Supplemental Material A

The Codebook for Political Information Identification

Units of Analysis

Coding unit: The title of the content card

Coding Categories

1 - Political

0 - Non-political

Definition

Titles relevant to issues of public concern were considered political. These not only include references to political figures, groups involved in political decisions, policies, elections, news events, and specific political events (e.g., impeachment inquiry, or the primaries), but also sociopolitical issues such as community activities, acusations of sexual assault, the regulation of large tech companies, issues related to racial, gender, sexual, ethnic, and religious idenitity, as well as crimes with significant social impact such as guns, shootings, corruption, large-scale fraud, or other crimes with significant social impact.

General Coding Instructions

1. Focus on the title only.
    a. Read each title twice before coding.
    b. Do not read the linked article to determine your coding.
2. Exception for unknown names. If a title contains a person's name or an organization's name you do not recognize, you may Google the name or briefly scan the text only to identify who/what they are.

    Example:

    "Bridget Maasland gesloopt in De Telegraaf: 'Ze is een vals kreng!'" – You may check who Bridget Maasland is to decide if the title is political.
3. Titles that mention royal families should be coded based on the context. If the activities are about personal events or family matters, code it as *0 - Non-political*. Otherwise, code

them as *1 - Political,* including (but not limited to) activities related to public affairs, governance, or political engagement.

Example:

"Koning Willem-Alexander en koningin Máxima bezoeken Noord-Groningen - RTV Noord". 1 - Political

"Koningin Máxima trekt alle aandacht in déze pikante trendkleur" *0 - Non-political*

4. Titles involving housing benefits, allowance, pensions, and taxes should be considered *1 - Political*, as they relate to social policy and public welfare systems.

   Example:

   "Pensioenleeftijd niet altijd gelijk aan de AOW-leeftijd: dit betekent het voor jou | Radar - het consumentenprogramma van AVROTROS". 1 - Political

5. If bitcoin news is about market regulation or crisis it might bring to affect society, it should be coded as *1 - Political*; However, if the news is about price analysis, personal stories, trading strategies, and other things irrelevant to the public concern, it should be coded as *0 - Non-political*

   Example:

   "ECB-president Lagarde denkt dat inflatie in 2025 terug is op doelstelling van 2%". 1 - Political

   "Crypto analist verwacht gezonde correctie: 'Daarna weg vrij voor Bitcoin koers van $70.000' - Newsbit" 0 - Non-Political

Supplemental Material B

The Codebook for Manual Coding of Sentiment Level

Units of Analysis

Coding unit: The title of the content card in Google Discover

Introduction

Sentiment level refers to the tonality of politics in this study. Each title will be coded into five categories: *-2 = Predominantly negative, -1 = Slightly negative, 0 = Balanced or neutral, 1 = Slightly positive, 2 = Predominantly positive*. A title is categorized as 'negative' if it delivers political failure, crisis, frustration, or disappointment, whereas a 'positive' tonality reflects political success, achievement, improvement, hope, or gain. A title is coded as neutral if it states facts without making any evaluative judgment, or both positive and negative tonality are evenly balanced.

Please refer to the following coding scheme and principles before you start:

Code	Coding Scheme	Examples
-2/2 Predominantly negative/positive	Very explicit and strong evaluative judgement without any ambiguity.	-2: Harde kritiek op goedkeuring Spot ETFs, Bitcoin koers daalt - Newsbit (Harsh criticism of Spot ETF approval, Bitcoin price drops - Newsbit) -2: Vechten tegen kerncentrales heeft niets uitgehaald, toch gaat Roy Pumfrey door - Omroep Zeeland (Fighting nuclear power plants has achieved nothing, yet Roy Pumfrey continues - Omroep Zeeland) 2: Oekraïense veiligheidsdienst: 'Een van de grootste Russische desinformatie-netwerken ontmanteld' - NRC (Ukrainian Security Service: 'One of the largest Russian disinformation networks

		dismantled' - NRC) 2: China troeft VS af met lancering geheimzinnig ruimtevliegtuig (China trumps US with launch of mysterious space plane)
-1/1 Slightly negative/positive	Evaluative cues are present but limited in salience and strength. They can be more implicit or weaker compared to cues in titles categorized as predominantly negative/positive.	-1: Bouwplannen Maassluis-West baren bewoners zorgen, gemeente gaat gesprek aan | WOS (Maassluis-West construction plans raise concerns for residents) -1: 'Niet meer naar Tilburg komen, het is te druk', zegt gemeente - Omroep Brabant ('Don't come to Tilburg anymore, it's too busy,' says the municipality - Omroep Brabant)
	Evaluative cues are absent, but the events presented are widely considered as negative/positive. - Negative: wars, crises, military escalation, surveillance, etc. - Positive: inflation slows down, victories, reforms, cheaper energy prices, etc.	-1: Russia sends North Korean soldiers to the front line - Joop - BNNVARA *(war)* -1: Weapons for Ukraine: Australia delivers dozens of Abrams tanks *(war)* 1: Must read: House price increases will slow next year | IEX *(stable house price)* 1: Tax authorities want to pay and are looking for people who still need to get money back | Economy | NU.nl

		(pay money back) 1: Spaargeldgarantiefonds voor als bank faalt, is helemaal gevuld - NRC (Savings guarantee fund for bank failure is fully funded - NRC) (*savings are funded*) 1: China releases American pastor imprisoned for nearly 20 years | International | gelderlander.nl *(release pastor)* 1: Vierdaagse werkweek kan juist voor méér productiviteit zorgen: 'Mannen kunnen best wat nutteloze werktijd inleveren' | Financieel | Telegraaf.nl (A four-day workweek can actually lead to greater productivity: "Men can certainly give up some useless work time" | Financial | Telegraaf.nl) *(Greater productivity is not an evaluative cue, but it is a positive trend; You may also argue that "give up useless work time" is an evaluative cue)*
0 Balanced or neutral	No evaluative judgement at all, **and** the events presented are not widely considered as negative/positive	0: Kamer wil opheldering over banden Nederlandse politici en Rusland (Parliament wants clarification about ties between Dutch politicians and Russia)

		0: Harris heeft geoefend tegen de bullebak én de beheerste Trump: kan zij het tv-debat dinsdag winnen? | de Volkskrant (Harris has practiced against the bully and the controlled Trump: can she win Tuesday's TV debate? | de Volkskrant)
	Tonality from both sides is evenly balanced.	0: Solar Magazine - Belastingplan 2025 | Salderen voor eigenaar zonnepanelen voordeliger dan verwacht, energiebelasting minder snel verlaagd (Solar Magazine - 2025 Tax Plan | Netting for solar panel owners more advantageous than expected, energy tax reduction slower) 0: Ambtenaren kunnen hoger pensioen tegemoetzien, zorgpersoneel | Economie | NU.nl (Civil servants can expect higher pensions, healthcare workers are out of luck | Economy | NU.nl)

Principle 1: Note that when there are multiple objects/events in the title, please code the negativity level based on the main object/event if one is dominant. Otherwise, consider all of them.

Example	Reasoning	Code
Example 1: Workplace Drentsche Aa investigates action against 'illegal' snack	This headline contains two events: (1) the investigation conducted by the workplace	0

truck at 't Nije Hemelriek - RTV Drenthe	authority, and (2) the reference to an "illegal" snack truck. While the phrase "illegal snack truck" carries a potentially negative tonality, the primary focus of the headline is the investigation itself. Therefore, the coding should be based on the investigation as the dominant object, rather than the negative undertone of "illegal." Since no evaluative cue is present for this investigation, code it as neutral An exception to this rule would be a case such as *"Jehovah's Witnesses sue the Dutch government for discrimination."* According to the coding logic described above, "A sues/investigates B" would normally be classified as neutral, regardless of what is investigated. However, because B in this example is the government, the event is widely perceived as negative. Therefore, when a government is the object of investigation or the party being sued, the event should be coded as negative.	
Example 2: Civil servants can expect higher pensions, healthcare workers are out of luck | Economy | NU.nl	This headline contains two objects: (1) civil servants, and (2) healthcare workers. Since both are equally important and there is no dominant one, the coding should be based on the	0

	combination of both objects.	
Example 3: Dijkgraaf waterschap: 'Ik had eerder moeten vertellen over salarissen directie Waternet' / Het Parool (Water board dike reeve: 'I should have told you about Waternet management salaries sooner' / Het Parool)	When there are quotes in the titles, first check the dominant event in the title; second, check whether the quote itself is the dominant event or equally important as the main event in the sentence. If yes, take the sentiment of quote in account; if not, don't take the sentiment of quote into account. In this case, the main event is reflected via the quote, focusing on the official's delayed disclosure of salary information, which reflects negatively on the water board's transparency; therefore, the sentiment is coded as –1 (negative).	-1

Principle 2: Code titles based on how the AUTHOR conveys information instead of your own feelings. One exception is that, if the author conveys it neutrally, but it is PREVALENTLY perceived as a positive/negative political event, then code it based on the prevalent perception.

Example	Reasoning	Code
Example 1: China trumps the US with launch of mysterious space plane	Please note that for this title, even though from the US perspective, it can be perceived as negative, what we code here is how the author of the information conveys it.	2
Example 2: Russia sends North Korean soldiers to the	This headline describes the development of the war in a	-1

front line - Joop - BNNVARA	neutral way, but in real life, war is generally associated with negative consequences. Therefore, what we code here is based on the prevalent negative perception of such moves in a war.	

Supplemental Material C

Fine-tuning LLM for Coding Sentiment Level

Data Preparation

Two native Dutch coders manually coded the sentiment level of a subset of 1,000 Dutch titles based on the codebook in Appendix B. The distribution check showed that titles in different categories are imbalanced. There are 536 titles categorized as slightly negative (-1), 273 articles as neutral (0), 155 titles as slightly positive (1), 33 titles as predominantly negative (-2), and three titles categorized as very predominantly positive (2).

If we directly draw 500 titles completely randomly, what we would get is an imbalanced training dataset, where positive political titles would be underrepresented. Fine-tuning Large Language Model (LLM) on such an imbalanced dataset risks biased outputs, as manifested by its bias towards the majority class and poor performance on minority cases (M. S. & S. S., 2025).

To ensure that the model can learn sufficiently during fine-tuning, we constructed a relatively more, even though not entirely, balanced training dataset by setting target sizes for each class and then randomly drew titles until targets were met. Specifically, we took an undersampling approach to adjust the class distribution (M. S. & S. S., 2025), i.e., reducing the majority classes' examples in our dataset to align more closely with the minority ones. In our case, since negative sentiment (-1) is a dominant class, we undersampled it. Besides, to make fine-tuned models differentiate between two dominant classes, neutral sentiment and negative sentiment, more accurately, we set a slightly higher target size for neutral titles before sampling.

Following this procedure, we constructed a training dataset composed of 200 slightly negative titles, 200 neutral titles, 73 slightly positive titles, 25 predominantly negative titles, and 2 predominantly positive titles. The remaining 500 titles formed the final validation set, containing 336 slightly negative titles, 73 neutral titles, 82 slightly positive titles, 8 predominantly negative titles, and 1 predominantly positive title.

Model Fine-tuning

We chose GPT-4.1-2025-04-14 as the base model for fine-tuning and compared both zero-shot and few-shot approaches. We set the batch size to 20, the epoch number to 2 to avoid overfitting, and the learning rate multiplier to 0.5.

The prompt for the zero-shot model is:

Prompt:

Classify the following political titles using the sentiment codebook:

-2: Predominantly negative - strong, explicit negative judgment.

-1: Slightly negative — weak or implicit negative cues, or neutral tone describing events widely perceived as negative (e.g., wars, crises, inflation).

0: Neutral — factually presented without evaluative judgment, or includes both positive and negative elements but they balance each other.

1: Slightly positive — weak or implicit positive cues, or neutral tone describing events widely perceived as positive (e.g., reforms, recoveries, economic improvement).

2: Predominantly positive — strong, explicit positive judgment.

Consider the main event and how the author conveys it. Respond only with one numeric code (-2, -1, 0, 1, or 2).

The prompt for the few-shot model is:

Classify the following political titles using the sentiment codebook:

Sentiment Scale

-2 = Predominantly negative, explicit and strong negative judgment.

-1 = Slightly negative, weak or implicit negative cues; or neutral tone describing events widely perceived as negative (e.g., wars, crises, inflation).

0 = Neutral or balanced, factually presented without evaluative judgment, or includes both positive and negative elements but they balance each other.

1 = Slightly positive, weak or implicit positive cues, or neutral tone describing events widely perceived as positive (e.g., reforms, recoveries, economic improvement).

2 = Predominantly positive, strong, explicit positive judgment.

Examples

-2: Vechten tegen kerncentrales heeft niets uitgehaald, toch gaat Roy Pumfrey door - Omroep Zeeland

-1: Bouwplannen Maassluis-West baren bewoners zorgen, gemeente gaat gesprek aan | WOS

0: Kamer wil opheldering over banden Nederlandse politici en Rusland

0: Ambtenaren kunnen hoger pensioen tegemoetzien, zorgpersoneel | Economie | NU.nl

1: Spaargeldgarantiefonds voor als bank faalt, is helemaal gevuld - NRC

2: China troeft VS af met lancering geheimzinnig ruimtevliegtuig

Consider the main event and how the author conveys it. Respond only with one numeric code (-2, -1, 0, 1, or 2).

Fine-tuned Model Performance

The final report on fine-tuning shows that there was no overfitting during the process for either approach, because as training loss decreases, validation loss decreases as well. Two indices directly reported on the fine-tuning dashboard both show that zero-shot approach (full validation accuracy: 0.905; full validation loss: 0.233) performs equally well as the few-shot approach (full validation accuracy: 0.912; full validation loss: 0.229).

Following the performance measures proposed by Alizadeh et al. (2025), we ran the model over the same validation set twice and checked model performance via two other metrics. One is average accuracy, which is the average of the accuracy of two trials of the same model. Accuracy represents the percentage of correct predictions among all predictions using the evaluation set as the gold standard. The other is the internal consistency of the model, which is measured with Krippendorff's α for inter-coder reliability. In our case, each trial of the same model was treated as an independent coding practice. The results show that fine-tuned zero-shot GPT-4.1 (average accuracy: 70.9%, Krippendorff's $\alpha = 0.97$) slightly outperformed the few-shot GPT-4.1 (average accuracy: 70.0%, Krippendorff's $\alpha = 0.95$) on both metrics.

For robustness, we also checked the precision, recall, and F1 score. It turns out that overall, fine-tuned zero-shot GPT4.1 (trial 1: precision: 0.76, recall: 0.71, F1 score: 0.73; trial 2: precision: 0.75, recall: 0.70, F1 score: 0.72) performs equally well as the fine-tuned few-shot GPT 4.1 (trial 1: precision: 0.77, recall: 0.71, F1 score: 0.73; trial 2: precision: 0.76, recall: 0.70, F1 score: 0.72). This aligns with previous research which find no consistent improvement in the performance of few-shot learning compared to zero-shot learning on the same task (Alizadeh et al., 2025).

Given the results above and for parsimony, we decided to use fine-tuned zero-shot GPT4.1 to compute sentiment scores for the whole dataset.

References

Alizadeh, M., Kubli, M., Samei, Z., Dehghani, S., Zahedivafa, M., Bermeo, J. D., Korobeynikova, M., & Gilardi, F. (2024). Open-source LLMs for text annotation: A practical guide for model setting and fine-tuning. *Journal of Computational Social Science*, *8*(1), 17. https://doi.org/10.1007/s42001-024-00345-9

M.S., N., & S.S., V. C. (2025). A review of unlabeled and imbalanced data challenges in machine learning: Strategies and solutions. *WIREs Data Mining and Knowledge Discovery*, *15*(3), e70043. https://doi.org/10.1002/widm.70043

Supplemental Material D

Exploratory Analyses on Users' Selection of Negative Political Content

An information diet dominated by more extremely negative content differs meaningfully from one with the same proportion of negative content but at lower levels of negativity. Therefore, as exploratory analyses for users' selection of negative political content, we also checked the sentiment level of each individual's political information diet.

The sentiment level of each individual's political information diet (M = - 0.42, SD = 0.49) was the average of the sentiment of all political titles they selected while using Discover. Besides, the sentiment of an individual's political information diet was also calculated weekly by averaging the sentiment scores of all political titles selected within each week.

We first examined sentiment-related differences in individuals' overall political information diet. With three socio-demographics and political interest controlled, an OLS regression shows no significant difference in sentiment of overall political information diet among people with different levels of internal efficacy (b = -0.01, β = -0.02, p = .771) and external efficacy (b = 0.04, β = 0.05, p = .251).

We then turn to the longitudinal perspective (see the table below). With the same controls included, we first fit a linear mixed-effects model with the sentiment level of either weekly political information diet as the dependent variable, interaction terms between time and two political efficacy variables as predictors, and both the intercept and slopes were allowed to vary by individual. The model showed that people with different internal efficacy (b = -0.03, β = -0.03, p = .543) and external efficacy levels (b = 0.03, β = 0.06, p = .087) did not differ significantly in how they change over weeks in terms of their political information diet's sentiment.

Table 1

Results of the mixed-effects models with the sentiment level of the weekly political information diet as the dependent variable

	DV = Sentiment level of political clicks			
Predictors	**b**	**β**	**SE**	**p**
Intercept	-0.413	-0.407	0.115	0.014
Internal efficacy	-0.032	-0.035	0.036	0.543

External efficacy	0.034	0.042	0.026	0.087
Time (relative week)	-0.003	-0.037	0.039	0.240
Interaction terms				
Internal efficacy * Time	-0.001	-0.024	0.043	0.497
External efficacy * Time	0.000	0.032	0.037	0.685
Controls				
Political interest	0.016	0.020	0.034	0.466
Biological sex	0.049	0.027	0.050	0.523
Age	-0.003	-0.025	0.023	0.549
Education	-0.028	-0.021	0.018	0.408
Model performance				
Marginal R^2	0.017			
Conditional R^2	0.185			

Note: *Unstandardized coefficients (b) are based on a model with mean-centered interaction terms. The other estimates, including standardized coefficients (β), are all based on a model with standardized predictors and an unstandardized dependent variable. Marginal R^2 is the variance explained by fixed factors. Conditional R^2 is the variance explained by both fixed and random factors.*