Iterative Topic Taxonomy Induction with LLMs: A Case Study of Electoral Advertising

Alexander Brady*
Department of Computer Science
ETH Zurich
Zurich, Switzerland
bradya@ethz.ch

Tunazzina Islam*
Department of Computer Science
Purdue University
West Lafayette, IN 47907, USA
islam32@purdue.edu

Abstract

Social media platforms play a pivotal role in shaping political discourse, but the scale and rapid evolution of online content make systematic analysis difficult. We introduce an end-to-end framework for inducing an interpretable topic taxonomy from unlabeled text corpora. The framework combines embedding-based clustering with iterative large language model (LLM) inference to construct a topic taxonomy without requiring predefined labels or seed topics. It first synthesizes candidate topics from document clusters and then uses the resulting taxonomy to assign consistent topic labels across clusters. We evaluate the approach through a case study of political advertising ahead of the 2024 U.S. presidential election. We use the induced taxonomy to support downstream analyses of issue prevalence, moral framing, advertising spend, and demographic exposure patterns. These results suggest that iterative taxonomy construction can provide a scalable and interpretable approach to organizing large unlabeled text corpora while supporting substantive downstream analysis.

1 Introduction

Modern political campaigns have been fundamentally transformed by the rise of social media platforms. Platforms such as Facebook and Instagram have become important venues for political advertising, enabling campaigns and advocacy organizations to reach voters with issue-specific appeals. Prior work has shown that voters respond more favorably to advertisements addressing issues they personally prioritize, irrespective of partisanship (Chu et al., 2024). Understanding the content and structure of these messages is critical for studying contemporary political communication.

The scale and rapid evolution of social media content make manual analysis difficult. Traditional topic modeling methods such as Latent Dirichlet Allocation (LDA) (Blei et al., 2003) and non-negative matrix factorization (NMF) (Lee and Seung, 1999) provide unsupervised mechanisms for discovering latent structure in text corpora, but their outputs often require additional interpretation or manual labeling. Prior work has therefore explored automated topic labeling through probabilistic methods (Magatti et al., 2009), summarization-based approaches (Wan and Wang, 2016; Cano Basave et al., 2014), and sequence-to-sequence models (Alokaili et al., 2020).

More recently, embedding-based approaches such as BERTopic (Grootendorst, 2022) have combined contextualized text representations with density-based clustering to produce more coherent topic groupings. Large language models (LLMs) have also enabled more semantically expressive approaches to topic discovery and labeling (Islam,

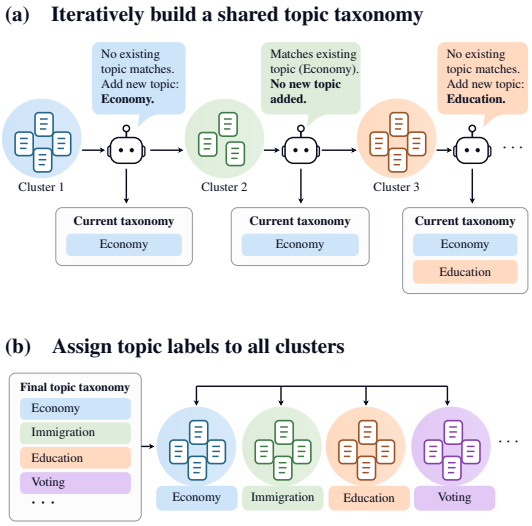


Figure 1: Overview of the proposed framework. Documents are first embedded and clustered. (a) Cluster representatives are processed sequentially by the LLM to induce a shared topic taxonomy. (b) The LLM assigns one topic label to each cluster from the completed taxonomy.

*Equal contribution.

Findings of the Association for Computational Linguistics: IJCNLP-AACL 2026, pages 1976–1990
November 6-10, 2026 ©2026 Association for Computational Linguistics

2026a; Sudhoff et al., 2026; Pham et al., 2024). However, several recent LLM-based approaches assume either an initial seed set of themes or a predefined set of topics (Islam and Goldwasser, 2025b,d).

In this work, we introduce a seed-free framework for inducing an interpretable topic taxonomy from unlabeled text corpora (Figure 1). Our method first uses embedding-based clustering to identify groups of semantically related documents. It then processes these clusters sequentially with an LLM, which determines whether each cluster is adequately represented by an existing topic or whether a new topic should be added to the taxonomy. In a second pass, the complete induced taxonomy is used to assign a single topic label to each cluster. This iterative procedure is designed to promote consistency across cluster labels while allowing the taxonomy to expand as new themes are encountered.

Unlike methods that require human-provided seed labels, our framework constructs the taxonomy directly from the corpus. The method is also distinct from single-shot LLM labeling: rather than generating each cluster label independently, previously induced topics remain available throughout the synthesis process, creating a shared and evolving label space across clusters. The resulting taxonomy can then be used for downstream annotation and analysis.

We evaluate the framework through a case study of political advertisements collected from the Meta Ad Library during the month preceding the 2024 U.S. presidential election[1]. The induced taxonomy captures a range of political issues represented in the advertising corpus. We further extend our analysis by examining the moral framings underlying these ads through the lens of Moral Foundations Theory (Iyer et al., 2012; Haidt and Graham, 2007; Haidt and Joseph, 2004). By examining political issue framing and its associations with observed audience demographics, the case study illustrates the substantive analyses enabled by the proposed framework.

Our contributions are:

1. A general, seed-free, LLM-guided framework for inducing interpretable topic taxonomies from unlabeled corpora without human intervention.
2. An iterative topic synthesis mechanism that yields more consistent and better-aligned topic labels than classical topic models and single-shot LLM labeling under human evaluation.
3. A case study of 2024 U.S. political advertising demonstrating how the induced taxonomy can support downstream analyses of issue salience, moral framing, ad spending, and demographic exposure.
4. Release of a curated dataset of 2024 U.S. presidential election ads to support future research.

[1]Code, prompts and dataset are available at: https://github.com/alexander-brady/llm-topic-synthesis

2 Related Work

Political advertising and microtargeting. Social media has become an important venue for political communication, campaign messaging, and voter mobilization (Islam, 2025b; Aggarwal et al., 2023; Capozzi et al., 2021; Goldberg et al., 2021; Ribeiro et al., 2019; Silva et al., 2020; Fulgoni et al., 2016; Teresi and Michelson, 2015; Hersh, 2015). A substantial body of work has examined political microtargeting and the ways campaigns tailor messages to specific audiences (Islam and Goldwasser, 2025c; Islam, 2025a; Tappin et al., 2023; Zuiderveen Borgesius et al., 2018; Hersh, 2015).

Related studies have analyzed the strategic dissemination of campaign content across platforms and the relationship between targeted advertising and voter responses (Hackenburg and Margetts, 2024; Islam et al., 2023a; Hirsch et al., 2024). These settings motivate scalable methods for organizing and interpreting large collections of political messages.

Topic modeling and automated topic labeling. Traditional topic modeling methods have been widely used to analyze large text corpora in domains including social media (Chen et al., 2019; Ozer et al., 2016) and political discourse (Mathaisel and Comm, 2021; Lahoti et al., 2018). However, classical approaches can produce topics that are difficult to interpret and often require human inspection or post-hoc labeling (Chang et al., 2009; Mei et al., 2007). More recent embedding-based approaches such as BERTopic (Grootendorst, 2022) combine contextualized representations with density-based clustering to improve topic coherence and interpretability.

Large language models have enabled a further shift toward direct topic synthesis and labeling. TopicGPT (Pham et al., 2024), for example, uses

LLMs to construct and refine topic taxonomies and has demonstrated strong performance in terms of coherence and interpretability. More broadly, LLMs have also shown strong performance on annotation tasks traditionally performed by human coders (Gilardi et al., 2023). These results motivate the use of LLM inference not only for labeling existing clusters but also for constructing a shared topic inventory.

Recent LLMs-in-the-loop methods have explored fine-grained theme and argument discovery, for example using an initial seed set of themes to guide topic discovery or uncover associated arguments (Islam and Goldwasser, 2025d,b; Islam et al., 2023b; Islam and Goldwasser, 2022). Lam et al. (2024) introduced LLooM, which iteratively induces and scores higher-level concepts from text, with optional analyst-provided seeds. In contrast, our framework incrementally constructs a shared topic inventory over document clusters and then relabels all clusters using the completed taxonomy.

Moral framing in political discourse. Moral Foundations Theory provides a widely used framework for analyzing the moral dimensions of political language (Iyer et al., 2012; Haidt and Graham, 2007; Haidt and Joseph, 2004). Earlier computational approaches typically relied on manually annotated datasets to train classifiers for detecting moral appeals (Pacheco et al., 2022; Roy et al., 2021). More recently, LLM-based approaches have been used to generate moral labels and explanations directly from political text (Islam and Goldwasser, 2025a). We build on this line of work by using the induced topic taxonomy as the basis for a downstream analysis of moral framing in political advertisements. In our setting, moral annotations are used as an exploratory analytical layer rather than as independent validation of the induced topics.

3 Methodology

We propose a framework for inducing an interpretable topic taxonomy from an unlabeled text corpus. This framework consists of three main stages: (1) embedding-based clustering, (2) LLM-based taxonomy induction, (3) cluster-level topic assignment. We formalize the iterative topic synthesis and annotation procedure in Algorithm 1.

3.1 Embedding-Based Clustering

Given a corpus of documents, we first encode each document into a dense vector representation using a pretrained sentence embedding model such as Sentence-BERT (Reimers and Gurevych, 2019). We then apply UMAP (McInnes et al., 2020) to reduce the dimensionality of the embedding space before clustering.

We cluster the reduced embeddings using HDBSCAN (McInnes et al., 2017), a density-based clustering algorithm that can identify clusters of varying density while leaving low-confidence points unassigned. This property is useful for noisy text corpora in which not every document necessarily belongs to a well-defined topical cluster.

For each cluster C_i, we select the five documents with the highest HDBSCAN membership probabilities as its representative set $R(C_i)$. These representative documents are used as the textual input to the LLM during both taxonomy induction and subsequent topic assignment. Restricting LLM inference to a small representative set reduces the amount of text that must be processed while emphasizing documents that are most characteristic of each cluster.

Algorithm 1 Iterative Topic Taxonomy Induction

Require: Document clusters $\mathcal{C} = \{C_1, \ldots, C_n\}$, cluster representatives $R(C_i)$, LLM $\mathcal{M}$
Ensure: Topic taxonomy $\mathcal{T}$ and cluster-topic assignments

Initialize topic list $\mathcal{T} \leftarrow \emptyset$
for each cluster $C_i \in \mathcal{C}$ **do**
  if $\mathcal{T} = \emptyset$ **then**
    Prompt $\mathcal{M}$ to generate a topic t from $R(C_i)$
    $\mathcal{T} \leftarrow \mathcal{T} \cup \{t\}$
  else
    Prompt $\mathcal{M}$ with $R(C_i)$ and topics $\mathcal{T}$
    Constrain output to {`yes`, `no`}
    if `no` **then**
      Prompt $\mathcal{M}$ to generate a new topic t from $R(C_i)$
      $\mathcal{T} \leftarrow \mathcal{T} \cup \{t\}$
    end if
  end if
end for
for each cluster $C_i \in \mathcal{C}$ **do**
  Prompt $\mathcal{M}$ with $R(C_i)$ and topics $\mathcal{T}$
  Constrain output to select one topic $t \in \mathcal{T}$
  Assign topic t to cluster C_i
end for
return $\mathcal{T}$ and cluster-topic assignments

3.2 Iterative Topic Taxonomy Induction

The goal of the next stage is to construct a shared topic inventory directly from the document clusters, without requiring human-provided seed labels. We initialize the topic taxonomy $\mathcal{T}$ as an empty set and process the clusters sequentially.

For the first cluster, the LLM is prompted to generate a concise topic label from its representative documents. This label becomes the first element of $\mathcal{T}$. For each subsequent cluster, the LLM is shown the cluster representatives together with the topics currently in $\mathcal{T}$ and asked whether any existing topic adequately describes the cluster. We use constrained decoding (Beurer-Kellner et al., 2024) to restrict this decision to `yes` or `no`. If the model returns `no`, it is then prompted to generate a new concise topic label, which is then added to $\mathcal{T}$.

In this way, previously generated labels remain available while later clusters are processed. Unlike independent, single-shot labeling, the procedure therefore constructs an evolving shared label space and only introduces a new topic when the model determines that the existing inventory is insufficient.

For our experiments, clusters were processed in ascending HDBSCAN label order. Because HDBSCAN cluster identifiers are arbitrary rather than semantically ordered, this corresponds to a fixed but otherwise arbitrary traversal of the induced clusters. We treat robustness to cluster ordering as a limitation and direction for future work.

3.3 Cluster-Level Topic Assignment

The first pass determines which topics belong in the taxonomy but does not necessarily provide the final assignment for every cluster. After the complete topic inventory has been constructed, we therefore perform a second pass over all clusters.

For each cluster, the LLM receives its representative documents together with the complete taxonomy $\mathcal{T}$. We use constrained decoding to require the model to select exactly one label from the available topic inventory. This second pass allows every cluster to be evaluated against the same final label space, including topics that may have been introduced only after that cluster was first encountered. The output of the framework is therefore (1) an induced topic taxonomy and (2) a topic assignment for each document cluster.

4 Case Study: Experimental Setup

We apply the proposed framework to political advertisements from the 2024 U.S. election cycle. The goal of the case study is twofold: first, to evaluate the quality of the induced topic taxonomy; and second, to demonstrate how the resulting annotations can support downstream analyses.

4.1 Dataset

We collected political advertisements running on Facebook and Instagram in the United States during October 2024 using the Meta Ad Library, starting one month before the 2024 presidential election. The resulting dataset contains $8{,}047$ unique advertisements. The ads cover election-related messaging at both the local and federal levels. 99% of the advertisements are in English, with a small fraction in Spanish.

For each advertisement, the Meta Ad Library provides a unique ad identifier, title, body text, URL, creation time, campaign duration, page author, funding entity, and advertising spend reported as a range. We represent spend using the midpoint of the reported range. Meta also reports impressions as a range, for which we similarly use the midpoint, together with aggregate audience distributions by gender, age group, and U.S. state.

Because these demographic statistics describe observed ad delivery rather than necessarily the advertiser's intended targeting criteria, we refer to them throughout as *audience exposure* or *delivery patterns* rather than direct targeting decisions.

4.2 Taxonomy Induction

We embed the advertisement body texts using the `all-MiniLM-L6-v2` sentence-transformer model. To reduce near-duplicate content, we remove advertisements whose embedding has cosine similarity greater than 0.95 to another advertisement. We then apply HDBSCAN to the UMAP-reduced embeddings with a minimum cluster size of 15.

This procedure assigns $4{,}510$ advertisements to 73 clusters. For each cluster, we select the five advertisements with the highest HDBSCAN membership probabilities as cluster representatives. These are given to the LLM during taxonomy induction and cluster-level topic assignment.

We use a locally hosted Llama-3.2-3B model (Grattafiori et al., 2024) to execute our framework, without GPU acceleration.

We initialize the topic inventory with no seed

labels. Processing the 73 clusters using the procedure described in Section 3 produces a taxonomy containing 15 generated topics. During the second-pass cluster assignment stage, the model is additionally allowed to assign the label *other* when none of the induced topics adequately describes a cluster. Although *border security* appears in the generated taxonomy, no cluster is ultimately assigned this label. The final topic set and cluster counts are shown in Table 1. The complete prompts are provided in App. A.

4.3 Human Evaluation: Topic Labels

To evaluate the quality of the induced labels, we compare our method with two baselines. The first is BERTopic (Grootendorst, 2022). We use the same underlying clusters for both approaches and provide the same LLM with the cluster representative documents; for the BERTopic condition, the model additionally receives the ten highest-ranked cluster keywords.

The second baseline is inspired by TopicGPT (Pham et al., 2024). For each cluster, the LLM receives the representative advertisements and generates a concise topic label independently, without access to previously generated labels or an evolving shared taxonomy. This comparison is intended to isolate the effect of iterative taxonomy construction relative to single-shot LLM labeling.

Two volunteer annotators (one female and one male) with backgrounds in natural language processing (NLP) and computational social science (CSS), age range 25-40, evaluate the resulting labels. Each annotator is given the same set of 10 advertisements. For each evaluated advertisement, the labels produced by BERTopic, the TopicGPT-style baseline, and our framework are presented in randomized order. Annotators select the label that best represents the advertisement and rate each candidate label from 1 (irrelevant) to 5 (perfect fit).

4.4 Moral Foundation Annotation

We use Moral Foundations Theory (Iyer et al., 2012; Haidt and Graham, 2007; Haidt and Joseph, 2004) as an exploratory framework for examining the moral framing of political advertisements. We consider six foundations: Care/Harm, Fairness/Cheating, Loyalty/Betrayal, Authority/Subversion, Sanctity/Degradation, and Liberty/Oppression. For more information on Moral Foundations Theory, see App. B.

For each topic-labeled cluster, we first prompt the LLM to summarize the primary argument expressed by its representative advertisements with respect to the assigned topic. We then provide the model with definitions of the six moral foundations and use constrained decoding to require selection of a single foundation. The complete prompts are provided in App. A.

Because these moral labels are generated using the topic annotation as part of the prompting procedure, we treat subsequent relationships between topics and moral foundations as exploratory downstream analyses rather than independent validation of the topic taxonomy.

4.5 Document-Level Classification

The taxonomy-induction framework produces cluster-level labels, while HDBSCAN leaves a substantial fraction of advertisements unassigned. To examine whether the induced topic labels can be propagated to individual advertisements, we conduct a separate supervised classification experiment.

We construct a pseudo-labeled training set from advertisements assigned to clusters with HDBSCAN membership probability greater than 0.98. This produces 2,337 advertisements. The resulting class distribution is highly imbalanced: the four largest classes—*voting rights* (26%), *crime/justice* (19%), *education* (17%), and *abortion* (11%)—account for 73% of the pseudo-labeled advertisements, while the smallest classes, *healthcare access* and *property taxes*, each account for 0.64%.

We additionally sample 200 advertisements from among the unassigned documents and annotate them manually with topic labels. These annotations serve as human-labeled evaluation data for the document-level classification task.

We compare Logistic Regression, XGBoost (Chen and Guestrin, 2016), RoBERTa (Liu et al., 2019), and SetFit (Tunstall et al., 2022). For classifiers operating on fixed embeddings, advertisement texts are represented using `all-MiniLM-L6-v2`. Hyperparameters are selected using GridSearchCV, and performance is reported using macro F1 and accuracy. The manually annotated advertisements were held out entirely for final evaluation.

4.6 Ad-Level Annotation

For downstream descriptive analyses, we additionally assign topic and moral-foundation labels to individual advertisements using the induced tax-

Issue	#Clusters	Example Ad
economy	15	*Molly Buck's agenda is failing Iowa families. We can't afford Molly Buck in the State House.*
voting rights	13	*Make Your Vote Count. RE-register to Vote in the District of Your Second Home.*
crime/justice	12	*Orange County Firefighters Trust Dave Min To Keep Our Communities Safe.*
personal freedom	4	*Tim Sheehy will always fight for Montana Values in the Senate!*
voting	4	*Vote Rebecca for State Rep by Nov 5th!*
education	4	*Vote for students. Vote for teachers. Support our public schools.*
affordable housing	4	*Catherine has supported more multi-family housing so we can afford homes in the communities we grew up in. Vote Stefani!*
environmental protection	4	*We can count on Lucy Rehm to be a responsible leader and protect our clean air and water.*
immigration	3	*Yadira Caraveo supported efforts to defund ICE and the border patrol. Now Colorado is being overwhelmed with illegal immigration.*
election integrity	3	*Question 3 will change our state's Constitution and weaken our democracy by throwing away thousands of ballots over minor errors. Vote NO on Question 3.*
abortion	2	*Defend reproductive freedom - VOTE on Tuesday, November 5th!*
healthcare access	1	*Lucy Rehm cares about improving conditions in your local hospital. That's why Minnesota Nurses endorsed Lucy Rehm!*
property taxes	1	*Caleb will work to stop reckless spending and oppose tax and fee increases.*
sewer sale	1	*Get the facts on the Gloucester Township sewer sale, visit VoteYesGloucesterTownship.com!*
other	2	*LISCOMB - thanks for welcoming me at the little street where old and new friends meet!*

Table 1: Topics identified by the LLM in the political ads dataset. '#Clusters' column gives the number of unique clusters assigned to each topic by the LLM.

Model	Average Score	Best Label
BERTopic	1.2	3
TopicGPT-style	2.7	5
Our Method	2.8	12

Table 2: Annotation results. The average score is out of 5, and the best label is the number of times that model's label was selected as the most fitting (among all annotators).

Model	Macro F1	Accuracy
Logistic Regression	0.37	0.55
XGBoost	0.31	0.53
RoBERTa	0.32	0.53
SetFit	0.36	0.60

Table 3: Document-level classification performance for topic assignment.

onomy. We use `mistralai/Mistral-7B-v0.1` (Jiang et al., 2023) for this ad-level annotation stage. This model is used only for the downstream case-study analyses and is separate from the Llama-3.2-3B model used to induce and evaluate the cluster-level taxonomy.

Using these document-level annotations together with Meta's spend and impression data, we examine issue prevalence, advertising expenditure, funding concentration, moral framing, and demographic exposure patterns.

To characterize associations between advertisement content and demographic groups, we compute positive pointwise mutual information (PPMI) (Church and Hanks, 1990) between topic or moral-foundation annotations and observed audience attributes. Higher PPMI indicates that a topic or moral frame occurs disproportionately among advertisements delivered to a given demographic group. We interpret these associations descriptively as exposure patterns rather than direct evidence of advertisers' targeting intent.

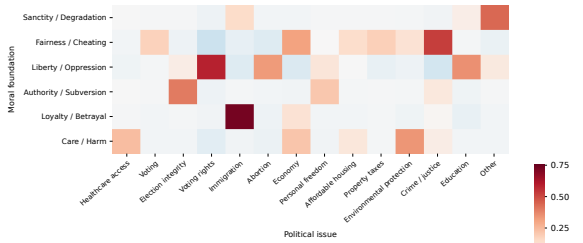
5 Results and Analysis

We first evaluate the quality of the induced topic labels and the feasibility of propagating those labels to individual advertisements. We use the resulting annotations for exploratory analyses of issue salience, moral framing, advertising expenditure, funding concentration, and demographic exposure.

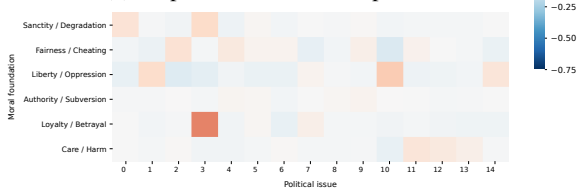
5.1 Topic Label Evaluation

Table 2 summarizes the human evaluation of topic labels generated by BERTopic, the TopicGPT-style single-shot baseline, and our iterative framework.

The iterative framework receives the highest average rating (2.8), compared with 2.7 for the single-shot baseline and 1.2 for BERTopic. The difference in mean rating between our method and the single-shot baseline is small. However, the labels produced by our method are selected as the best-fitting



(a) Proposed Framework Topics



(b) LDA Topics

Figure 2: Associations between moral-foundation labels and topic labels for (a) the proposed taxonomy and (b) LDA topics. Cells show Pearson correlations between one-hot topic and moral-foundation assignments.

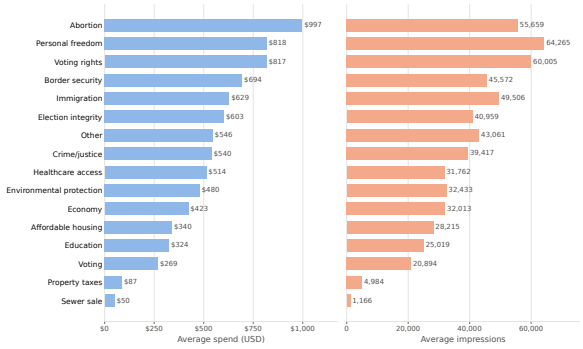


Figure 3: Average advertising spend and impressions per ad by annotated topic.

label in 12 judgments, compared with 5 for the single-shot baseline and 3 for BERTopic.

These results suggest that maintaining a shared, evolving topic inventory can improve label consistency relative to generating cluster labels independently, although the small scale of the human evaluation limits the strength of this conclusion. The strongest contrast is with BERTopic-based labeling, while the difference between iterative and single-shot LLM labeling is more modest in average rating.

The annotators achieve a Cohen's κ (Cohen, 1960) of 0.66, indicating substantial agreement under this evaluation setup. For 5 out of the 20 rounds, the annotators gave the same score for all models (in all cases, a score of 1).

5.2 Document-Level Classification Results

We next evaluate whether the cluster-derived labels can be propagated to individual advertisements. Table 3 reports performance across the four supervised classifiers. Logistic Regression achieves the highest macro F1 score (0.37), while SetFit achieves the highest overall accuracy (0.60). Performance is modest across all models, reflecting the difficulty of the task and the substantial class imbalance in the pseudo-labeled training data.

These results indicate that the induced labels can provide supervision for document-level classification, but they also show that propagating cluster-level topic labels to individual advertisements remains challenging. We therefore treat the supervised classifier as a downstream extension of the framework rather than evidence of high-quality annotation for all advertisements.

5.3 Topic Salience and Moral Framing

We next examine how the induced topics relate to the moral-foundation annotations. For each advertisement, we construct one-hot representations of its topic and moral-foundation labels and compute pairwise Pearson correlations.

Figure 2 compares these relationships for the induced taxonomy and an LDA baseline. In the induced taxonomy, several topic–foundation pairs show comparatively strong associations. For example, *Fairness/Cheating* is strongly associated with *crime/justice*, while *Loyalty/Betrayal* is strongly associated with *immigration*. The corresponding LDA topics show weaker associations with the moral-foundation labels.

Because the moral labels are themselves generated using topic-conditioned LLM inference, these relationships should be interpreted as descriptive properties of the resulting annotation pipeline rather than independent validation of the topic taxonomy. Nevertheless, they illustrate how an interpretable topic inventory can support subsequent analysis of moral framing.

5.4 Ad-Level Analysis

We further use the document-level annotations to examine advertising expenditure, audience exposure, funding concentration, and moral framing. These analyses are descriptive and are intended to demonstrate the analytical affordances of the induced taxonomy rather than establish causal claims about campaign strategy.

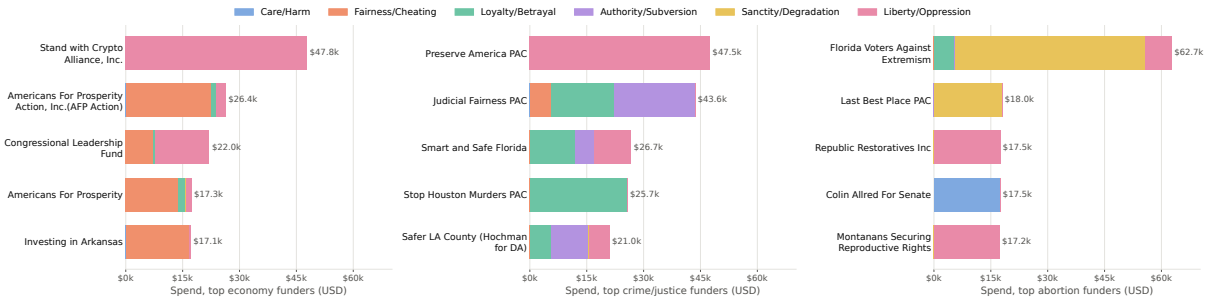


Figure 4: Spend by the top five funders of economy, crime/justice, and abortion ads, segmented by moral foundation.

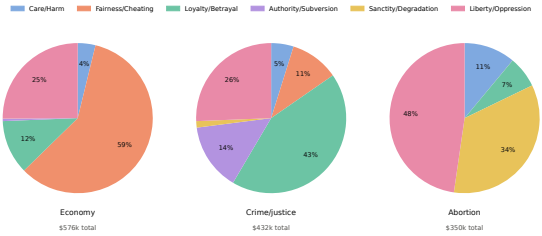


Figure 5: Distribution of advertising spending across moral foundations within economy, crime/justice, and abortion advertisements. Percentages indicate within-topic spending shares; total spending is shown beneath each chart.

5.4.1 Spending and Impressions by Topic

Figure 3 compares average advertising spend and average impressions across topics. *Abortion*, *personal freedom*, and *voting rights* have among the highest average spending levels in the annotated corpus. By contrast, *immigration* and *border security* receive comparatively high impressions relative to their average spending.

Lower-volume issues such as *property taxes* and local-service topics receive substantially less expenditure and exposure. Overall, the differences between spending and impressions illustrate that issue categories vary not only in advertiser investment but also in the amount of observed audience exposure associated with that investment.

5.4.2 Leading Funders by Issue

Figure 4 compares estimated advertising expenditure among the five highest-spending funders within each of three issue categories: *economy*, *abortion*, and *crime/justice*[2] . For economy and abortion advertisements, the largest displayed funder spends substantially more than each of the other four. In contrast, crime/justice has two leading funders with similar expenditures, followed by three funders with lower, broadly comparable spending.

[2]https://www.pewresearch.org/politics/2024/09/09/issues-and-the-2024-election/

These results illustrate differences in expenditure patterns among the leading funders across issue categories in the observed corpus. Because the comparison is restricted to the five highest-spending funders in each category, it does not establish differences in overall funding concentration or in the total number of organizations supporting each issue. Assessing those questions requires the complete funder-level expenditure distribution. Further analysis is provided in Appendix D.1.

5.4.3 Moral Framing and Advertising Spend

Figure 5 shows the distribution of estimated advertising expenditure across moral foundations within each issue category. Economic advertising primarily emphasizes *Fairness/Cheating* (59%) and *Liberty/Oppression* (25%), while crime/justice advertising primarily emphasizes *Loyalty/Betrayal* (43%) and *Liberty/Oppression* (26%). Abortion advertising primarily emphasizes *Liberty/Oppression* (48%) and *Sanctity/Degradation* (34%).

Figure 4 shows that the composition of expenditure also varies among the highest-spending funders within each category. For example, the displayed abortion funders include expenditure associated with *Care/Harm*, in addition to the two largest aggregate foundations. These findings describe differences in spending associated with LLM-generated moral annotations; they do not independently establish funder motives or moral polarization. Further analysis is provided in Appendix D.2.

5.4.4 Demographic Exposure Patterns

Finally, we use PPMI to identify topic and moral-foundation labels that occur disproportionately among advertisements delivered to particular demographic groups.

Figure 6 shows several descriptive differences in exposure. In Florida, younger audiences receive a disproportionate share of *affordable housing* ad-



(a) Females: AL vs WA

(b) All age groups: FL

(c) Males: MT vs VA

(d) Moral framing across selected states

Figure 6: PPMI associations between political issue or moral-foundation labels and observed audience demographics. Higher values indicate that a label occurs disproportionately among advertisements delivered to the corresponding demographic group.

vertisements, while older audiences show stronger associations with *abortion*-related advertising. Geographic variation is also visible among male audiences: advertisements delivered to males in Montana show stronger associations with *environmental protection* and *personal freedom*, whereas advertisements delivered to males in Virginia show stronger associations with *voting*.

These patterns demonstrate that issue exposure varies across demographic groups in the observed Meta delivery data. Because the dataset describes realized audience exposure rather than the full set of advertiser-selected targeting parameters, we **do not** interpret these associations as direct evidence of advertiser targeting intent.

6 Conclusion

We introduced a seed-free framework for inducing interpretable topic taxonomies from unlabeled text corpora. The framework combines embedding-based clustering with iterative LLM inference, maintaining a shared topic inventory that expands as new clusters are processed and is subsequently used to assign labels across the complete set of clusters.

We evaluated the framework through a case study of political advertisements from the month preceding the 2024 U.S. presidential election. In a small-scale human evaluation, the induced labels received similar average ratings to a single-shot LLM baseline but were selected as the best-fitting labels more often. We further showed that the resulting taxonomy can support downstream analyses of issue prevalence, moral framing, advertising expenditure, funding concentration, and demographic exposure.

The framework is not specific to political advertising and can be applied to other large, unlabeled text corpora in which interpretable topic organization is useful. Future work should evaluate the stability of induced taxonomies across domains, clustering configurations, LLMs, prompts, and cluster-processing orders, and explore methods for constructing more robust or hierarchical topic inventories.

7 Limitations

Our framework depends on the quality of the underlying embeddings and clustering procedure, and the resulting taxonomy may be sensitive to clustering parameters and the order in which clusters are processed. The human evaluation is also small in scale, and the difference in average rating between our framework and the single-shot LLM baseline

is modest. Larger evaluations across datasets and models are needed to assess the robustness of the approach.

The framework also relies on LLM-generated labels, which may introduce bias, oversimplify ambiguous texts, or produce misleading interpretations (Islam, 2026b; Ji et al., 2023; Blodgett et al., 2020; Brown et al., 2020). The moral-foundation labels are generated using topic-conditioned inference and should therefore be interpreted as exploratory rather than as independent validation of the induced topics.

Finally, the case study covers approximately one month of Meta political advertising during the 2024 U.S. election cycle. Its substantive findings may not generalize to other periods, platforms, or political contexts. Meta reports spending and impressions as ranges, and its demographic statistics describe observed ad delivery rather than necessarily reflecting advertiser targeting intent.

8 Ethical Considerations

This study uses political advertising data made publicly available through the Meta Ad Library. The data contain aggregate audience statistics and do not include personally identifiable information.

As the topic and moral annotations are partly generated by LLMs, they may reflect model biases or simplify nuanced political discourse. We therefore interpret the resulting patterns descriptively and avoid treating demographic exposure patterns as evidence of individual political preferences or advertiser intent. We report the models, prompts, and processing details used in our experiments to support transparency and reproducibility.

9 Acknowledgments

We would like to thank Lightning AI Studio for providing compute resources. We also thank the anonymous reviewers for their thoughtful feedback.

References

Minali Aggarwal, Jennifer Allen, Alexander Coppock, Dan Frankowski, Solomon Messing, Kelly Zhang, James Barnes, Andrew Beasley, Harry Hantman, and Sylvan Zheng. 2023. A 2 million-person, campaign-wide field experiment shows how digital advertising affects voter turnout. *Nature Human Behaviour*, 7(3):332–341.

Areej Alokaili, Nikolaos Aletras, and Mark Stevenson. 2020. Automatic generation of topic labels. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, pages 1965–1968.

Luca Beurer-Kellner, Marc Fischer, and Martin Vechev. 2024. Guiding llms the right way: fast, non-invasive constrained generation. In *Proceedings of the 41st International Conference on Machine Learning*, pages 3658–3673.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Su Lin Blodgett and 1 others. 2020. Language (technology) is power: A critical survey of "bias" in nlp. In *ACL*.

Tom Brown and 1 others. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Amparo Elizabeth Cano Basave, Yulan He, and Ruifeng Xu. 2014. Automatic labelling of topic models learned from twitter by summarisation. Association for Computational Linguistics (ACL).

Arthur Capozzi, Gianmarco de Francisci Morales, Yelena Mejova, Corrado Monti, André Panisson, and Daniela Paolotti. 2021. Clandestino or rifugiato? anti-immigration facebook ad targeting in italy. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15.

Jonathan Chang, Sean Gerrish, Chong Wang, Jordan Boyd-Graber, and David Blei. 2009. Reading tea leaves: How humans interpret topic models. *Advances in neural information processing systems*.

Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Yong Chen, Hui Zhang, Rui Liu, Zhiwen Ye, and Jianying Lin. 2019. Experimental explorations on short text topic mining between lda and nmf based schemes. *Knowledge-Based Systems*, 163:1–13.

Xiaotong Chu, Lukas Otto, Rens Vliegenthart, Sophie Lecheler, Claes de Vreese, and Sanne Kruikemeier. 2024. On or off topic? understanding the effects of issue-related political targeted ads. *Information, Communication & Society*, 27(7):1378–1404.

Kenneth Church and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational linguistics*, 16(1):22–29.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

Gian M Fulgoni, Andrew Lipsman, and Carol Davidsen. 2016. The power of political advertising: Lessons for practitioners: How data analytics, social media, and creative strategies shape us presidential election campaigns. *Journal of Advertising Research*, 56(3):239–244.

Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. 2023. Chatgpt outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30):e2305016120.

Matthew H Goldberg, Abel Gustafson, Seth A Rosenthal, and Anthony Leiserowitz. 2021. Shifting republican views on climate change through targeted advertising. *Nature Climate Change*, 11(7):573–577.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, and 1 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Kobi Hackenburg and Helen Margetts. 2024. Evaluating the persuasive influence of political microtargeting with large language models. *Proceedings of the National Academy of Sciences*, 121(24):e2403116121.

Jonathan Haidt and Jesse Graham. 2007. When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20(1):98–116.

Jonathan Haidt and Craig Joseph. 2004. Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, 133(4):55–66.

Eitan D Hersh. 2015. *Hacking the electorate: How campaigns perceive voters*. Cambridge University Press.

Melanie Hirsch, Marlis Stubenvoll, Alice Binder, and Jörg Matthes. 2024. Beneficial or harmful? how (mis) fit of targeted political advertising on social media shapes voter perceptions. *Journal of Advertising*, 53(1):19–35.

Tunazzina Islam. 2025a. *Understanding and Analyzing Microtargeting Pattern on Social Media*. Ph.D. thesis, Purdue University.

Tunazzina Islam. 2025b. Understanding microtargeting pattern on social media. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 29269–29270.

Tunazzina Islam. 2026a. Reasoning-based refinement of unsupervised text clusters with LLMs. In *Findings of the Association for Computational Linguistics: ACL 2026*, pages 9899–9917, San Diego, California, United States. Association for Computational Linguistics.

Tunazzina Islam. 2026b. Who gets which message? auditing demographic bias in LLM-generated targeted text. In *Findings of the Association for Computational Linguistics: ACL 2026*, pages 9270–9297, San Diego, California, United States. Association for Computational Linguistics.

Tunazzina Islam and Dan Goldwasser. 2022. Understanding covid-19 vaccine campaign on facebook using minimal supervision. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 585–595. IEEE.

Tunazzina Islam and Dan Goldwasser. 2025a. Can llms assist annotators in identifying morality frames? - case study on vaccination debate on social media. In *Proceedings of the 17th ACM Web Science Conference 2025*, Websci '25, page 169–178, New York, NY, USA. Association for Computing Machinery.

Tunazzina Islam and Dan Goldwasser. 2025b. Discovering latent themes in social media messaging: A machine-in-the-loop approach integrating llms. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 19, pages 859–884.

Tunazzina Islam and Dan Goldwasser. 2025c. Post-hoc study of climate microtargeting on social media ads with LLMs: Thematic insights and fairness evaluation. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 15838–15859, Suzhou, China. Association for Computational Linguistics.

Tunazzina Islam and Dan Goldwasser. 2025d. Uncovering latent arguments in social media messaging by employing LLMs-in-the-loop strategy. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 7412–7444, Albuquerque, New Mexico. Association for Computational Linguistics.

Tunazzina Islam, Shamik Roy, and Dan Goldwasser. 2023a. Weakly supervised learning for analyzing political campaigns on facebook. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, pages 411–422.

Tunazzina Islam, Ruqi Zhang, and Dan Goldwasser. 2023b. Analysis of climate campaigns on social media using bayesian model averaging. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pages 15–25.

Ravi Iyer, Spassena Koleva, Jesse Graham, Peter Ditto, and Jonathan Haidt. 2012. Understanding libertarian morality: The psychological dispositions of self-identified libertarians. *PLoS ONE*, 7(8).

Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *ACM computing surveys*, 55(12):1–38.

Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego

de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, and 1 others. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.

Preethi Lahoti, Kiran Garimella, and Aristides Gionis. 2018. Joint non-negative matrix factorization for learning ideological leaning on twitter. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*.

Michelle S Lam, Janice Teoh, James A Landay, Jeffrey Heer, and Michael S Bernstein. 2024. Concept induction: Analyzing unstructured text with high-level concepts using lloom. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–28.

Daniel D Lee and H Sebastian Seung. 1999. Learning the parts of objects by non-negative matrix factorization. *nature*, 401(6755):788–791.

Yinhan Liu and 1 others. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Davide Magatti, Silvia Calegari, Davide Ciucci, and Fabio Stella. 2009. Automatic labeling of topics. In *2009 Ninth International Conference on Intelligent Systems Design and Applications*, pages 1227–1232. IEEE.

Dennis FX Mathaisel and Clare L Comm. 2021. Political marketing with data analytics. *Journal of Marketing Analytics*, 9(1).

Leland McInnes, John Healy, and Steve Astels. 2017. hdbscan: Hierarchical density based clustering. *JOSS*.

Leland McInnes, John Healy, and James Melville. 2020. Umap: Uniform manifold approximation and projection for dimension reduction. *Preprint*, arXiv:1802.03426.

Qiaozhu Mei, Xuehua Shen, and ChengXiang Zhai. 2007. Automatic labeling of multinomial topic models. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 490–499.

Mert Ozer, Nyunsu Kim, and Hasan Davulcu. 2016. Community detection in political twitter networks using nonnegative matrix factorization methods. In *2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, pages 81–88. IEEE.

Maria Leonor Pacheco, Tunazzina Islam, Monal Mahajan, Andrey Shor, Ming Yin, Lyle Ungar, and Dan Goldwasser. 2022. A holistic framework for analyzing the covid-19 vaccine debate. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5821–5839.

Chau Minh Pham, Alexander Hoyle, Simeng Sun, Philip Resnik, and Mohit Iyyer. 2024. TopicGPT: A prompt-based topic modeling framework. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 2956–2984, Mexico City, Mexico. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Filipe N Ribeiro and 1 others. 2019. On microtargeting socially divisive ads: A case study of russia-linked ad campaigns on facebook. In *Proceedings of the conference on fairness, accountability, and transparency*.

Shamik Roy, Maria Leonor Pacheco, and Dan Goldwasser. 2021. Identifying morality frames in political tweets using relational learning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9939–9958.

Mirela Silva, Luiz Giovanini, Juliana Fernandes, Daniela Oliveira, and Catia S Silva. 2020. Facebook ad engagement in the russian active measures campaign of 2016. *arXiv preprint arXiv:2012.11690*.

Samantha Sudhoff, Pranav Perumal, Zhaoqing Wu, and Tunazzina Islam. 2026. Paid voices vs. public feeds: Interpretable cross-platform theme-based analysis of climate discourse. *arXiv preprint arXiv:2601.13317*.

Ben M Tappin and 1 others. 2023. Quantifying the potential persuasive returns to political microtargeting. *Proceedings of the National Academy of Sciences*.

Holly Teresi and Melissa R Michelson. 2015. Wired to mobilize: The effect of social networking messages on voter turnout. *The Social Science Journal*, 52(2):195–204.

Lewis Tunstall, Nils Reimers, Unso Eun Seo Jo, Luke Bates, Daniel Korat, Moshe Wasserblat, and Oren Pereg. 2022. Efficient few-shot learning without prompts. *arXiv preprint arXiv:2209.11055*.

Xiaojun Wan and Tianming Wang. 2016. Automatic labeling of topic models using text summaries. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2297–2305.

Frederik Zuiderveen Borgesius and 1 others. 2018. Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review*, 14(1):82–96.

A Prompts

Below are the prompts used in the various steps of the framework. Constrained decoding was implemented with the guidance framework[3].

System Prompt

[3]https://github.com/guidance-ai/guidance

You are a political analyst. You are given the following set of political ads to analyze. You will assign the ads a topic that best summarizes the key issue discussed in these ads.

A.1 Topic Synthesis

Binary Output

Is there a topic in the following list that well describes the key issue discussed in these ads?

Topics:

{{ topics }}

Can the ads be summarized by one of the previous topics? Answer with "yes" or "no".

Generate New Topic (User)

In three words or less, describe the issue that ads discuss, summarizing the topic. Examples: "Abortion", "Climate Change", etc.

Generate New Topic (Assistant)

A better topic for these ads is: "

The LLM would finish the above, with the quote (") as a stop token.

A.2 Annotation

System Prompt

You are a political analyst. You are given the following set of political ads to analyze: {{ ads }}

Select Topic (User)

Which of the following topics best describes the key issue discussed in these ads? Select exactly one topic from the provided list.

Topics:

{{ topics }}

Summarize (User)

Summarize the main talking point of the ads. Do so in a single sentence.

Select Moral Foundation (User)

Which of the following moral foundations best describes the arguments used in the ads about {{ topic }}?

{{ MFs_with_definitions }}

B Moral Foundations Theory

Moral Foundations Theory (MFT) (Iyer et al., 2012; Haidt and Graham, 2007; Haidt and Joseph, 2004) suggests a theoretical framework for analyzing *six* moral values (i.e., foundations, each with a positive and a negative polarity) central to human moral sentiment. MFT states that political attitudes are shaped by *six* core moral dimensions: care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, sanctity/degradation, and liberty/oppression.

The definitions given in Table 4 correspond to the same definitions given to the LLM when assigning moral foundations.

C Substantive Insights: Topic Salience and Moral Framing

We analyze the distribution of moral framings and political issues across different demographics to uncover political microtargeting strategies. We used the positive pointwise mutual information (PPMI) to detect topics that appear disproportionately in ads shown to specific demographics, potentially indicating targeted messaging. We visualize the results of the PPMI analysis in Figure 6. The plots show the distribution of moral framings and political issues across different demographics. Some interesting observations include the differences in ads targeting different states, or the shift in topics among different age groups in Florida (Figure 6b). For example, *young* people in Florida are more likely to be targeted with ads related to *affordable housing*, while *older* people are more likely to be targeted with ads related to *abortion*. This shows generational targeting strategies, where `economic concerns` are emphasized for younger groups, while older groups are mobilized on `moralized issues` like abortion.

On the other hand, Figure 6c shows that *environmental protection* and *personal freedom* are heavily overrepresented among *male* audiences in Montana, while *abortion* also appears moderately elevated. By contrast, *male* audiences in Virginia are disproportionately exposed to *voting*-related advertising. These contrasts highlight

CARE/HARM: It suggests that someone other than the speaker is worthy of compassion or is experiencing harm, grounded in the values of kindness, tenderness, and care.
FAIRNESS/CHEATING: Emphasizes justice, personal rights, and independence; involves comparing with other groups. Advocates for equal opportunity and resists those who benefit without contributing ("Free Riders").
LOYALTY/BETRAYAL: Based on the values of loyalty to one's country and willingness to sacrifice for the group. Activated by a sense of unity and collective responsibility—"one for all, and all for one".
AUTHORITY/SUBVERSION: Centers on showing respect (or resistance) toward established authority and following long-standing traditions. It includes maintaining social order and fulfilling the duties tied to hierarchical roles, such as obedience, respect, and role-based responsibilities.
SANCTITY/DEGRADATION: Beyond religion, this value highlights respect for human dignity and aversion to moral or physical corruption, promoting purity, self-control, and the belief that the body is sacred and vulnerable to defilement.
LIBERTY/OPPRESSION: Captures the feelings of reactance and resentment people experience when their freedom is restricted, often leading to collective disdain for authoritarian figures and motivating unity and resistance against oppression.

Table 4: Six basic moral foundations (Iyer et al., 2012; Haidt and Graham, 2007; Haidt and Joseph, 2004).

how campaign strategies selectively target localized demographic niches: `progressive-leaning issue frames resonate more strongly in Montana male` audiences, whereas mobilization around `law-and-order` and `electoral process is more pronounced among Virginia males`.

D Substantive Insights: Ad-Level Analysis

Using the synthesized topics, we annotate individual ads to enable ad-level analysis. Cluster-level annotations reveal latent issue and moral structure, while ad-level annotations enable analysis of spending, funder concentration, and targeting strategies.

D.1 Top Funders

We examine the concentration of advertising spend among funding entities across major issue domains. Figure 4 presents the top 5 funding entities by total spend for three central issues in the 2024 election—*economy*, *abortion*, and *crime/justice*.

For the *economy*, Figure 4 shows substantial variation in advertising expenditure among the five highest-spending funders in the observed corpus. Stand with Crypto Alliance, Inc. is the largest displayed spender, at approximately $47,800, followed by Americans for Prosperity Action, Inc. ($26,400) and the Congressional Leadership Fund ($22,000). Americans for Prosperity and Investing in Arkansas spend approximately $17,300 and $17,100, respectively. These results describe spending differences among the leading funders; assessing concentration across all economy-related advertising requires accounting for expenditure by the remaining funding entities.

For *abortion*, Figure 4 shows a substantial spending gap among the five highest-spending funders in the observed corpus. Florida Voters Against Extremism[4] is the largest displayed spender, with estimated expenditure of approximately $62,700. The other four funders spend comparatively similar amounts: Last Best Place PAC ($18,000), Republic Restoratives Inc. ($17,500), Colin Allred for Senate ($17,500), and Montanans Securing Reproductive Rights[5] ($17,200). These results highlight differences in expenditure among the leading funders, with a pronounced drop from the largest spender to the remaining four. Assessing concentration across all abortion-related advertising requires accounting for expenditure by the remaining funding entities.

In contrast, for *crime/justice*, the two highest-spending funders have similar estimated expenditures: Preserve America PAC ($47,500) and Judicial Fairness PAC ($43,600). The remaining displayed funders—Smart and Safe Florida ($26,700), Stop Houston Murders PAC ($25,700), and Safer LA County (Hochman for DA) ($21,000)—have lower, broadly comparable expenditures. These amounts show a gap between the leading pair and the other three funders within the displayed group.

Overall, Figure 4 illustrates differences in expenditure patterns among the five highest-spending funders within each issue category. For *economy* and *abortion*, the largest displayed funder spends substantially more than each of the other four, whereas *crime/justice* has two leading funders with similar expenditures. These comparisons characterize spending differences among the leading funders in

[4] https://dos.elections.myflorida.com/committees

[5] `https://www.influencewatch.org/organization/montanans-securing-reproductive-rights/`

the observed corpus. They do not establish whether overall spending is more concentrated in one issue category or whether an issue involves a broader set of funding entities; evaluating those questions requires the complete funder-level expenditure distribution.

D.2 Moral Foundations in Funded Messaging

We examine the relationship between annotated moral framing and estimated advertising expenditure across three issue categories: *economy*, *crime/justice*, and *abortion*. Figure 5 presents the distribution of spending across moral foundations within each topic, together with the corresponding total expenditure. This aggregate view summarizes the moral-foundation composition of spending in the analyzed corpus without disaggregating it by funding entity.

Figure 4 complements this aggregate view by comparing the five highest-spending funders within each of the three issue categories. Each funder's expenditure is segmented according to the moral-foundation labels assigned to its advertisements. These funder-level comparisons describe the displayed organizations rather than the complete population of funders.

For the *economy*, *Fairness/Cheating* accounts for approximately 59% of topic-level expenditure, followed by *Liberty/Oppression* at 25%. The displayed funders differ in the composition of their spending. Stand with Crypto Alliance, Inc. has expenditure predominantly associated with *Liberty/Oppression*-annotated advertisements. In contrast, Americans for Prosperity Action, Inc., Americans for Prosperity, and Investing in Arkansas have expenditure primarily associated with *Fairness/Cheating*. The Congressional Leadership Fund's expenditure includes both *Fairness/Cheating* and *Liberty/Oppression*, while smaller *Loyalty/Betrayal* components appear for some displayed funders. These patterns show variation in the annotated moral framing associated with expenditure among the leading economy-related funders.

For *crime/justice*, the largest aggregate spending shares are associated with *Loyalty/Betrayal* (43%) and *Liberty/Oppression* (26%). The funder-level breakdown reveals different compositions within this issue category. Preserve America PAC's expenditure is predominantly associated with *Liberty/Oppression*, whereas Stop Houston Murders PAC's expenditure is predominantly associated with *Loyalty/Betrayal*. Judicial Fairness PAC has substantial expenditure associated with both *Authority/Subversion* and *Loyalty/Betrayal*. Thus, the aggregate distribution does not imply that all displayed funders exhibit the same moral-framing profile.

For *abortion*, *Liberty/Oppression* and *Sanctity/Degradation* account for approximately 48% and 34% of topic-level expenditure, respectively. Among the displayed funders, Florida Voters Against Extremism and Last Best Place PAC have expenditure primarily associated with *Sanctity/Degradation*. Republic Restoratives Inc. and Montanans Securing Reproductive Rights have expenditure predominantly associated with *Liberty/Oppression*. However, Colin Allred for Senate's displayed expenditure is associated with *Care/Harm*. The funder-level results therefore include additional moral foundations beyond the two largest aggregate categories and should not be reduced to a uniform split between liberty-based and sanctity-based framing.

Taken together, Figures 4 and 5 show differences in the moral-foundation composition of advertising expenditure across issue categories and among their highest-spending funders. These findings describe spending associated with LLM-generated annotations; they do not independently establish funder motives, specific policy arguments, or moral polarization. We therefore interpret them as exploratory patterns in the annotated corpus rather than as independently validated accounts of the organizations' moral or political positions.