Supplementary Online Appendix for "Adding Fuel to the (Gun)Fire: How Politicians Polarize the Public Debate" by L. Zhang, G. Gratton, P. Grosjean and H. Yousaf

B Data Construction and Measurement

B.1 Mass-shooting events and tweet collection

We closely follow the data collection and pre-processing methodology of Demszky et al. (2019) to build our database of online public debates.

We use the Python package *snscrape*[1] to retrieve public, English-language tweets, excluding retweets. For each event listed in Table B.1, we define two sets of search terms. The first contains event-specific keywords, such as references to the event's location. The second contains event type specific keywords, i.e. terms related to mass shootings and other characteristics of the event. Columns 3 and 4 of Table B.2 reports the complete set of terms used. We keep all events with more than 3000 tweets after cleaning.

[1] https://github.com/JustAnotherArchivist/snscrape. *snscrape* is a commonly used alternative to the Twitter API for retrieving Twitter data; see Ridhwan and Hargreaves (2021); Yousefinaghani et al. (2021).

Table B.1: Mass shooting event characteristics

Event #	Event Name	Location	Event Start Time	Event End Time	Location Type	Shooter Ethnicity	# Deaths
1	UCLA shooting	California, CA	1/06/16 16:49	1/06/16 17:00	school	Non-White	3
2	Orlando nightclub shooting	Orlando, FL	12/06/16 6:02	12/06/16 9:14	community other	Non-White	80
3	Townville Elementary School shooting	Townville, SC	28/09/16 17:44	28/09/16 18:00	school	White	2
4	Palm Springs shooting	Palm Springs, CA	8/10/16 20:13	9/10/16 7:50	residential	Non-White	2
5	Fort Lauderdale airport shooting	Broward County, FL	6/01/17 17:53	6/01/17 17:54	other	Non-White	5
6	North Park Elementary School shooting	San Bernardino, CA	10/04/17 17:27	10/04/17 17:27	school	Non-White	3
7	Congressional baseball shooting	Alexandria, VA	14/06/17 11:09	14/06/17 11:20	other	White	1
8	Bronx-Lebanon Hospital attack	Bronx, NY	30/06/17 18:45	30/06/17 19:55	community other	Non-White	2
9	Las Vegas shooting	Paradise, NV	2/10/17 5:05	2/10/17 5:15	community other	White	61
10	Marshall County High School shooting	Benton, KY	23/01/18 13:57	23/01/18 14:06	school	White	2
11	Parkland high school shooting	Parkland, FL	14/02/18 19:21	14/02/18 19:27	school	White	17
12	Great Mills high school shooting	Great Mills, MD	20/03/18 11:55	20/03/18 11:55	school	White	2
13	YouTube headquarters shooting	San Bruno, CA	3/04/18 19:46	3/04/18 19:48	other	Non-White	1
14	Nashville Waffle House shooting	Nashville, TN	22/04/18 9:25	23/04/18 19:07	retail	White	4
15	Santa Fe High School shooting	Santa Fe, TX	18/05/18 12:32	18/05/18 13:02	school	White	10
16	Art All Night shooting	Trenton, NJ	17/06/18 6:45		community other	Non-White	1

Event #	Event Name	Location	Event Start Time	Event End Time	Location Type	Shooter Ethnicity	# Deaths
17	Capital Gazette shooting	Annapolis, MD	28/06/18 18:34	28/06/18 18:38	other	White	5
18	Jacksonville Landing shooting	Jacksonville, FL	26/08/18 17:30	26/08/18 17:36	community other	White	3
19	Aberdeen, Maryland shooting	Aberdeen, MD	20/09/18 13:06	20/09/18 13:09	other	Non-White	4
20	Pittsburgh synagogue shooting	Pittsburgh, PA	27/10/18 13:54	27/10/18 15:08	place of worship	White	11
21	Thousand Oaks shooting	Thousand Oaks, CA	8/11/18 7:18	8/11/18 7:38	retail	White	13
22	Mercy Hospital shooting	Chicago, IL	19/11/18 21:00	19/11/18 21:20	community other	Non-White	4
23	Sebring shooting	Sebring, FL	23/01/19 17:30	23/01/19 19:28	retail	White	5
24	Aurora, Illinois shooting	Aurora, IL	15/02/19 19:24	15/02/19 20:59	other	Non-White	6
25	Poway synagogue shooting	Poway, CA	27/04/19 18:23	27/04/19 18:33	place of worship	White	1
26	University of North Carolina at Charlotte shooting	Charlotte, NC	30/04/19 21:43	30/04/19 21:44	school	White	2
27	STEM School Highlands Ranch shooting	Douglas County, CO	7/05/19 19:53	7/05/19 20:07	school	White	1
28	Virginia Beach shooting	Virginia Beach, VA	31/05/19 20:08	31/05/19 20:44	other	Non-White	13
29	Gilroy Garlic Festival shooting (28/7/19, utc is 29/7)	Gilroy, CA	29/07/19 0:40	29/07/19 0:45	community other	White	4
30	El Paso shooting	El Paso, TX	3/08/19 16:39	3/08/19 16:45	retail	White	23
31	Dayton shooting	Dayton, OH	4/08/19 5:05	4/08/19 5:06	retail	White	10
32	Midland-Odessa shootings	Midland/Odessa, TX	31/08/19 20:17	31/08/19 21:20	other	White	8
33	Saugus High School shooting	Santa Clarita, CA	14/11/19 15:38	14/11/19 15:38	school	Non-White	3
34	Fresno shooting	Fresno, CA	18/11/19 2:00		residential	Non-White	4

Event #	Event Name	Location	Event Start Time	Event End Time	Location Type	Shooter Ethnicity	# Deaths
35	Naval Air Station Pensacola shooting	Pensacola, FL	6/12/19 11:43	6/12/19 23:58	other	Non-White	4
36	Jersey City shooting	Jersey City, NJ	10/12/19 17:21	10/12/19 20:47	retail	Non-White	6
37	West Freeway Church of Christ shooting	White Settlement, TX	29/12/19 16:50	29/12/19 16:50	place of worship	White	3
38	NYPD officers shooting	Bronx, NY	10/02/20 0:30	10/02/20 12:00		Non-White	0
39	Milwaukee brewery shooting	Milwaukee, WI	26/02/20 20:08	26/02/20 20:08	other	Non-White	6
40	Medical Clinic Attack	Buffalo, MN	9/02/21 16:52	9/02/21 17:00	community other	White	1
41	Atlanta Spa Shootings	Atlanta, GA	16/03/21 20:50	17/03/21 0:30	retail	White	8
42	Boulder Supermarket Shooting	Boulder, CO	22/03/21 20:30	22/03/21 21:28	retail	Non-White	10
43	Indianapolis FedEx Shooting	Indianapolis, IN	16/04/21 3:00	16/04/21 3:04	other	White	9
44	San Jose VTA Rail Yard Shooting	San Jose, CA	26/05/21 13:33	26/05/21 13:43	other	White	10
45	Collierville Kroger Shooting	Colleyville, TX	23/09/21 18:30	23/09/21 18:34	retail	Non-White	2
46	Oxford High School Shooting	Oakland, MI	30/11/21 17:51	30/11/21 17:55	school	White	4
47	Dumas Car Show Shooting	Dumas, AR			community other		1
48	Downtown Sacramento Shooting	Sacramento, CA	3/04/22 9:01	3/04/22 9:01	other		6
49	NYC Subway attack	Sunset Park, NY	12/04/22 12:24	13/04/22 17:42	community other	Non-White	0
50	Buffalo Supermarket Shooting	Buffalo, NY	14/05/22 18:31	14/05/22 18:36	retail	White	10
51	Laguna Woods Church Shooting	Orange County, CA	15/05/22 20:26	15/05/22 20:26	place of worship	Non-White	1
52	Robb Elementary School shooting	Uvalde, TX	24/05/22 16:28	24/05/22 17:50	school	Non-White	22

Event #	Event Name	Location	Event Start Time	Event End Time	Location Type	Shooter Ethnicity	# Deaths
53	Saint Francis Hospital Shooting	Tulsa, OK	1/06/22 21:56	1/06/22 22:00	community other	Non-White	5
54	South Street Philadelphia Shooting	Philadelphia, PA	5/06/22 3:31		other	Non-White	3
55	Chattanooga Shooting	Chattanooga, TN	5/06/22 6:42		other		3
56	Highland Park Parade Shooting	Chicago, IL	4/07/22 15:14	4/07/22 23:03	other	White	7
57	Greenwood Park Mall shooting	Greenwood, IN	17/07/22 21:56	17/07/22 21:56	retail	White	4

This table contains event-level characteristics for each event in our dataset. It contains the event name, location, start and end date and time of the mass shooting event, location type, shooter ethnicity, and number of deaths.

Table B.2: Mass shooting event retrieval and cleaning

Event #	Total Tweets	Event Specific Keywords	Event Type Specific Keywords	Politician Int	Celeb. Int
1	63304	ucla, mainak sarkar, mainaksarkar, william scott klug, williamscottklug, william klug, williamklug	MS*, engineering, sarkar	1	1
2	1027551	orlando, pulse nightclub, pulsenightclub, omar mateen, omarmateen	MS*, pulse, nightclub	1	1
3	6049	townville, jesse osborne, jesseosborne	MS*, elementary, school	1	0
4	15487	palm springs, palmsprings, john felix, johnfelix, jose gilbert vega, josegilbertvega, josevega, jose vega, lesley zerebny, lesleyzerebny	MS*, police, officer	1	0
5	56373	broward county, santiago-ruiz, fort lauderdale	MS*, airport, santiago	1	1
6	30884	north park elementary, northpark, san bernardino, sanbernardino	MS*, school	1	1
7	56161	alexandria, congressional baseball, james hodgkinson, jameshodgkinson, eugene simpson stadium	MS*, hodgkinson	1	0
8	15697	bronx-lebanon hospital, bronx, bronx lebanon, bronxlebanon	MS*, hospital	1	0
9	582748	nevada, las vegas, lasvegas, steven paddock, stevenpaddock, route 91 harvest	MS*, route 91, harvest, paddock, festival	1	1
10	6710	draffenville, marshall county, marshallcounty	MS*, school	1	0
11	176467	stoneman douglas, stonemandouglas, parkland, nikolas cruz, nikolascru, nikolas jacob cruz, nikolasjacobcruz	MS*, school, marjory, stoneman	1	1
12	12681	st. mary's county, st marys, st mary's, great mills, greatmills	MS*, school	1	0
13	40519	san bruno, sanbruno, youtube headquarters, youtube hq	MS*	1	0
14	17491	nashville, waffle house, antioch, travis jeffrey reinking, travisjeffreyreinking	MS*, waffle	0	0
15	60898	santa fe, santafe, dimitrios pagourtzis, dimitriospagourtzis	MS*, school	1	0
16	3579	trenton, art all night, artallnight, amir armstrong, amirarmstrong, davonewhite, davone white, tahaij wells, tahaijwells	MS*, festival	0	0
17	84015	annapolis, capital gazette, capitalgazette, jarrod ramos, jarrodramos	MS*, capital	1	0

Event #	Total Tweets	Event Specific Keywords	Event Type Specific Keywords	Political Int	Celeb. Int
18	68292	jacksonville, david katz, davidkatz, elijah clayton, elijahclayton, taylor robertson, taylorrobertson, madden nfl 19, maddennfl19	MS*, madden, tourney, tournament, twitch, stream, game, qualifier	1	0
19	6570	aberdeen, snochia moseley, snochiamoseley, riteaid, rite aid	MS*	1	0
20	214476	pittsburgh, robert gregory bowers, robertgregorybowers, tree of life, treeoflife	MS*, synagogue	1	0
21	61777	thousand oaks, thousandoaks, ian david long, iandavidlong	MS*	1	1
22	8921	mercy hospital, mercyhospital, juan lope, juanlopez	MS*	0	0
23	5476	sebring, zephen allen xaver, zephenallenxaver, zephen xaver, zephenxaver	MS*, bank, suntrust	1	0
24	23044	aurora, gary montez martin, garymontezmartin, henry pratt company, henryprattcompany	MS*	1	0
25	21700	poway, john timothy earnest, johntimothyearnest, johnearnest, john earnest, lori gilbert-kaye, lorigilbert-kaye	MS*, synagogue	1	0
26	36183	charlotte, unc, trystan andrew terrell, trystanandrewterrell, riley howell, rileyhowell, ellis parlier, ellisparlier	MS*, school, unc	1	0
27	6455	highlands ranch, highlandsranch, alec mckinney, alecmckinney, devon erickson, devonerickson	MS*, school, academy	1	0
28	41316	virginia beach, virginiabeach, dewayne craddock, dewaynecraddock	MS*, courthouse	1	0
29	56500	gilroy, garlic festival, santino william legan, santinowilliamlegan	MS*, garlic, festival	1	0
30	317337	el paso, elpaso, patrick wood crusius, patrickwoodcrusius, patrickcrusius, patrick crusius	MS*, walmart	1	0
31	173004	dayton, connor betts, connorbetts, connor stephen betts, connorstephenbetts	MS*	1	0
32	34385	odessa, midland odessa, midlandodessa, seth ator, sethator, seth aaron ator, sethaaronator	MS*	1	0
33	26196	santa clarita, santaclarita, saugus high school, saugushighschool, nathaniel berhow, nathanielberhow	MS*, school, saugus	1	1
34	5214	fresno	MS*	1	0

Event #	Total Tweets	Event Specific Keywords	Event Type Specific Keywords	Political Int	Celeb. Int
35	35692	pensacola, mohammed saeed alshamrani, mohammed-saeedalshamrani, mohammed alshamrani, mohammedal-shamrani	MS*	1	0
36	27034	jersey city, jerseycity, david anderson, davidanderson, francine graham, francinegraham, joseph seals, joseph-seals	MS*	1	0
37	8278	white settlement, west freeway church, westfreeway-church, whitesettlement, keith thomas kinnunen, kei-ththomaskinnunen, jack wilson, jackwilson	MS*, church	1	0
38	3672	bronx, robert williams, robertwilliams	MS*, cop, police, nypd	1	0
39	17619	milwaukee, molson coors, molsoncoors, anthony ferrill, anthonyferrill	MS*, miller, molson, campus	1	0
40	3865	allina health, buffalo, buffalo crossroads	MS*, allina, clinic	1	0
41	88862	cherokee, acworth, atlanta, gold spa	MS*	1	1
42	87600	boulder	MS*, supermarket, king soopers	1	0
43	16212	fedex facility, indiana	MS*, fedex	1	0
44	21877	santa clara, vta, valley transportation authority, san jose	MS*, vta, yard	1	0
45	10613	collierville, tennessee, kroger, memphis	MS*	0	0
46	24805	oxford high, oxfordhighschool, oxford high school	MS*	1	0
47	4641	arkansas, dumas	MS*	0	0
48	23962	sacramento	MS*	1	0
49	50333	sunset park, brooklyn	MS*, subway, train, station	1	0
50	215514	buffalo	MS*, supermarket	1	1
51	8103	laguna woods	MS*	1	0
52	521065	uvalde, robb elementary	MS*	1	1
53	32031	saint francis hospital, tulsa	MS*	1	0
54	15376	south street, philadelphia	MS*	1	0
55	3749	chattanooga	MS*	1	0
56	153700	highland park	MS*	1	0
57	10467	greenwood park, greenwood	MS*, mall	1	0

This table contains information about the total number of tweets in each event, Event Specific Keywords and Event Type Specific Keywords. Political Int =1 if there is an intervention by

a politician and Celeb = 1 if a non-political celebrity intervenes. MS* = gun, shoot, kill, attack, massacre, victim, terror, violence, crime. Keywords are chosen to reflect the methodology of Demszky et al. (2019).

B.2 Dictionary construction and keyword frequencies

B.2.1 List of Partisanship phrases identified by Gentzkow et al. (2019) between 2005-2016 Congressional Sessions

afford care, african american, al qaeda, american energi, american peopl, busi owner, care act, care bill, care plan, care reform, children health, civil war, clean air, climat chang, colleagu join, colleagu support, comprehens immigr, continent shelf, credit card, death tax, depart homeland, dog coalit, employ mandat, farm bill, fornia madam, general kagan, god pleas, govern spend, govern takeov, gun violenc, hate crime, homeland secur, human traffick, hurrican katrina, illeg immigr, immigr reform, insur compani, interest rate, job creation, job creator, men women, mental health, middl class, million american, minimum wage, muslim brotherhood, nation debt, nation guard, natur gas, nobid contract, oil compani, outer continent, plan parenthood, pleas bless, presid health, progress caucus, public health, puerto rico, radic islam, rais tax, recoveri act, religi freedom, reserv balanc, side aisl, stem cell, stimulus bill, student loan, tax break, tax increas, tax rate, tax relief, taxpay dollar, troop iraq, unemploy benefit, unemploy insur, vote right, war iraq, war terror

B.2.2 List of Tribalism words from the loyalty/betrayal dimension of Moral Foundation theory by Graham et al. (2009) as used by Enke (2020)

abandon, ally, apostasy, apostate, betray, cadre, cliqu, cohort, collectiv, communal, commune, communis, communit, comrad, deceiv, deserted, deserter, deserting, devot, disloyal, enem, familial, families, family, fellow, foreign, group, guild, homeland, immigra, imposter, individual, insider, jilt, joint, loyal, member, miscreant, nation, patriot, renegade, segregat, sequester, solidarity, spy, terroris, together, traitor, treacher, treason, unison, unite

B.3 Construction and validation of the model-based issue-polarization measure

To verify that our main results are not an artifact of dictionary definitions, we build a second, semantic measure of issue polarization using a supervised language model trained on large-language-model (LLM) annotations. The construction proceeds in five stages: (1) a dictionary pre-filter to concentrate annotation effort on plausibly issue-relevant tweets; (2) stratified sampling for labeling; (3) labeling with GPT-4.1-mini under a structured prompt; (4) fine-tuning a DeBERTa-v3 classifier on those labels and scoring the full corpus of 4.75 million tweets; and (5) converting the model outputs into binary indicators,

with probabilities calibrated by isotonic regression. Each tweet is characterized along two dimensions: an *issue domain* (one of ten categories, including "none") and an *ideological direction* (left / neutral / right). The headline indicator, *polarized*, equals one when a tweet takes a non-neutral ideological stance within an identified issue domain.

Conceptual definition and label space. A tweet exhibits *issue polarization* when it expresses a one-sided stance on a specific political or social issue aligned with partisan/ ideological camps, typically with zero-sum, moralized, or conflict framing. Pure grief, generic hostility toward a group, and neutral news reporting are *not* coded as issue polarization unless tied to an issue stance. Tweets are assigned to one of ten issue domains: gun *regulation, crime policing, mental health, race and hate crime, immigration, terror and national security, gender LGBT, school policy, elite media blame,* or *none*, and an ideological direction coded: -1 (US left/liberal), $+1$ (US right/ conservative), or 0 (neutral, mixed, or not inferable; forced to 0 whenever the domain is *none*).

Stage 1: Dictionary pre-filter. We first match every tweet against an issue-domain phrase dictionary using the Aho-Corasick multi-pattern string-matching algorithm, recording, for each tweet, the matched domain(s), the number of matches, and the event. Of the $5{,}098{,}220$ tweets read, $1{,}850{,}197$ (36.3%) contain at least one dictionary trigger. The pre-filter is used only to allocate labeling budget toward issue-relevant content; it does not by itself define the measure, and the labeling sample deliberately retains a fraction of non-matching tweets (Stage 2) so the downstream classifier also sees genuine negatives.

Stage 2: Stratified sampling for labeling. From the pre-filtered pool we draw a labeling sample designed to cover rare domains and many distinct events. We oversample the rare domains (*gender LGBT, immigration, terror and national security*), balance across events, and include a 10% admixture of non-hit (likely-neutral) tweets so the classifier is not trained only on dictionary-positive text. The procedure produces a pilot sample of $2{,}000$ tweets and a full sample of $50{,}000$ tweets ($45{,}000$ dictionary hits $+$ $5{,}000$ non-hits; fixed random seed). The pilot is labeled and inspected first to validate the schema and rare-domain coverage before committing to the full run.

Stage 3: LLM labeling with GPT-4.1-mini. The sampled tweets are labeled by GPT-4.1-mini at temperature 0, using a single expert-annotator system prompt with explicit domain definitions, tie-break rules, and worked examples, and a per-tweet user prompt. Outputs are constrained to a strict JSON schema (enumerated *issue domain* and *ideology*), submitted through the Batch API. In the full labeled sample, *issue domain* is distributed as *none* 33.5%, *gun regulation* 24.5%, *race hatecrime* 8.7%, *terror and national security* 8.5%, *elite media blame* 8.3%, *gender LGBT* 6.1%, *crime policing* 4.5%, *immigration* 2.3%, *school policy*

2.2%, *mental health* 1.4%; ideology is 56.7% neutral, 27.7% left, 15.6% right.

System Prompt for LLM Annotation

```
You are an expert political communication researcher.

Task:
Given a tweet, assign:
1) issue_domain ∈ {gun_regulation, crime_policing, mental_health, race_hatecrime, immigration,
    terror_nationalsecurity, gender_lgbt, school_policy, elite_media_blame, none}
2) ideology ∈ {-1, 0, 1}

Core Principle:
Identify the PRIMARY issue frame expressed in the tweet. If multiple issues appear, choose the
    domain that reflects the main causal or policy claim.

------------------------------------------------------------
DEFINITIONS

Issue polarization is present when the tweet expresses a one-sided stance on a specific political
    /social issue aligned with partisan/ideological camps, often using zero-sum, moralized, or
    conflict framing.

Do NOT classify:
- Pure grief, sympathy, or sadness without an issue stance.
- Pure insults or hostility toward groups unless it advances a specific issue position.

If the tweet is primarily grief/anger/insults without a discernible issue stance:
→ issue_domain = "none"
→ ideology = 0

------------------------------------------------------------
TIE-BREAK RULES (when multiple domains appear)

1. If the tweet makes a concrete policy demand (ban X, arm teachers, close borders), prefer the
    domain tied to that policy action.
2. If both blame and policy appear, prefer the domain tied to policy over elite_media_blame.
3. If racial or hate ideology is the central explanatory claim (e.g., white supremacy caused this
    ), prefer race_hatecrime.
4. If policing response, sentencing, or prosecution is central, prefer crime_policing.
5. If media/politician blame is the dominant frame without a policy proposal, use
    elite_media_blame.
6. If no clear issue stance dominates, use "none".

Always choose the BEST SINGLE domain.

------------------------------------------------------------
ISSUE DOMAIN GUIDANCE

- gun_regulation: gun bans, background checks, concealed carry, 2A rights, assault weapons.
- crime_policing: law enforcement strength/weakness, sentencing, prosecutors, "law and order."
- mental_health: mental illness as cause or healthcare framing as solution.
- race_hatecrime: hate crime framing, white supremacy, racial ideology, anti-Asian hate, BLM.
- immigration: borders, deportations, asylum, migrant crime.
- terror_nationalsecurity: terrorism, homeland security, FBI/foreign threats.
```

```
- gender_lgbt: LGBTQ/trans/culture war framing.
- school_policy: school security, armed teachers, SROs, curriculum blame.
- elite_media_blame: media/political blame without a primary policy claim.
- none: no clear issue stance.

------------------------------------------------------------
IDEOLOGY CODING

ideology = -1 → US liberal/left alignment within the chosen issue_domain
ideology =  1 → US conservative/right alignment within the chosen issue_domain
ideology =  0 → neutral/mixed/unclear OR ideology not inferable from tweet alone OR issue_domain
    ="none"

If ideology is ambiguous from the text alone, use 0.

------------------------------------------------------------
EXAMPLES (Scored)

Example 1:
Tweet: "@tedcruz @IlhanMN @nytimes This is Lori Gilbert Kaye who was killed yesterday in America
    by an American white supremacist. Nothing to say about that literal anti-Semitic murder that
    happened right here? Fuck you and your FALSE OUTRAGE THAT IS LITERALLY POLITICAL SPIN."
Output:
issue_domain = "race_hatecrime"
ideology = -1

Example 2:
Tweet: "So UCLA incident is murder-suicide. Bet mental health issues are still not made into
    focus of chatter but gun control will be."
Output:
issue_domain = "mental_health"
ideology = 1

Example 3:
Tweet: "@tariqnasheed The FBI on them white supremacists sites be funding them. The dems want gun
     control to be the topic but can't use buffalo because they're Black and we will start
    demanding protections. So they sent this patsy to get Black people out the news cycle."
Output:
issue_domain = "elite_media_blame"
ideology = 1

Example 4:
Tweet: "(Reuters) #Orlando gunman told 911 he was strapping explosive vests to hostages..."
Output:
issue_domain = "none"
ideology = 0

Example 5:
Tweet: "Report: Father saved others from gunfire in Las Vegas before being shot in neck - CBS
    News https://t.co/SPqXrP9Ik4 https://t.co/6BoAvtjcXl"
Output:
issue_domain = "none"
ideology = 0

Output MUST be strictly valid JSON that matches the provided JSON Schema.
```

```
Return only the JSON object.
```

Stage 4: Fine-tuning DeBERTa-v3 and scoring the corpus. Labeling 4.75 million tweets directly with a frontier LLM is cost-prohibitive, so we distill the GPT-4.1-mini labels into a supervised classifier. We fine-tune `microsoft/deberta-v3-base` with two heads—a 10-class `issue domain` classifier and a 3-class `ideology` classifier—on the LLM-labeled sample, using stratified train/validation/test splits (max sequence length 128, batch size 16, learning rate 2×10^{-5}, 4 epochs, fixed seed). We evaluate with macro-F1, per-class precision/recall, and confusion matrices on the held-out splits. The trained models then score the entire corpus, producing, for every tweet, a full probability vector over issue domains and over ideological directions.

As an internal consistency check, DeBERTa's predictions reproduce the GPT-4.1-mini labels on the $50{,}000$-tweet labeled set with 94.0% agreement on `issue domain` (Cohen's $\kappa = 0.926$), 95.1% on the three-way ideology label ($\kappa = 0.914$), and 96.3% on the binary polarized indicator ($\kappa = 0.925$). Because this set includes training rows, these figures are an upper bound on fit.

On the held-out test split ($5{,}000$ tweets per task, never seen in training or validation), DeBERTa reproduces the GPT-4.1-mini labels with macro-F1 of 0.813 for `issue domain` (accuracy 0.862, $\kappa = 0.829$) and 0.842 for the three-way `ideology` label (accuracy 0.875, $\kappa = 0.782$); the derived binary polarized indicator reaches accuracy 0.905, F1 $= 0.889$, $\kappa = 0.806$ (Table B.3). As expected, these generalization figures sit below the in-sample agreement above, confirming the internal-consistency numbers are an upper bound rather than held-out performance.

Table B.3: DeBERTa held-out test-split performance ($n = 5{,}000$ per task).

Task	Accuracy	Macro-F1	Cohen's κ
`issue_domain` (10-class)	0.862	0.813	0.829
`ideology` (3-class)	0.875	0.842	0.782
polarized (binary, derived)	0.905	0.889	0.806

Held-out test split from the fixed-seed stratified partition; per-class F1 and confusion matrices.

Per-class F1 on the held-out test split is highest for the common, well-defined domains—*gun regulation* (0.92), *none* (0.89), *race hate crime* (0.86), *gender LGBT* (0.84), *terror and national security* (0.83), *immigration* (0.82)—and lowest for the rarest domains, *school policy* (0.67, $n = 109$) and *mental health* (0.73, $n = 68$). For ideology, *neutral* (F1 $= 0.92$) and *left* (0.85) are recovered more reliably than the rarer *right* class (0.76).

Table B.4: Threshold sensitivity of the *polarized* measure (held-out test, raw probabilities, $n = 5{,}000$).

Threshold	Accuracy	Precision	Recall	F1	Pred. positive rate
0.3	0.901	0.881	0.893	0.887	0.439
0.4	0.903	0.889	0.887	0.888	0.432
0.5	0.905	0.899	0.880	0.890	0.424
0.6	0.905	0.905	0.871	0.888	0.417
0.7	0.903	0.912	0.858	0.884	0.408

Stage 5: Binary measures and probability calibration. We convert the classifier's probabilities into the binary variables used in the analysis: *polarized* ($= 1$ if the tweet is left or right, i.e. non-neutral), *ideology_left*, *ideology_right*, and a *domain_** indicator for each issue domain. Raw model probabilities are recalibrated on the labeled data using isotonic regression (fit on an 80% calibration split, validated on the remaining 20%), separately for the polarization, ideology-left, ideology-right, and per-domain measures, so that predicted probabilities match empirical rates and a common 0.5 threshold is interpretable.

On the held-out test split ($n = 5{,}000$; empirical polarized rate 0.433), the raw DeBERTa polarized probability is sharply bimodal—54% of tweets fall in the lowest probability bin (mean predicted 0.004, mean actual 0.067) and 38% in the highest (mean predicted 0.992, mean actual 0.930)—and increases monotonically with the empirical rate, indicating the model separates polarized from neutral tweets well but is mildly over-confident at the extremes; isotonic recalibration corrects this. The binary measure is insensitive to the exact cutoff: accuracy stays within 0.901–0.905 and F1 within 0.884–0.890 across thresholds from 0.3 to 0.7, peaking at 0.5–0.6 (Table B.4), supporting a common 0.5 threshold. The final dataset, with both calibrated scores and binary indicators, is exported to Stata and merged with the rest of the project data.

B.3.1 External validation and robustness to the annotating model

We validate the measure against human annotation, comparing the production pipeline (GPT-4.1-mini $\rightarrow$ DeBERTa) against human annotated and against three independent open-weight LLMs, to show that the measure is not an artifact of the specific model used for labeling. Two hand-annotated samples anchor the exercise: (i) a 251-tweet sample (single annotator; human binary polarized and left/neutral/right ideology available for ≈ 150) used for the polarized and ideology comparisons, and (ii) a domain-balanced 100-tweet sample, hand-labeled on the full 10-domain taxonomy, used for the `issue_domain` comparison. GPT-4.1-mini, DeBERTa, and the three open models are all scored on the same 251 tweets; the open models were not run on the balanced-100, so their domain cells

Table B.5: Agreement with human annotation, by annotator and dimension.

	Polarized		Ideology (3-way)		Domain	
Annotator	Acc	κ	Acc	κ	Acc	κ
GPT-4.1-mini (production labeler)	0.873	0.746	0.813	0.667	0.830	0.811
DeBERTa-v3 (distilled)	0.901	0.802	0.796	0.635	0.820	0.799
gemma4:31b	0.893	0.787	0.813	0.683	—	—
qwen3.5:35b	0.873	0.746	0.793	0.637	—	—
llama3.1:8b	0.880	0.760	0.727	0.555	—	—
Dictionary (baseline)	0.612	0.220	—	—	—	—

Accuracy and Cohen's κ against human annotation. Polarized and 3-way ideology are on the 251-tweet sample ($n \approx 150$–152 with human labels); domain is on the balanced-100 ($n = 100$). Open models were not run on the balanced-100 (—). Confusion matrices in `paper/validation_tables/`.

are blank.

The same prompt (see "System Prompt for LLM Annotation" above) was used across all models. Results are in Table B.5. First, on the main binary polarized measure, every LLM—the production GPT-4.1-mini labeler, the distilled DeBERTa, and all three open models—agrees with human coders at $\kappa = 0.75$–0.80, far above the dictionary baseline ($\kappa = 0.22$); the semantic measure is therefore robust to the choice of annotating model rather than specific to GPT-4.1-mini. Second, both models validated on domain (GPT-4.1-mini and DeBERTa) recover the 10-way `issue_domain` taxonomy against human labels at $\kappa \approx 0.80$, and DeBERTa's distilled labels track its GPT-4.1-mini teacher closely. Third, ideology direction is the hardest dimension for every model ($\kappa \approx 0.55$–0.68), driven by the rare `right`-leaning class.

B.4 Politicians

Table B.6: Politicians' Twitter Accounts

Twitter Username	# Followers	Party	Gender	Role	State/ Territory	Account Type
alfranken	1.1 M	D	M	Senator	Minnesota	Personal
amyklobuchar	1.9 M	D	F	Senator	Minnesota	Personal
aoc	13 M	D	F	Representative	New York	Personal
ayannapressley	1.1 M	D	F	Representative	Massachusetts	Personal
barackobama	132 M	D	M			Personal
berniesanders	15.5 M	I	M	Senator	Vermont	Personal
betoorourke	2.4 M	D	M	Representative	Texas	Personal
chrismurphyct	1.1 M	D	M	Senator	Connecticut	Office
corybooker	4.9 M	D	M	Senator	New Jersey	Personal
dancrenshawtx	1.2 M	R	M	Representative	Texas	Personal

devinnunes	1.3 M	R	M	Representative	California	Official
ewarren	5.9 M	D	F	Senator	Massachusetts	Personal
gavinnewsom	2 M	D	M	Governor	California	Personal
gopleader	1.5 M	R	M			
govrondesantis	2.5 M	R	M	Representative, Governor	Florida	Official
hillaryclinton	31.5 M	D	F			Personal
ilhan	1.3 M	D	F	Representative	Minnesota	Office
ilhanmn	3.1 M	D	F	Representative	Minnesota	Personal
jerrybrowngov	1 M	D	M	Governor	California	Office
jim_jordan	2.8 M	R	M	Representative	Ohio	Official
joebiden	34.4 M	D	M	VP, President	Delaware	Personal
johnkerry	3.4 M	D	M	Cabinet		Personal
kamalaharris	19.7 M	D	F	Senator, VP	California	Official
laurenboebert	1.4 M	R	F	Representative	Colorado	Personal
leadermcconnell	2.1 M	R	M	Senator	Kentucky	Official
lindseygrahamsc	2.1 M	R	M	Senator	South Carolina	Official
marcorubio	4.4 M	R	M	Senator	Florida	Personal
markmeadows	1 M	R	M	Representative	North Carolina	Official
mattgaetz	1.5 M	R	M	Representative	Florida	Personal
mike_pence	5.8 M	R	M	VP, Governor	Indiana	Personal
mikepompeo	1.4 M	R	M	Representative	Kansas	Personal
mittromney	2.1 M	R	M	Senator	Utah	Personal
nycmayor	1.6 M	D	M	Mayor	New York	Official
nygovcuomo	2.4 M	D	M	Governor	New York	Office
ossoff	1.3 M	D	M	Senator	Georgia	Personal
petebuttigieg	3.6 M	D	M	Cabinet		Personal
potus	22.7 M	D	M	President		Office
presssec	2.4 M	D	F			Office
presssec45	5.8 M	R	F			Office
randpaul	3.9 M	R	M	Senator	Kentucky	Personal
rashidatlaib	1.4 M	D	F	Representative	Michigan	Personal
realbencarson	2.2 M	R	M	Cabinet		Personal
realdonaldtrump	87.4 M	R	M	President		Personal
repadamschiff	3.1 M	D	M	Representative	California	Office
repjohnlewis	1.1 M	D	M	Representative	Georgia	Official
repkatieporter	1.2 M	D	F	Representative	California	Office
repmattgaetz	1.6 M	R	M	Representative	Florida	Official
repmaxinewaters	1.6 M	D	F	Representative	California	Office
repswalwell	1.3 M	D	M	Representative	California	Office
reverendwarnock	1.1 M	D	M	Senator	Georgia	Personal

rondesantis	1.8 M	R	M	Representative, Governor	Florida	Personal
secblinken	1.5 M	D	M	Cabinet		Office
secpompeo	2.9 M	R	M	Representative, Cabinet	Kansas	Office
senfeinstein	1.4 M	D	F	Senator	California	Official
sengillibrand	1.6 M	D	F	Senator	New York	Official
senjohnmccain	2.8 M	R	M	Senator	Arizona	Official
sensanders	12.5 M	I	M	Senator	Vermont	Office
senschumer	3.4 M	D	M	Senator	New York	Office
sentedcruz	2.7 M	R	M	Senator	Texas	Official
senwarren	7.1 M	D	F	Senator	Massachusetts	Office
speakermccarthy	2 M	R	M	Representative	California	Official
speakerpelosi	7.4 M	D	F	Representative	California	Office
speakerryan	3.4 M	R	M	Speaker	Wisconsin	Office
tedcruz	5.2 M	R	M	Senator	Texas	Personal
tedlieu	1.6 M	D	M	Representative	California	Personal
tgowdysc	1.2 M	R	M	Representative	South Carolina	Personal
tulsigabbard	1.5 M	D	F	Representative	Hawaii	Personal
vp	12.9 M	R	M	Senator, VP	California	Official
vp45	9.9 M	R	M	Vice President		Office
whitehouse	7.1 M	D				Office

Institutional accounts change holder during the sample period. Party and gender in the analysis follow the holder at the time of each intervention, rather than a permanent account classification. In this sample, `vp` spans Mike Pence (Republican, male, 2017–2020) and Kamala Harris (Democrat, female, 2021–2022).

This table contains the $N = 70$ politicians' Twitter accounts which have over one million followers (as of July 11, 2022). In Party column, D = Democrat, R = Republican, I = Independent. In Gender column, F = Female, M = Male. For accounts that are used by sitting Party, the relevant Party at the time of intervention is used, e.g., whitehouse is Republican during the time period 2017-2021.

Table B.7: Celebrities' twitter accounts

Twitter Username	Followers
elonmusk	159.8 M
justinbieber	111.7 M
cristiano	109.9 M
rihanna	108.5 M
katyperry	107.2 M
taylorswift13	94.6 M
arianagrande	85.3 M
ladygaga	83.9 M
kimkardashian	75.3 M
ellendegeneres	75.1 M
selenagomez	66.7 M
billgates	64 M
neymarjr	62.9 M
jtimberlake	61.4 M
imvkohli	58.4 M
britneyspears	55.1 M
shakira	53.8 M
ddlovato	53 M
kingjames	52.7 M
jimmyfallon	50.2 M
bts_twt	48.6 M
mileycyrus	46.5 M
akshaykumar	46.2 M
beingsalmankhan	45.4 M
jlo	44.9 M
iamsrk	43.8 M
bts_bighit	43.6 M
brunomars	42.6 M
oprah	42.1 M
niallofficial	40.4 M
kyliejenner	40.2 M
drake	39.4 M
sachin_rt	39.2 M
harry_styles	37.7 M
kevinhart4real	37 M
wizkhalifa	36.3 M
louis_tomlinson	35.3 M
liltunechi	34.4 M
liampayne	33.5 M
iamcardib	32.3 M
ihrithik	32.3 M
kendalljenner	31.9 M
kanyewest	31.7 M
chrisbrown	31.6 M
pink	30.9 M
zaynmalik	30.5 M
khloekardashian	30.3 M
aliciakeys	29.5 M
kaka	29.3 M
nickiminaj	28.1 M
conanobrien	27.7 M
priyankachopra	27.7 M
emmawatson	27.6 M
adele	27.4 M
whindersson	27.1 M
deepikapadukone	26.9 M
aamir_khan	26.8 M
kourtneykardash	26.5 M
m10	26.4 M
shawnmendes	26.1 M
andresiniesta8	25.4 M

We retrieve the list of the top 100 accounts from the social media analytics website Social Blade: https://socialblade.com/twitter/top/100 retrieved 14 October 2023. We then eliminate the accounts of politicians (including non-U.S. and non-active/past politicians) and organizations. This leaves us with a list of 61 accounts.

Table B.8: News media organizations twitter accounts

Twitter Username	Followers	Twitter Username	Followers
cnnbrk	63.9 M	SkyNewsBreak	4.9 M
CNN	61.8 M	politico	4.6 M
nytimes	55.1 M	CNNPolitics	4.5 M
BBCBreaking	51.7 M	thehill	4.4 M
espn	49.1 M	BusinessInsider	4 M
SportsCenter	41.9 M	latimes	3.9 M
BBCWorld	40.3 M	guardiannews	3.8 M
TheEconomist	27.2 M	Independent	3.6 M
Reuters	25.7 M	CBCNews	3.5 M
FoxNews	24.3 M	Newsweek	3.5 M
WSJ	20.6 M	nypost	3 M
washingtonpost	20 M	nprpolitics	2.9 M
Forbes	19.7 M	Variety	2.9 M
TIME	19.3 M	MailOnline	2.8 M
ABC	17.8 M	Xnews	2.7 M
AP	16 M	itvnews	2.6 M
BBCNews	15.2 M	Channel4News	2.4 M
cnni	14.7 M	abcnews	2.3 M
SkySportsNews	12.5 M	AJENews	2.2 M
XHNews	11.9 M	PBS	2.1 M
enews	11.7 M	usweekly	2.1 M
TheOnion	11.6 M	BreitbartNews	2 M
guardian	10.9 M	OANN	2 M
HuffPost	10.9 M	TheSun	2 M
TimesNow	10.2 M	CTVNews	1.9 M
BreakingNews	9.4 M	VICE	1.9 M
mashable	9.4 M	foxnewspolitics	1.8 M
NBCNews	9.4 M	YahooFinance	1.7 M
ABSCBNNews	9.1 M	ABCWorldNews	1.6 M
CBSNews	8.9 M	BuzzFeedNews	1.3 M
NewYorker	8.8 M	HuffPostPol	1.3 M
AJEnglish	8.7 M	thedailybeast	1.3 M
NPR	8.7 M	Refinery29	1.2 M
SkyNews	8.4 M	ABCPolitics	1.1 M
people	7.6 M	chicagotribune	1.1 M
TMZ	7.4 M	NewsHour	1.1 M
ARYNEWSOFFICIAL	5.6 M	BBCPolitics	1 M
USATODAY	5 M	YahooNews	1 M

To create this list of all news media organization Twitter accounts with more than one million followers we aggregated several sources including the list of top 100 accounts from Social Blade used for the list of celebrities, https://memeburn.com/2010/09/the-100-most-influential-news-media-twitter-accounts/, https://viralpitch.co/topinfluencers/twitter/top-200-twitter-influencers/, and https://en.wikipedia.org/wiki/List_of_most-followed_Twitter_accounts. We retrieved the number of followers for each account on 6 December 2023. The final list of 76 accounts include all news media organizations that cover U.S. news stories, in English.

B.4.1 List of Political Identity Words from Rogers and Jones (2021)

activist, all lives matter, all_lives_matter, alllivesmatter, alt-right, anarchist, black lives matter, black_lives_matter, blacklivesmatter, blm, blue lives matter, blue_lives_matter, bluelivesmatter, communist, conservative, democrat, deplorable, feminist, gop, leftist, lgbtq, liberal, libertarian, maga, marxist, men's rights, mens rights, mens_rights, mensrights, nasty woman, nasty_woman, nastywoman, progressive, red pill, republican, socialist, the 99%, the_99%, the99%, woke

References

Dorottya Demszky, Nikhil Garg, Rob Voigt, James Zou, Jesse Shapiro, Matthew Gentzkow, and Dan Jurafsky. Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. In *Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics: Human language technologies, volume 1 (long and short papers)*, pages 2970–3005, 2019.

Benjamin Enke. Moral values and voting. *Journal of Political Economy*, 128(10):3679–3729, 2020. doi: 10.1086/708857.

Matthew Gentzkow, Jesse M Shapiro, and Matt Taddy. Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica*, 87(4):1307–1340, 2019. doi: 10.3982/ecta16566.

Jesse Graham, Jonathan Haidt, and Brian A Nosek. Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5):1029–1046, 2009. doi: 10.1037/a0015141.

Khairiyah Mohamed Ridhwan and Carol Anne Hargreaves. Leveraging Twitter data to understand public sentiment for the COVID-19 outbreak in Singapore. *International Journal of Information Management Data Insights*, 1(2):100021, 2021. doi: 10.1016/j.jjimei.2021.100021.

Nick Rogers and Jason J Jones. Using Twitter bios to measure changes in self-identity: Are Americans defining themselves more politically over time? *Journal of Social Computing*, 2(1):1–13, 2021. doi: 10.23919/jsc.2021.0002.

Samira Yousefinaghani, Rozita Dara, Samira Mubareka, Andrew Papadopoulos, and Shayan Sharif. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. *International Journal of Infectious Diseases*, 108:256–262, 2021. doi: 10.1016/j.ijid.2021.05.059.