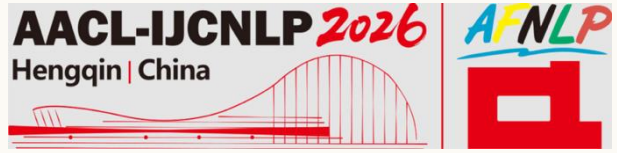
AACL-IJCNLP 2026



Iterative Topic Taxonomy Induction with LLMs

A Case Study of Electoral Advertising

Alexander Brady*

ETH Zurich

Tunazzina Islam*

Purdue University
Email: islam32@purdue.edu
Website: https://tunazislam.github.io/

*Equal contribution

ETH zürich

PURDUE UNIVERSITY



MOTIVATION

8,047 political ads ran on Facebook and Instagram in the month before Election Day 2024

IMMIGRATION

“Yadira Caraveo supported efforts to defund ICE and the border patrol. Now Colorado is being overwhelmed with illegal immigration.”

ABORTION

“Defend reproductive freedom - VOTE on Tuesday, November 5th!”

SEWER SALE

“Get the facts on the Gloucester Township sewer sale …”

What issues? · How are they framed? · Who sees them?

Example ads from the dataset (Meta Ad Library, October 2024)

THE GAP

Three ways current topic tools fall short

Topic models

LDA · NMF · BERTopic

Return keyword lists or clusters that still need human to name them.

Single-shot LLM labels

Our TopicGPT-style baseline

Name each cluster on its own, so labels for similar clusters need not agree.

Seeded LLM pipelines

e.g., Islam & Goldwasser 2025

Start from a seed set of themes or a predefined topic list.

GOAL A **seed-free**, **shared**, interpretable taxonomy induced from the corpus itself.

KEY IDEA

Grow one shared taxonomy. Then relabel every cluster against it.

No seed topics · no human labels · a 3B model running on a CPU

METHOD

Three stages, no seed topics

STAGE 1

Cluster

Embed ads with MiniLM, reduce with UMAP, cluster with HDBSCAN. The 5 ads with the highest membership probability represent each cluster.

PASS 1

Grow the taxonomy

Visit clusters in turn. The LLM answers yes or no: does an existing topic fit? Only on “no” does it write a new topic of 3 words or fewer.

PASS 2

Relabel

Show each cluster the final taxonomy. The LLM picks exactly one topic, so early clusters can take topics added later.

Constrained decoding for the yes/no and pick-one steps · Llama-3.2-3B run locally, no GPU · no seeds, no labeled data

PASS 1, STEP BY STEP

A topic is added only when nothing in the taxonomy fits

Cluster 1

Does an existing topic fit?

No existing topic matches.
Add new topic: Economy.

CURRENT TAXONOMY

- Economy

Cluster 2

Does an existing topic fit?

Matches existing topic (Economy).
No new topic added.

CURRENT TAXONOMY

- Economy

Cluster 3

Does an existing topic fit?

No existing topic matches.
Add new topic: Education.

CURRENT TAXONOMY

- Economy
- Education

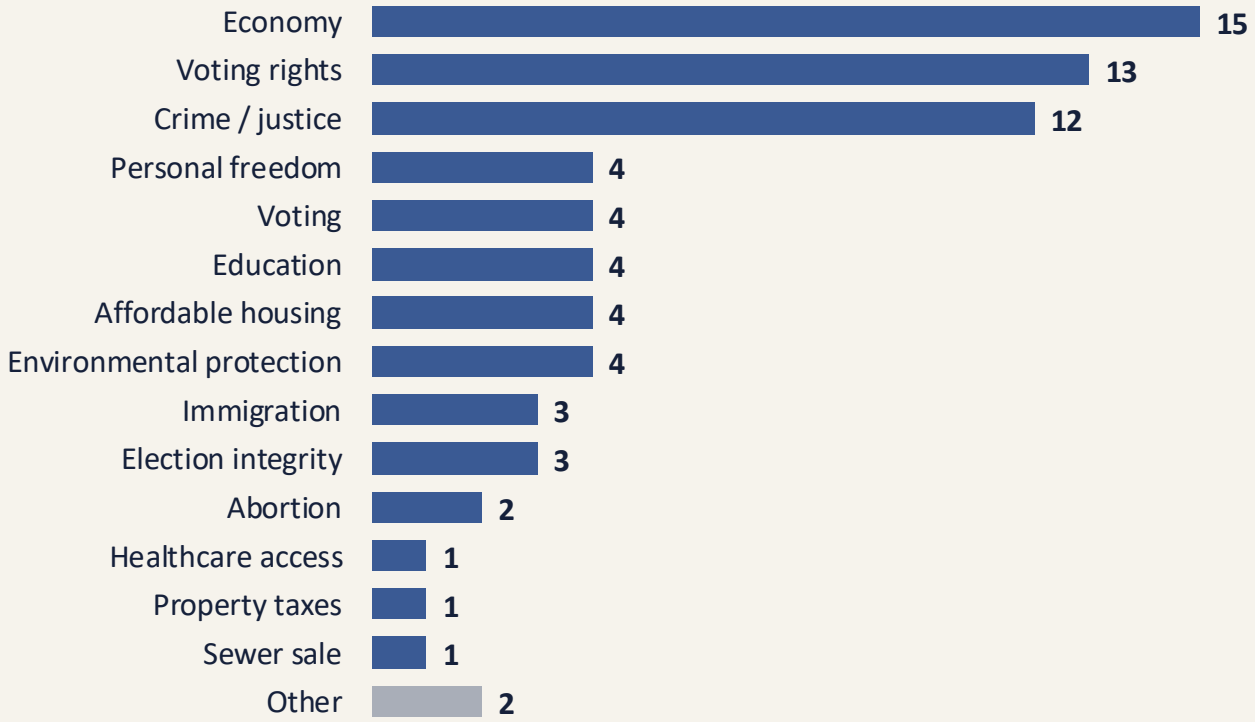
Pass 2 relabels every cluster against the final taxonomy, so early clusters can also get later topics.

Illustrative walk-through (paper Figure 1) · each cluster is shown to the LLM as its 5 representative ads

CASE STUDY · THE INDUCED TAXONOMY

73 clusters became 15 topics

Clusters assigned to each topic in pass 2



8,047
ads collected from the Meta Ad Library, October 2024

4,510
ads placed in clusters after near-duplicate removal

73
HDBSCAN clusters (min. size 15)

15
topics induced with no seeds; one, border security, got no clusters

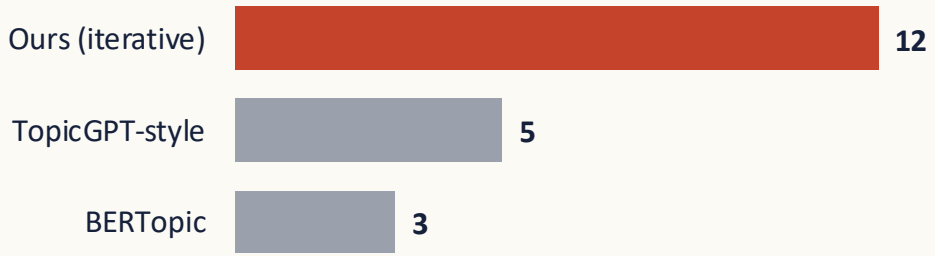
Paper Table 1 · clusters processed in ascending HDBSCAN id order

HUMAN EVALUATION

Annotators chose our labels as the best fit most often

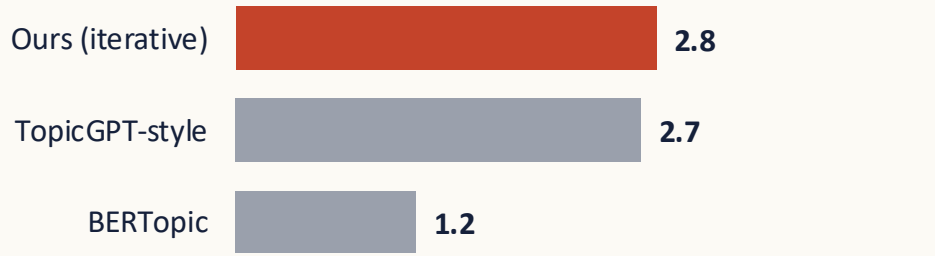
Picked as the best label

out of 20 judgments: 2 annotators × 10 ads



Mean fit rating

1 = irrelevant · 5 = perfect fit



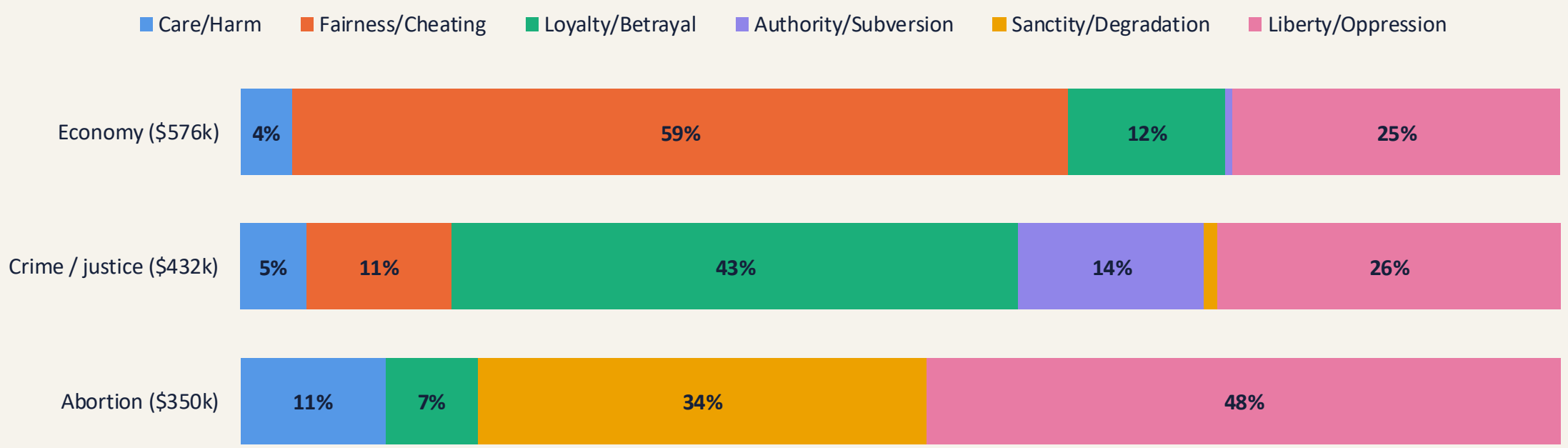
READ WITH CARE

A small study (Cohen's κ = 0.66). Mean ratings sit close to single-shot labeling; the clearer gap is in which label annotators preferred.

Same clusters and LLM for all three; labels shown in random order and each rated 1–5

CASE STUDY · MORAL FRAMING

Each issue leans on different moral foundations

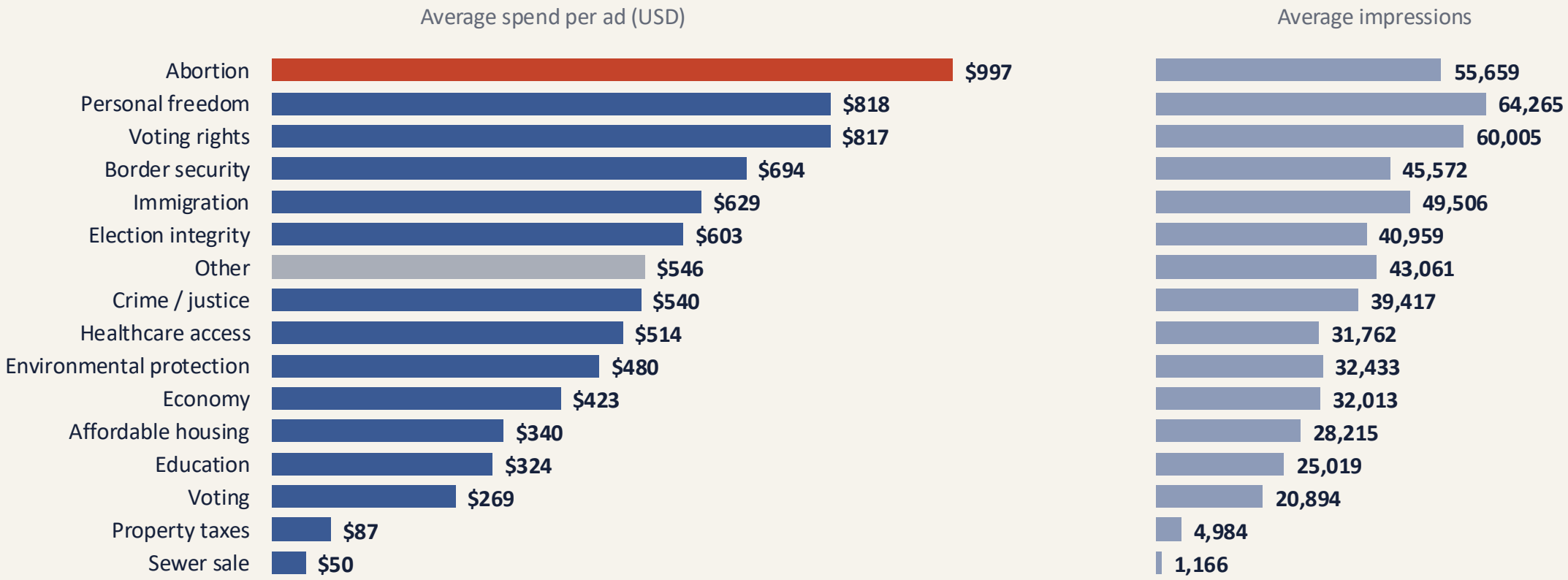


Share of estimated ad spend by moral foundation; slivers too small to label are drawn but unlabeled (economy authority, under 1%; crime/justice sanctity, about 1%). Foundations are LLM-assigned and conditioned on the topic label, so this is a descriptive layer, not validation of the taxonomy.

Paper Figure 5 · Moral Foundations Theory (Haidt & Graham 2007; Iyer et al. 2012)

CASE STUDY · AD SPEND

Abortion ads drew the highest spend per ad



Paper Figure 3 · midpoints of Meta's reported ranges · ad-level labels from Mistral-7B

CASE STUDY · WHO SAW WHICH ISSUES

Issue exposure differs by age and by state

Florida, by age

Strongest association per group

18–24	Affordable housing
55–64, 65+	Abortion

Men, by state

Strongest associations per group

Montana	Environmental protection, personal freedom
Virginia	Voting

EXPOSURE, NOT INTENT

PPMI between topic labels and Meta's reported audience delivery. It shows who saw the ads, not whom advertisers chose to target.

Paper Figure 6 · positive pointwise mutual information (PPMI)

WRAP-UP

Takeaways

1 **Seed-free:** a yes/no gate grows one shared taxonomy, then every cluster is relabeled against the final list.

2 **Preferred** over single-shot LLM labels and BERTopic in a small human study, with a 3B model on a CPU.

3 **Useful downstream:** the taxonomy supports analyses of moral framing, ad spend and audience exposure.

Open questions: sensitivity to cluster order · larger human evaluations · propagating cluster labels to unclustered ads (best macro-F1 0.37)

Code, prompts and dataset: github.com/alexander-brady/llm-topic-synthesis

Thank You

Questions?

ALEXANDER BRADY
bradya@ethz.ch

TUNAZZINA ISLAM
islam32@purdue.edu

CODE, PROMPTS, DATASET
github.com/alexander-brady/llm-topic-synthesis

SCAN ME