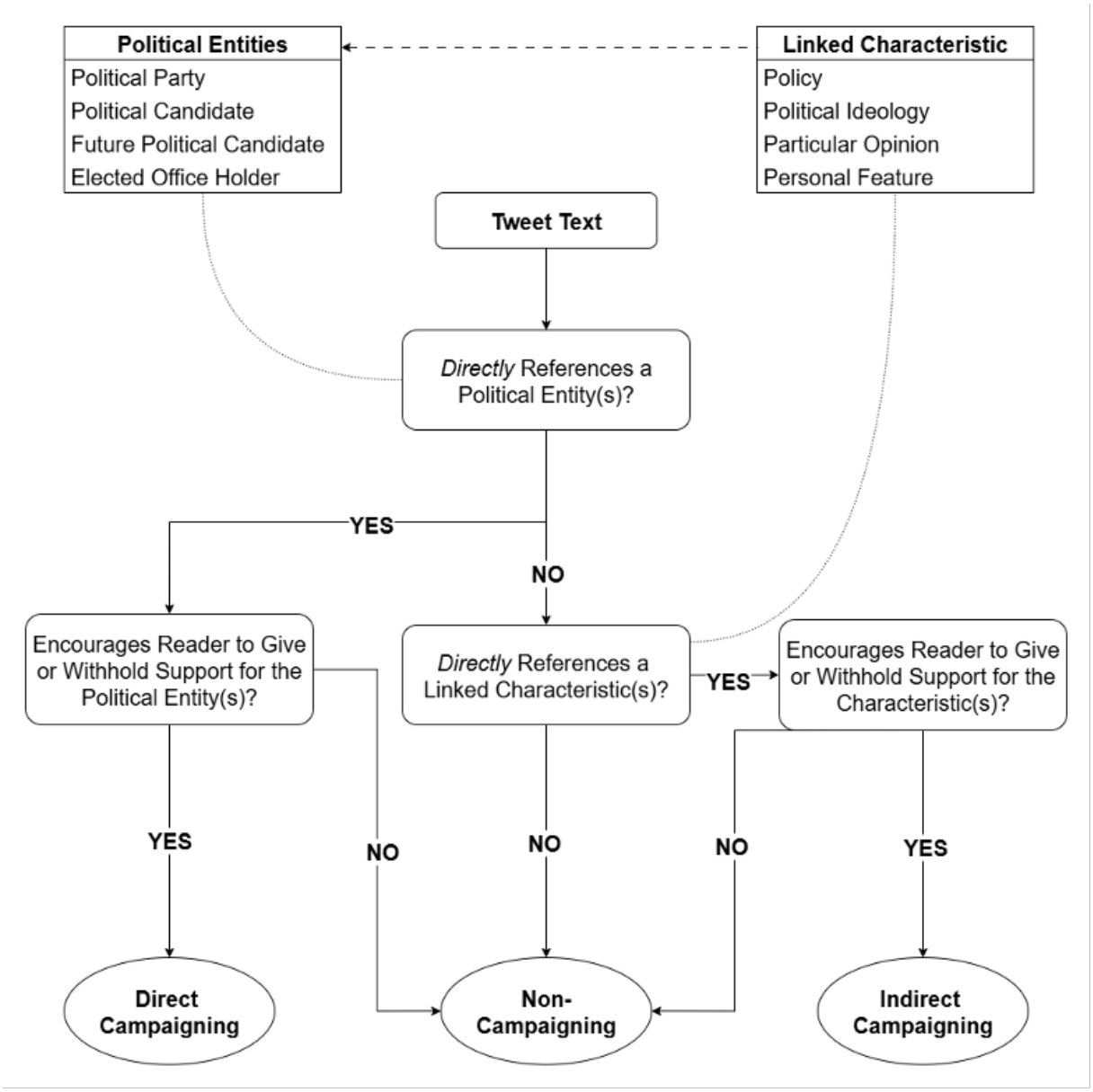

Figure 1: The decision tree guiding the classification of any given tweet according to our annotation scheme.

Our annotation guidelines complement the decision tree by formalising each binary decision with explicit definitions, decision rules and boundary conditions for common sources of ambiguity, such as self-referential language. They specify how political entities and related characteristics are identified within the text of a given tweet and how evaluative language is distinguished from neutral description when assessing persuasive intent.

5.3 Data Annotation

To validate our annotation protocol prior to large-scale labelling using an LLM, we conducted a two-round pilot exercise to assess inter-annotator agreement (IAA) and identify sources of ambiguity. Given the cost and irreversibility of full-scale annotation, this step ensured that the decision tree and guidelines yielded consistent and reliable labels across annotators.

A random sample of 200 tweets was selected for the pilot. In each round, two political science specialists independently annotated 100 tweets following a brief training session with worked examples. Beyond assigning labels, annotators provided structured feedback on the clarity of the DT, instances of hesitation or backtracking, and ambiguities. After each round, the protocol was refined based on class-level agreement patterns and annotator feedback. After the final round, the standard (unweighted) Cohen's κ IAA score between the two human anno-

tators reached 0.57, indicating moderate agreement (Landis and Koch, 1977).

In addition to human annotation, we employed an LLM as a third expert annotator to enable scalable dataset labelling. To ensure consistency with human annotators, the decision tree and annotation guidelines were adapted into a structured prompting scheme, with each decision node represented as a separate prompt. For reproducibility, we provide all prompts in Appendix C.

Among several locally hosted models evaluated, `Llama-3.3-70B-Instruct` achieved the best alignment with human annotations (on the same 100 randomly sampled tweets), attaining an overall accuracy (matching labels) of 0.67 and a Cohen's κ of 0.50.

The agreement between the LLM and one human annotator (κ: 0.50) was comparable to that between the two human annotators (κ: 0.57). Our human–human IAA is lower than that reported by (Grimminger and Klinger, 2021) (human-human Cohen's κ: 0.61–0.88 for stance detection on 2020 US election tweets). It is, however, worth noting that stance detection relies on explicit evaluation towards named candidates, yielding clearer class boundaries, whereas our task in some cases requires inferring persuasive intent without direct political references, making it inherently more subjective. These results indicate that our LLM-generated annotations are sufficiently reliable for large-scale training set labelling.

The final annotation strategy combined human and LLM annotations: The training set was fully annotated by the LLM; the validation set was annotated by one human annotator; and the test set was annotated by two humans with LLM annotations only used for majority vote in cases where the two human annotators disagree.

6 Methodology

6.1 Models and Evaluation Strategy

We fine-tune and evaluate three transformer-based encoders: BERT (Devlin et al., 2019), BERTweet (Nguyen et al., 2020) and PoliBERTweet (Kawintiranon and Singh, 2022). This allowed us to examine how progressive domain adaptation, from general English to Twitter, and then to electoral discourse, affects model performance on the task of campaign content detection.

BERT (110M parameters), which follows the BERT-base architecture and is pre-trained on large-scale English corpora, serves as our general-purpose baseline, though its pre-training data does not reflect the stylistic features of Twitter or political discourse. BERTweet retains the BERT-base architecture but follows the RoBERTa pre-training procedure and is further trained on 850M English tweets, effectively modelling the informal syntax, abbreviations, and platform-specific conventions of Twitter. BERTweet was shown to outperform general-purpose models on tasks such as sentiment analysis of tweets (Nguyen et al., 2020). Meanwhile, PoliBERTweet further pre-trains BERT-base on 83M tweets from the US 2020 presidential election, incorporating election-specific vocabulary and campaigning rhetoric. It has shown strong performance on politics-focused tasks, such as stance detection in relation to presidential candidates (Kawintiranon and Singh, 2022), which, like campaigning detection, requires nuanced interpretation of evaluative and persuasive language.

Our dataset exhibits class imbalance, with the sample of 100 tweets (from the second round of our pilot exercise) showing uneven distribution for `dir_camp` (54%), `ind_camp` (29%) and `non-camp` (17%). As all classes are equally important, we adopt macro-averaged F1-score (macro F1) as the primary metric for model selection and evaluation, ensuring equal weighting across classes. We additionally report per-class metrics for diagnostic analysis and use Cohen's κ to contextualise model performance relative to human–human and human–LLM agreement. This comparison distinguishes genuine model limitations from apparent underperformance due to annotation ambiguity. Given the interpretative nature of the task, model performance can be compared against observed human agreement rather than an assumed noise-free gold standard.

6.2 Training Configurations

All our fine-tuned models were trained under identical configurations. We set `hidden_dropout_prob` = 0.1, `attention_probs_dropout_prob` = 0.1 and `classifier_dropout` = 0.1. Parameter-efficient fine-tuning was applied using LoRA, with rank ($r = 16$), scaling factor ($\alpha = 32$) and dropout of 0.1 applied to the `query`, `key` and `value` projection matrices. LoRA was employed in lieu of full fine-tuning for computational efficiency, as it substantially reduces the number of trainable parameters without significant degradation in downstream task performance (Hu et al., 2021).

Tokenisation was performed using each model's corresponding tokeniser. The maximum sequence length was set to 104 tokens, which was decided by rounding up the 99th percentile of our tokenised sequence lengths to the nearest multiple of 8 for efficient batching. Truncation was enabled and dynamic padding was applied.

To address class imbalance, we applied oversampling using `WeightedRandomSampler`, where class weights were computed as the inverse class frequency, resulting in higher sampling probability for rarer classes. Optimisation was performed using AdamW with a learning rate of 5×10^{-5}, linear scheduling and a warmup ratio of 0.06 followed by linear decay. The batch size was set to 128, with a maximum of 15 training epochs. Early stopping was based on the macro F1 score, with a patience of 4 epochs and a threshold of 1×10^{-4}. Weight decay was set to 0.01, and all experiments were run with a fixed random seed (17) on an NVIDIA A100-SXM4-80GB GPU.

6.3 Hyperparameter Optimisation

While most of our configurations used the recommended values, we experimented with different learning rates, sampling techniques, and cross-entropy loss calculation methods. All hyperparameter optimisation experiments used our human-annotated validation set. We evaluated learning rates in the range of 5×10^{-5} to 3×10^{-4}, with 5×10^{-5} yielding the best validation macro F1 score and was thus selected for all experiments.

Performance on the `ind_camp` class was consistently lower than other classes. We therefore experimented with class-weighted cross-entropy, but this did not improve results. Since loss re-weighting only adjusts gradient magnitudes after batches are formed, batches dominated by the majority class still contained relatively few minority examples per update step. We also explored re-weighting the loss based on measured LLM–human annotation reliability, but this also did not yield improvements, likely because the class performance gap was primarily driven by overlapping decision boundaries rather than uniform label noise that could be mitigated through global weight adjustments. We therefore applied minority class over-sampling during training, where sampling weights were computed as the inverse of class frequency.

7 Results and Discussion

7.1 Comparison of Models

In this section, we compare our different pre-trained transformer models that were fine-tuned to assess the impact of domain-specific pre-training on campaign content detection and to establish a baseline for future work on this task. Table 1 shows the performance of the fine-tuned models on the validation set.

Model	Macro F1
BERT	0.684
BERTweet	**0.703**
PoliBERTweet	0.685

Table 1: Model performance comparison on the validation set.

PoliBERTweet performs almost identically to the base BERT model, while BERTweet outperforms both alternatives. An analysis of per-class performance in Table 2 shows that BERTweet achieves higher scores across all three classes, indicating that its overall ranking is robust and not driven by skewed performance on a single class. Its largest advantage emerges in the `ind_camp` class, suggesting more consistent identification of this more challenging boundary class.

BERTweet outperforms PoliBERTweet despite the latter being pre-trained on a larger set of political tweets. This difference can be explained by variations in pre-training scale, initialisation and training procedure. BERTweet was pre-trained on 850 million English tweets using the RoBERTa pre-training procedure, which includes dynamic masking, removal of next sentence prediction, large batch training and byte-level BPE tokenisation. In contrast, PoliBERTweet was initialised from BERT-base and further pre-trained on approximately 5 million English tweets related to the 2020 US Presidential Election. Although PoliBERTweet benefits from political domain specificity, BERTweet's substantially larger tweet corpus and more robust pre-training strategy likely provide stronger general representations of the language used in Twitter. These advantages appear to outweigh the benefits of political topic specialisation after fine-tuning.

To assess the impact of domain-specific pre-training in a zero-shot setting, we compare the performance of raw BERTweet and PoliBERTweet on the validation set, without fine-tuning (see Table 5 in Appendix D). BERTweet was able to produce

reasonable predictions only for the non-camp class, achieving perfect recall (P = 0.31, R = 1.00), but fails to detect dir_camp and ind_camp, resulting in near-zero precision and recall for those categories. This bias towards non-camp is consistent with the predominantly non-political nature of the general Twitter corpus on which BERTweet was pre-trained, which makes it insensitive to campaign-specific language without fine-tuning.

In contrast, PoliBERTweet (pre-trained on political election tweets) demonstrates substantially stronger zero-shot detection of campaign-related content, particularly for the ind_camp class, where it achieves high recall (R = 0.97). However, it performs poorly on non-camp tweets (R = 0.05), indicating a bias towards predicting political content, likely due to its specialised pre-training on US election data. Where a tweet is not classified as dir_camp, the model defaults to ind_camp. This is likely due to its specialised pre-training on US election data, which likely resulted in limited exposure to non-camp instances, causing the model to default to ind_camp in the absence of clear dir_camp signals.

Overall, these results suggest that political domain pre-training provides a clear advantage in zero-shot campaign detection, although this advantage diminishes once both models are fine-tuned on task-specific data, as fine-tuning exposes both models to the same campaign-related data.

7.2 Analysis of the Best-performing Model

Table 2 presents the detailed performance of the best-performing model. BERTweet achieves strong results on the dir_camp class (F1 = 0.816). However, ind_camp remains the primary bottleneck (F1 = 0.582), with the model substantially underperforming relative to the other classes. Our error analysis indicates that ind_camp is frequently misclassified as either dir_camp or non-camp, with many instances predicted as the former, confirming the blurred boundaries between categories.

This limitation likely stems less from model capacity and more from taxonomy ambiguity and class imbalance, as ind_camp constitutes only 16.14% of the training data. Moderate performance (F1 = 0.710) is demonstrated for the non-camp class, though some confusion with dir_camp remains. Overall, the main challenge lies in the definition and representation of ind_camp rather than in the model architecture.

Class	Precision	Recall	F1
dir_camp	0.778	0.859	0.816
ind_camp	0.559	0.607	0.582
non-camp	0.810	0.632	0.710
Overall Macro F1		0.703	
Overall Accuracy		0.744	

Table 2: Per-class Precision, Recall, and F1 scores obtained by the fine-tuned BERTweet model on the validation set.

7.3 Hierarchical Classification

To better align the model architecture with the human annotation decision tree, we decompose the task into a two-stage (one-vs-rest) setup and fine-tune two instances of BERTweet. Stage A distinguishes between the campaigning and non-campaigning (non-camp) classes, while Stage B takes the tweets predicted as campaign-related and classifies them as either dir_camp or ind_camp.

Performance on the dir_camp class remains the same as the single classifier, while slight improvement can be observed for non-camp (with macro F1 increasing from 0.53 to 0.58). However, performance on ind_camp dropped substantially (from 0.58 to 0.53 macro F1), with many ind_camp instances misclassified as non-camp.

These findings indicate that structurally mirroring the annotation decision tree does not resolve the persistent confusion surrounding the ind_camp class, supporting our hypothesis that improvement requires annotation scheme refinement and enhancing class-specific representation. Consequently, the single multiclass classifier remains preferable, as it provides better overall performance with lower computational cost.

7.4 Performance on the Test Set

Table 3 presents the macro F1 scores of the three models on the held-out test set. BERTweet achieves the highest performance (0.776), outperforming BERT (0.753) and PoliBERTweet (0.762). The ranking observed during validation is maintained on the test set, indicating stable generalisation and confirming BERTweet as the most effective model for our campaign detection task. As previously noted, BERT and PoliBERTweet show comparable performance.

Per-class analysis of BERTweet results (see Table 4) revealed a consistent pattern: dir_camp is the most reliably detected category (F1=0.882), whereas ind_camp remains the most challenging

Model	Macro F1
BERT	0.753
BERTweet	**0.776**
PoliBERTweet	0.762

Table 3: Model performance comparison on the test set.

Class	Precision	Recall	F1
`dir_camp`	0.951	0.823	0.882
`ind_camp`	0.583	0.831	0.685
`non-camp`	0.709	0.819	0.760

Table 4: Per-class Precision, Recall, and F1 scores obtained by the fine-tuned BERTweet model on the test set.

(F1=0.685). BERTweet attains high recall for `ind_camp` (0.831) but lower precision (0.583), indicating that while most indirect instances are captured, false positives persist due to blurred boundaries with neighboring classes. Although the consistency across the validation and test sets demonstrates robust generalisation, the difficulty with `ind_camp` remains, suggesting the models' limitation in understanding this category.

7.5 Political Campaigning in the 2020 US Presidential Election

We applied our best-performing model at scale on the full US 2020 dataset containing 27,620 tweets, covering the six weeks prior to and the week of the Election Day (3 November 2020). This allowed us to examine role-based and temporal patterns in political campaigning behaviour during the final phase of the election.

In Figure 2, one can observe how the distribution of the tweets (in the full dataset) changed over the six weeks in the run-up to the election (in Week 7). Interestingly, political actors seem to have used direct campaigning at a fairly consistent level in the first four weeks of the six-week period; however, this seems to have declined in the two weeks right before the election. Upon producing the breakdown of the weekly distribution according to role groups (see Figures 3 and 4 in Appendix E, we observed that this decline can be attributed to the presidential and vice-presidential candidates seemingly using much less of direct campaigning and more of indirect campaigning two weeks right before the election.

8 Conclusion and Future Work

In this paper, we describe the development of a new annotation scheme for detecting whether any given social media post (i.e., a tweet) can be considered as political campaigning content. We conducted a two-round annotation exercise that confirms that humans are able to apply the scheme consistently on a subset of tweets, and that an LLM can obtain agreement with a human that is close enough to human-human agreement. Afterwards, we fine-tuned three baseline transformer-based encoder models, namely, BERT, BERTweet and PoliBERTweet, for the political campaign content detection task. The fine-tuned BERTweet model obtained the best performance, with a macro-averaged F1-score of 0.776 on the held-out test set.

Indirect campaigning tweets pose a challenge to all baseline models. Future work will explore the use of more advanced models, as well as data augmentation strategies to enhance the representation of this class. Additionally, we will explore detecting a distinct "Call to Action" class, capturing messages mobilising readers to undertake specific actions, which is a more targeted form of support than direct or indirect campaigning.

Limitations

While the two-round pilot test iteratively refined the taxonomy, the moderate IAA reflects an inherent limitation of the classification scheme: the boundary between `ind_camp` and `non-camp` is partly subjective and could not be fully resolved through improved annotation guidelines alone. We consider the current taxonomy to be a sound conceptual framework grounded in electoral law, but further iterations informed by larger-scale annotation studies are needed to sharpen class boundaries. As a consequence, performance differences between models should be interpreted with caution, as some errors may still reflect taxonomy ambiguity rather than model limitations.

The baseline models we used consistently struggle with the indirect campaigning (`ind_camp`) class, likely due to its limited representation in the training data. We did not explore state-of-the-art generative large language models. This is due to our envisioned end-users of automated tools that detect political campaign content: regulatory bodies, NGOs and not-for-profit organisations, who might not necessarily have access to computational or financial resources required by generative LLMs

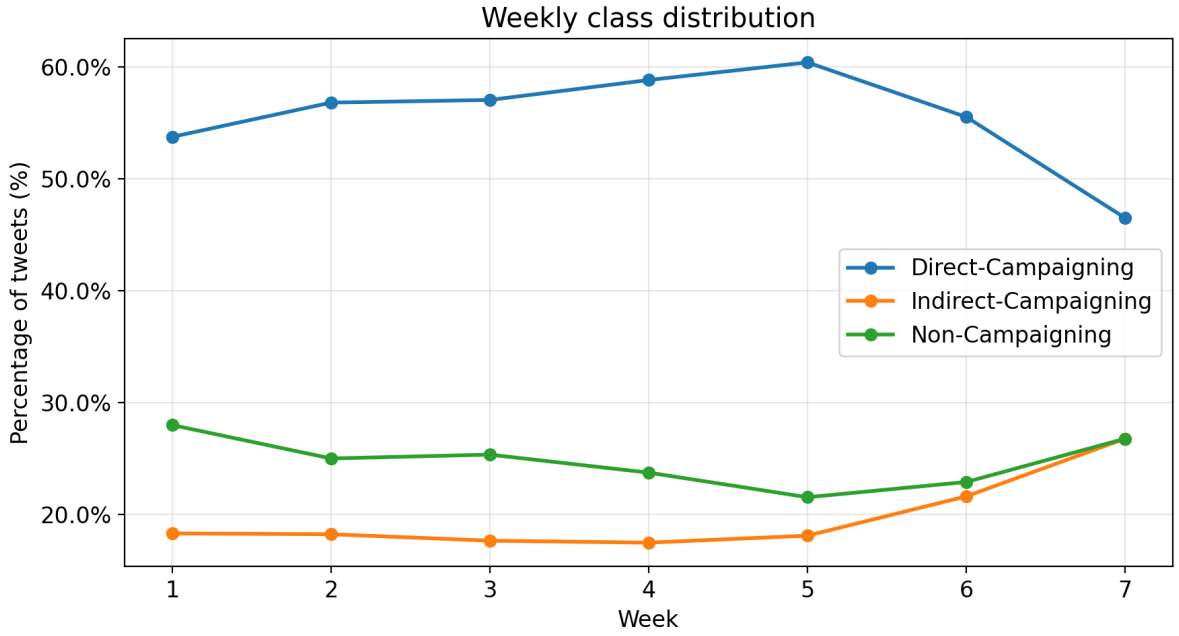

Figure 2: Distribution of the tweets in the weeks right before the US 2020 election, as classified by the best-performing model (fine-tuned BERTweet). Election Day is in Week 7.

in analysing large social media datasets. Lastly, the performance of our best-performing baseline model has not been evaluated across different countries, election cycles or election types beyond the US 2020 presidential elections. Evaluating performance under these settings would help further assess the robustness and generalisability of our proposed approach. We also encourage future work to experiment with using more advanced models like

Ethics statement

The dataset of tweets used in this study was collected as part of a previous project under terms and conditions that prevent us from sharing the data with other researchers. To provide other researchers with insights on how our proposed annotation scheme can be consistently applied, we provided examples of tweets from public, political figures in our annotation guidelines (see Appendix A and B).

References

Michael Achmann-Denkler, Jakob Fehle, Mario Haim, and Christian Wolff. 2024. Detecting calls to action in multimodal content: Analysis of the 2021 German federal election campaign on Instagram. In *Proceedings of the 4th Workshop on Computational Linguistics for the Political and Social Sciences: Long and short papers*, pages 1–13, Vienna, Austria. Association for Computational Linguistics.

Mateusz Baran, Mateusz Wójcik, Piotr Kolebski, Michał Bernaczyk, Krzysztof Rajda, Lukasz Augustyniak, and Tomasz Kajdanowicz. 2022. Electoral agitation dataset: The use case of the Polish election. In *Proceedings of the LREC 2022 workshop on Natural Language Processing for Political Sciences*, pages 32–36, Marseille, France. European Language Resources Association.

Emily Chen, Ashok Deb, and Emilio Ferrara. 2021. election2020: the first public twitter dataset on the 2020 us presidential election. *Journal of Computational Social Science*, 5(1):1–18.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *Preprint*, arXiv:1810.04805.

Electoral Commission. 2026. Statutory guidance on digital imprints.

Erika Franklin Fowler, Michael M Franz, and Travis N Ridout. 2020. Online political advertising in the united states. *Social media and democracy: The state of the field, prospects for reform*, pages 111–138.

Lara Grimminger and Roman Klinger. 2021. Hate towards the political opponent: A Twitter corpus study of the 2020 US elections on the basis of offensive speech and stance detection. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 171–180, Online. Association for Computational Linguistics.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *Preprint*, arXiv:2106.09685.

Andreas Jungherr. 2016. Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics*, 13(1):72–91.

Kornraphop Kawintiranon and Lisa Singh. 2022. PoliBERTweet: A pre-trained language model for analyzing political content on Twitter. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 7360–7367, Marseille, France. European Language Resources Association.

J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Michael Magin, Nicole Podschuweit, Jan Haßler, and Uwe Russmann. 2017. Campaigning in the fourth age of political communication: A multi-method study on the use of facebook by german and austrian parties in the 2013 national election campaigns. *Information, Communication & Society*, 20(11):1698–1719.

Dat Quoc Nguyen, Thanh Vu, and Anh Tuan Nguyen. 2020. Bertweet: A pre-trained language model for english tweets. *Preprint*, arXiv:2005.10200.

Andrea Römmele and Rachel K. Gibson. 2020. Scientific and subversive: The two faces of the fourth era of political campaigning. *New Media & Society*, 22(4):595–610.

Vera Sosnovik, Romaissa Kessi, Maximin Coavoux, and Oana Goga. 2023. On detecting policy-related political ads: An exploratory analysis of meta ads in 2022 french election. In *Proceedings of the ACM Web Conference 2023*, WWW '23, page 4104–4114, New York, NY, USA. Association for Computing Machinery.

Jesper Strömbäck. 2008. Four phases of mediatization: An analysis of the mediatization of politics. *The International Journal of Press/Politics*, 13(3):228–246.

Mark Vergeer. 2015. Twitter and political campaigning. *Sociology Compass*, 9(9):745–760.

Prashanth Vijayaraghavan, Soroush Vosoughi, and Deb Roy. 2021. Automatic detection and categorization of election-related tweets. *Proceedings of the International AAAI Conference on Web and Social Media*, 10(1):703–706.

Emi Yoshikawa and Franziska Roesner. 2025. Exploring political ads on news and media websites during the 2024 u.s. elections. *Preprint*, arXiv:2503.02886.

Appendix

A Annotation Guidelines

STEP 1

BINARY QUESTION: Does the tweet text directly reference a political entity?

• This strictly includes political parties, political candidates, future/prospective political candidates, and elected office holders (e.g. MPs, Mayors, Councillors, Governors, Senators, Prime Ministers, Presidents, etc.)
• The entity(s) must be directly named in the text, and this can also include @mentioned user accounts and relevant hashtags. For example, #trump2020 or #DemsLose would classify as direct references for Donald Trump and the Democratic Party respectively.
• Where specific political entities are not refenced directly by name but by their specific role/title (e.g. "Prime Minister", "the President", "POTUS", "Governor for X"), this classifies as a direct reference.
• It does not matter how many entities are referenced as long as there is at least one.
• Make sure to double-check where possible if the referenced entity meets one of the four entity categories. Former political candidates and elected office holders (at the time of the post) do not account as a direct reference.
• There are certain cases where the use of self-referential pronouns may also count as a direct reference, but this is contextual based on the language of the text. Given that every account in the tweet data was a US senator, where they refer to themselves as "I" "Me", "My" this counts as a direct reference to a political entity. E.g.
"I am the best person for the job. Lend me your vote and I promise it will not go to waste!"
"I have fought against the growing tide of fascism for over decade, and I will not give in today. No matter how hard my opponents try to slander to me."
"My job is to make sure that the voices of ordinary people are heard when it comes to concerns over immigration."
• Where they use plural self-referential pronouns like "we" or "us" or "our", whether this classifies as a direct reference is contingent on the context they are used. If it appears they are using the term in reference to their political party, this will count as a direct reference. E.g.
"We are the party on the side of the people; they are the party of the rich and corrupt."
"If you give us your vote this November, we promise not to let you down!"
"We are working tirelessly to improve transport up and down the country."
• However, if it used in reference to society or to make an appeal to the general population more broadly, this would not count as a direct reference. E.g.
"We deserve better as a nation. This is not what America stands for." "We should not accept this new policy proposal from our Government; we need to do more to help protect the poorest communities in our society."
"It's up to us to save the planet, together, united."

If answer to STEP 1 is YES: proceed to STEP 2.A
If answer to STEP 1 is NO: proceed to STEP 2.B

STEP 2.A

BINARY QUESTION: Does the tweet text encourage the reader to give to or withhold support from the directly reference entity(s)?
• Encouraging/discouraging support for a political entity can be explicit: e.g.

"The only way to secure a fairer future is with the Green Party. Join us this election—vote Green, vote for change!"
"Our community needs strong leadership. That's why I'm voting for Sarah Thompson on May 5th. She's the only candidate who will fight for working families. ThompsonForMayor"
"President Alvarez has delivered on jobs, healthcare, and education. Let's keep building on that progress—re-elect Alvarez this November!"
• Or encouraging/discouraging support for a political entity can be more suggestive based on sentiment: e.g.
"Hard not to notice how much better things have gotten since Mayor Patel took office. Feels like someone's finally listening to us."
"Sure, let's give the Democrats that caused the housing crisis another term. What could possibly go wrong?"
"Another broken promise from the Conservatives today... starting to lose count. Guess some things never change."
• It can also include instances where the author is attempting to elicit support via non-political or non-partisan dimensions like appeals to family, religious, national or cultural values:
"There is nothing more important to me in this world than my faith. I am proud to represent all devoted Hindus up and down the country."
"As a true American patriot, Senator Green is committed to upholding the values embedded in our constitution."
"The Republicans claim to be the party of family, but really they're the party of greed and self-interest."

If answer to STEP 2.A is YES: This is DIRECT CAMPAIGNING
If answer to STEP 2.A is NO: This is NON-CAMPAIGNING

STEP 2.B

BINARY QUESTION: Does the tweet text directly reference a linked characteristic of a political entity(s)?
• This includes policies, political ideologies, or political opinions, as well as other categories not based on policies or opinions such as sociodemographics or personal qualities.
• The referenced characteristic should be reasonably attributable to one of the four political entities:
• References to policies can include:
o Explicitly named policies like "Online Safety Act 2023" or the "One Big Beautiful Bill Act 2025".
o Policy areas like "healthcare", "crime", "foreign aid", "welfare", "education", "gun ownership", "human rights", "taxation" and so on.
• References to political ideologies can include:
o Explicitly named ideologies like "socialism", "conservatism", "capitalism", "liberalism", "populism", "fascism" and their variations.
o Or general political leanings like "left-wing", "right-wing", "centrists" "far-right", "far-left" and their variations.
o It can also include broader references to the facets of an ideology such as "free markets", "small government", "personal liberties" and so on.
• References to political opinions can include:
o Subjective takes, judgments, reactions, values, or feelings without direct reference to a policy or political entity. I.e.
"It seems that politics over the last few years has become more about bickering than dealing with pressing issues facing this country."

"The country is more divided now than ever before"
"The media seems more interested in pushing an agenda than actually covering the issues that people care about"
"Well, that debate was a complete joke. No wonder people are losing faith in the system!"
• References to personal features or characteristics can include:
o Explicit socio-demographic or personal features like gender, ethnicity, age, educational background, geographic location or occupation.
o Or more vague references to positive and negative traits about certain groups. E.g.
"The country is being ran into the ground by a small group of corrupt individuals who think they are above the law"
"99% of the population are hardworking, honest, and decent individuals. It's the 1% that are the problem"
"There are some people in politics who take their jobs seriously, and others who see it as a chance to progress their own careers."
• It does not matter how many characteristics are referenced as long as there is at least one.
• Characteristics can either be directly referenced within the tweet text or via relevant hashtags. For example, BuildTheWall or PlayIn4Climate can be considered as support for Trump's policy to build a wall along the US-Mexico border and support for reduction in air pollution respectively. Forthe99% or AgainstCapitalists would be considered support references to personal characteristics and an ideology.

If answer to STEP 2.B is NO: This is NON-CAMPAIGNING
If answer to STEP 2.B is Yes: proceed to STEP 3

STEP 3

BINARY QUESTION: Does the tweet text encourage the reader to give to or withhold support from the refenced characteristic(s)?
• Encouraging/discouraging support for a linked characteristic can be explicit: e.g.
"Support the Clean Air for Schools Act — every child deserves classrooms free from toxic pollution."
"Do not stand with those who dismiss the struggles of others — empathy is essential in public life."
"Stand behind the Affordable Homes Guarantee Bill to ensure safe housing is within everyone's reach."
"Reforms to vital medical provisions for the elderly is cruel and callous."
• Or encouraging/discouraging support for a characteristic can be more suggestive based on sentiment: e.g.
"Free markets shouldn't be allowed to decide everything — community values must matter too."
"Our right to the freedom of expression is crucial to everything this country stands for. It needs to be protected at all costs."
"Politics is dominated by private school kids who have no idea about the real world. We need more people who know about the real world."
"Veterans fought for this country and now they have been thrown to wayside. Don't let their sacrifices be in vain."

If answer to STEP 3 is YES: This is INDIRECT CAMPAIGNING
If answer to STEP 3 is NO: This is NON-CAMPAIGNING

B Example Tweets for All Decision Tree Outcomes

Pathway 1: Direct Campaigning

Directly References an Entity(s): YES
Tweet Encourages to Give/Withold Support: YES

"Senator Davidson is a fantastic leader and an even better person. The people of Arizona are lucky to have him!"
"I have always stood for working people and I always will"

Pathway 2: Direct Reference, Not Campaigning (Non-Campaigning)

Directly References an Entity(s): YES
Tweet Encourages to Give/Withold Support: NO

"I have scheduled a live Q&A session on Twitch tonight at 7pm"
"Senator Grey has passed away this evening, aged 78."

Pathway 3: Indirect Campaigning

Directly References an Entity(s): NO
Directly References Linked Characteristic: YES
Tweet Encourages to Give/Withold Support: YES

"Rolling back protections in the Digital Privacy Reform Act puts people at risk. That change should not go unchallenged. @PrivacyWatch #DataRights"
"Communities thrive when solidarity is put ahead of unchecked markets. Let's not lose sight of those values. #FairEconomy"

Pathway 4: Indirect Reference, Not Campaigning (Non-Campaigning)

Directly References an Entity(s): NO
Directly References Linked Characteristic: YES
Tweet Encourages to Give/Withold Support: NO

"The Green Transport Act was passed in 2021."
"The National Housing Strategy includes five chapters."

Pathway 5: No Reference (Non-Campaigning)

Directly References an Entity(s): NO
Directly References Linked Characteristic: NO

"Merry Christmas to all those who celebrate!"
"Anyone know why the Twitter server was down this afternoon?"

C LLM Annotation Prompts

system prompt STEP 1

You are an impartial and precise text classification model trained to detect whether a tweet (that is written by a political figure) directly references a political entity. You must read the tweet carefully, apply the rules below, and respond only with "YES" or "NO", no extra text.

Rules:
A political entity includes:
- Political parties
- Political candidates
- Future or prospective political candidates
- Elected office holders (e.g. MPs, Mayors, Councillors, Governors, Senators, Prime Ministers, Presidents, etc.)

Answer YES if any of the following conditions are met:

1) The tweet explicitly names a political entity. Example: "Democrats," "Donald Trump," "President Biden."
2) The tweet includes an @mention or hashtag referring to a political entity. Example: @JoeBiden, VoteLabour, trump2020, DemsLose
3) The tweet refers to a political entity by role or title. Example: "The Prime Minister", "the President", "Governor of Texas", "POTUS", "Governor for X".
4) The tweet author (who is always a political entity) refers to themselves using singular pronouns ("I", "me", "my"). Example: "I am honored to serve the people of Ohio", "I am the best person for the job. Lend me your vote and I promise it will not go to waste!".
5) The author uses plural pronouns ("we," "our," "us") in reference to their political party or campaign. Example: "We are the party of progress," "If you give us your vote this November, we won't let you down."
Answer NO in all other cases, including but not limited to the following cases:
- The tweet refers to "we," "our," or "us" in a societal or national sense. Example: "We deserve better as a nation", "It's up to us to save the planet, together, united.".
- The tweet mentions "government," "Congress," or similar institutions without specifying a political entity or title. Example: "The government should do more to protect the environment."

Respond strictly with "YES" or "NO"

system prompt STEP 2.A

You are an impartial and precise text classification model trained to detect whether a tweet (that directly references a political entity) encourages or discourages support for that entity. You must read the tweet carefully, apply the rules below, and respond only with "YES" or "NO", no extra text.

Rules:
A tweet encourages or discourages support for a political entity when it expresses, implies, or suggests that readers should:
- Give support to the referenced political entity (e.g. vote for, endorse, trust, or praise them), OR
- Withhold support from the referenced political entity (e.g. criticize, oppose, or reject them).

Answer YES if any of the following conditions are met:
1) The tweet explicitly urges the reader to support or reject the political entity.
Example: “Vote Green this election!”, “Re-elect Alvarez this November!”, “Our community needs strong leadership. That’s why I’m voting for Sarah Thompson.”
2) The tweet praises or criticizes a political entity in a way that clearly implies endorsement or opposition.
Example: “President Alvarez has delivered on jobs, healthcare, and education. Let’s keep building on that progress—re-elect Alvarez this November!”, “Another broken promise from the Conservatives today... starting to lose count.”
3) The tweet expresses sentiment or opinion that can reasonably be interpreted as encouraging or discouraging support for the entity.
Example: “Hard not to notice how much better things have gotten since Mayor Patel took office.”, “Sure, let’s give the Democrats that caused the housing crisis another term. What could possibly go wrong?”
4) The tweet appeals to non-political or cultural values (e.g. family, religion, patriotism) to elicit support or opposition for the political entity.
Example: “There is nothing more important to me in this world than my faith. I am proud to represent all devoted Hindus up and down the country.”, “As a true American patriot, Senator Green is committed to upholding the values embedded in our constitution.”, “The Republicans claim to be the party of family, but really they’re the party of greed and self-interest”

Answer NO in all other cases, including but not limited to the following:
- The tweet only provides factual or neutral information about the political entity without encouraging or discouraging support.
Example: “The President has confirmed that the annual conference will be held in NewYork this September.”, “Senator Grey has passed away this evening.”
- The tweet refers to a political entity in a ceremonial or administrative context without evaluative language.
Example: “The President met with politications in California to discuss trade policy.”

Respond strictly with “YES” or “NO”

system prompt STEP 2.B

You are an impartial and precise text classification model trained to detect whether a tweet (that does not directly reference a political entity) directly references at leaast one linked characteristic of a political entity. You must read the tweet carefully, apply the rules below, and respond only with “YES” or “NO”, no extra text.

Rules:
A tweet directly references a linked characteristic of a political entity if it mentions or implies a policy, political ideology, political opinion, or personal/sociodemographic characteristic that can be reasonably attributed to one or more political entities.

Answer YES if any of the following conditions are met:
• The tweet includes a reference to policies, political ideologies, or political opinions, as well as other categories not based on policies or opinions such as sociodemographics or personal qualities.
• The referenced characteristic should be reasonably attributable to one of the four political entities.
• Characteristics can either be directly referenced within the tweet text or via relevant hashtags. For example, BuildTheWall or PlayIn4Climate can be considered as support for Trump’s policy to

build a wall along the US-Mexico border and support for reduction in air pollution respectively. Forthe99% or AgainstCapitalists would be considered support references to personal characteristics and an ideology.

Explanation:
• References to policies can include:
o Explicitly named policies like “Online Safety Act 2023” or the “One Big Beautiful Bill Act 2025”.
o Policy areas like “healthcare”, “crime”, “foreign aid”, “welfare”, “education”, “gun ownership”, “human rights”, “taxation” and so on.
• References to political ideologies can include:
o Explicitly named ideologies like “socialism”, “conservatism”, “capitalism”, “liberalism”, and their variations.
o Or general political leanings like “left-wing”, “right-wing”, “centrists” “far-right”, and their variations.
o It can also include broader references to the facets of an ideology such as “free markets”, “small government”, “personal liberties” and so on.
• References to political opinions can include:
o Subjective takes, judgments, reactions, values, or feelings without direct reference to a policy or political entity. Examples:
“It seems that politics over the last few years has become more about bickering than dealing with pressing issues facing this country.”
“The media seems more interested in pushing an agenda than actually covering the issues that people care about”
“Well, that debate was a complete joke. No wonder people are losing faith in the system!”
• References to personal features or characteristics can include:
o Explicit socio-demographic or personal features like gender, ethnicity, age, educational background, geographic location or occupation.
o Or more vague references to positive and negative traits about certain groups. Examples:
“The country is being ran into the ground by a small group of corrupt individuals who think they are above the law”
“99“There are some people in politics who take their jobs seriously, and others who see it as a chance to progress their own careers.”

Answer NO in all other cases, including but not limited to the following:
- The tweet does not contain any reference to a policy, ideology, opinion, or personal/sociodemographic characteristic.
- The tweet is purely personal, social, or unrelated to politics or governance.

Respond strictly with “YES” or “NO”

system prompt STEP 3

You are an impartial and precise text classification model trained to detect whether a tweet (that directly references a linked characteristic of a political entity) encourages or discourages support for that characteristic. You must read the tweet carefully, apply the rules below, and respond only with “YES” or “NO”, no extra text.

Rules:
A tweet encourages or discourages support for a linked characteristic when it expresses, implies, or

suggests that readers should:
- Give support to the referenced characteristic (e.g. agree with, promote, or defend it), OR
- Withhold support from the referenced characteristic (e.g. criticize, oppose, or reject it).

Answer YES if any of the following conditions are met:

• Encouraging/discouraging support for a linked characteristic can be explicit: e.g.
"Support the Clean Air for Schools Act — every child deserves classrooms free from toxic pollution."
"Do not stand with those who dismiss the struggles of others — empathy is essential in public life."
"Stand behind the Affordable Homes Guarantee Bill to ensure safe housing is within everyone's reach."
"Reforms to vital medical provisions for the elderly is cruel and callous."
• Or encouraging/discouraging support for a characteristic can be more suggestive based on sentiment: e.g.
"Free markets shouldn't be allowed to decide everything — community values must matter too."
"Our right to the freedom of expression is crucial to everything this country stands for. It needs to be protected at all costs."
"Politics is dominated by private school kids who have no idea about the real world. We need more people who know about the real world."
"Veterans fought for this country and now they have been thrown to wayside. Don't let their sacrifices be in vain."

Answer NO in all other cases, including but not limited to the following:
- The tweet merely states or describes a characteristic without encouraging or discouraging support.
- The tweet provides factual or neutral information about a policy, ideology, or social characteristic without evaluative or persuasive language.

Respond strictly with "YES" or "NO"

D Raw Model Performance

Model	Class	Precision	Recall	F1
BERTweet	dir_camp	0.000	0.000	0.000
	ind_camp	0.000	0.000	0.000
	non-camp	0.312	1.000	0.476
PoliBERTweet	dir_camp	0.125	0.001	0.003
	ind_camp	0.181	0.967	0.305
	non-camp	0.315	0.055	0.093

Table 5: Per-class Precision, Recall, and F1 scores obtained on the validation set by the raw BERTweet and PoliBERTweet models, i.e., without any fine-tuning.

E Visualisation of Results from the Application of the Best-performing Model at Scale

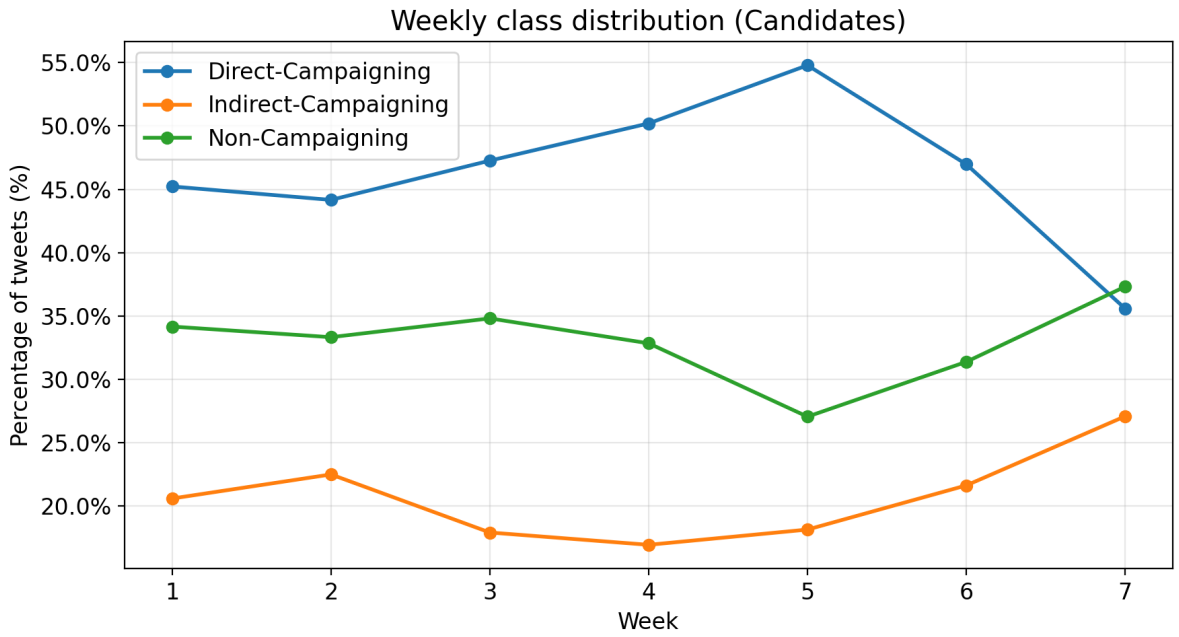

Figure 3: Distribution of the tweets from presidential and vice-presidential candidates in the weeks right before the US 2020 election, as classified by the best-performing model (fine-tuned BERTweet). Election Day is in Week 7.

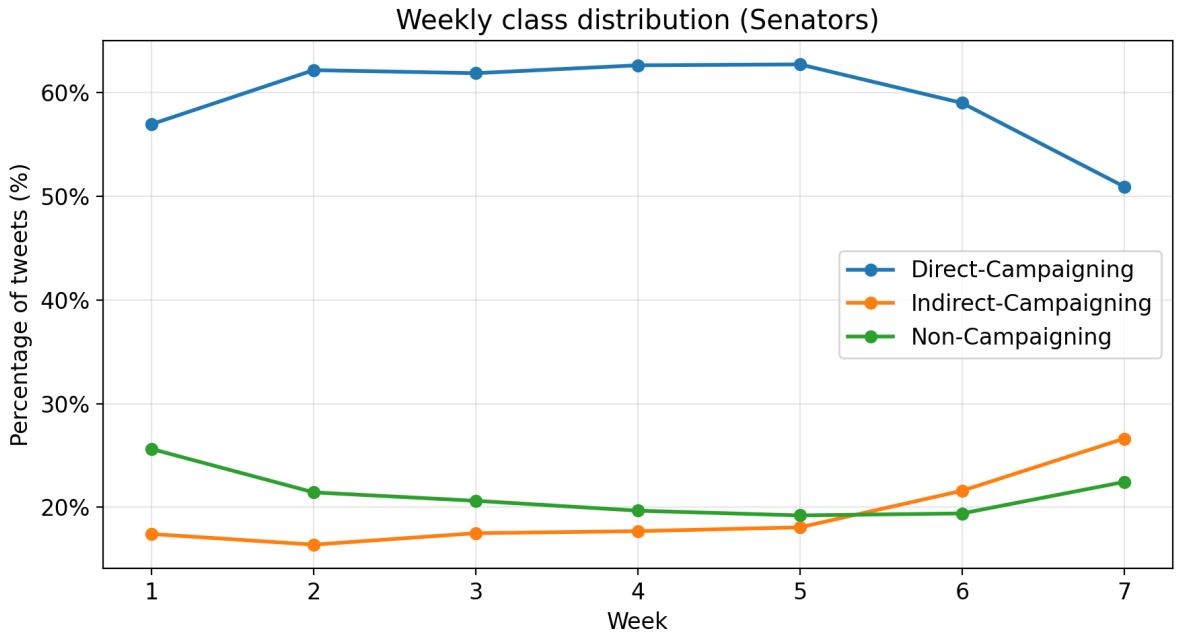

Figure 4: Distribution of the tweets from Democratic and Republican senators in the weeks right before the US 2020 election, as classified by the best-performing model (fine-tuned BERTweet). Election Day is in Week 7.

F Political Figures in the Dataset

Account Holder	Twitter Account	Position
Joe Biden	@JoeBiden	Presidential Candidate
Donald Trump	@realDonaldTrump	Presidential Candidate
Howie Hawkins	@HowieHawkins	Presidential Candidate
Jo Jorgensen	@Jorgensen4POTUS	Presidential Candidate
Kamala Harris	@KamalaHarris	Vice-Presidential Candidate
Mike Pence	@Mike_Pence	Vice-Presidential Candidate
Angela N. Walker	@AngelaNWalker	Vice-Presidential Candidate
Spike Cohen	@RealSpikeCohen	Vice-Presidential Candidate

Table 6: Twitter accounts of presidential and vice-presidential candidates included in the dataset.

Account Holder	Twitter Account
Abby Broyles	@abbybroyles
Dan Ahlers	@ahlers_dan
Amy McGrath	@AmyMcGrathKY
Barbara Bollier	@BarbaraBollier
Marquita Bradshaw	@Bradshaw2020
Cal Cunningham	@CalforNC
Mark Kelly	@CaptMarkKelly
Chris Coons	@ChrisCoons
CJ Ellison	@CJSenate2020
Cory Booker	@CoryBooker
Al Gross	@DrAlGrossAK
Paulette Jordan	@electpaulette
Theresa Greenfield	@GreenfieldIowa
Jaime Harrison	@harrisonjaime
John Hickenlooper	@Hickenlooper
Mark Warner	@MarkWarner
Monica Ben-David	@MBenDavid2020
Mike Espy	@MikeEspyMS
MJ Hegar	@mjhegar
Jon Ossoff	@ossoff
Paula Jean Swearengin	@paulajean2020
Ben Ray Luján	@repbenraylujan
Raphael Warnock	@ReverendWarnock
Ricky Torres	@RickyForSenate
Sara Gideon	@SaraGideon
Dick Durbin	@SenatorDurbin
Jeanne Shaheen	@SenatorShaheen
Doug Jones	@SenDougJones
Gary Peters	@SenGaryPeters
Jack Reed	@SenJackReed
Jeff Merkley	@SenJeffMerkley
Ed Markey	@SenMarkey
Tina Smith	@SenTinaSmith
Steve Bullock	@stevebullockmt

Table 7: Twitter accounts of Democratic senators included in the dataset.

Account Holder	Twitter Account
Bill Hagerty	@BillHagertyTN
Bryant “Corky” Messner	@CorkyForSenate
Cynthia Lummis	@CynthiaMLummis
Steve Daines	@DainesforMT
Daniel Gade	@gadeforvirginia
Jim Inhofe	@JimInhofe
John James	@JohnJamesMI
Kevin O’Connor	@KOCforSenate
Lauren Witzke	@LaurenWitzkeDE
Jason Lewis	@LewisForMN
Lindsey Graham	@LindseyGrahamSC
Mark Ronchetti	@MarkRonchettiNM
Mitch McConnell	@McConnellPress
Shane Perkins	@PerkinsForUSSen
Doug Collins	@RepDougCollins
Rik Mehta	@RikMehta_NJ
Susan Collins	@SenatorCollins
Kelly Loeffler	@SenatorLoeffler
Jim Risch	@SenatorRisch
Mike Rounds	@SenatorRounds
Shelley Moore Capito	@SenCapito
Cory Gardner	@SenCoryGardner
Dan Sullivan	@SenDanSullivan
David Perdue	@sendavidperdue
Cindy Hyde-Smith	@SenHydeSmith
Joni Ernst	@SenJoniErnst
Martha McSally	@SenMcSallyAZ
Ben Sasse	@SenSasse
Thom Tillis	@SenThomTillis
Tom Cotton	@SenTomCotton
John Cornyn	@TeamCornyn
Tommy Tuberville	@TTuberville
Joan Waters	@watersforsenate

Table 8: Twitter accounts of Republican senators included in the dataset.