

OATH-Frames: Characterizing Online Attitudes Towards Homelessness via LLM Assistants

Jaspreet Ranjit[♥] Brihi Joshi[♥] Rebecca Dorn[♥] Laura Petry[♦] Olga Koumoundouros[♦]

Jayne Bottarini[♥] Peichen Liu[♦] Eric Rice[♦] Swabha Swayamdipta[♥]

[♥]Thomas Lord Dept. of Computer Science, University of Southern California

[♦]Suzanne-Dwork School of Social Work, University of Southern California

jranjit@usc.edu

Abstract

Warning: This paper contains content that may be offensive or upsetting.

Homelessness in the U.S. is widespread; individual beliefs and attitudes towards homelessness—often expressed on social media—are complex and nuanced (e.g. critical as well as sympathetic). Such attitudes can be challenging to summarize at scale, obfuscating the broader public opinion which advocacy organizations use to guide public policy and reform efforts. Our work proposes an approach to enable a large-scale study on homelessness via two major contributions. First, with the help of domain experts in social work and their trainees, we characterize Online Attitudes towards Homelessness in nine hierarchical frames (**OATH-Frames**) on a collection of 4K social media posts. Further, in an effort to ease the annotation of these frames, we employ GPT-4 as an LLM assistant to the experts; GPT-4 + Expert annotation presents an attractive trade off owing to a $6.5 \times$ speedup in annotation time despite only incurring a 2 point F_1 difference in annotation performance. Our effort results in a collection of 8K social media posts labeled by domain and trained experts (with and without GPT-4 assistance). Second, using predicted OATH-Frames on a Flan-T5-Large model trained on our data, we perform a large-scale analysis on 2.4M posts on homelessness. We find that posts that contain mentions of west coast states express more harmful generalizations of people experiencing homelessness (PEH) compared to posts about east coast states. We also find marked differences in attitudes across vulnerable populations as they are compared to PEH as being either more- or less-deserving of aid.

1 Introduction

Despite being the largest economy on the planet, the United States struggles with a high level of homelessness—over 653,104 individuals experi-

enced homelessness in 2023.¹ The social, political and economic climate of the United States contributes to the complex state of homelessness (Robertson and Greenblatt, 1992) thus yielding nuanced attitudes toward PEH among broader society (Cronley, 2010; Moore et al., 1988; Narayan and Petesch, 2002).

Prior work from the fields of sociology and social work have studied discourse about homelessness grounded in ethnographic studies, qualitative surveys, and smaller scale content analyses (Kim et al., 2013; Tsai et al., 2019; Kingree and Daves, 1997). On the other hand, online social media data where individuals express their attitudes towards homelessness is available in much higher quantities. However, there exists little work to understand societal attitudes about homelessness *at a large scale*, as expressed in online discourse.

In this work, we address this issue by introducing **OATH-Frames**, a framing typology to characterize societal attitudes towards homelessness in the U.S., as expressed in social media posts on the X platform. Our frames are developed in collaboration with social work researchers with expertise in the domain of homelessness, and based on social science literature on framing theory (Entman, 1993) and grounded theory (Glaser and Strauss, 1967). Our OATH-Frames can be broadly categorized into three themes: **Critiques**, **Perceptions** and **Responses** to people experiencing homelessness (PEH), and further subcategorized into nine finer-grained Issue-specific frames; Table 1 summarizes our typology and corresponding definitions.

Using OATH-Frames, we manually annotate a corpus of 4.1K posts on homelessness from the platform X². We find that most posts receive multiple labels across the three broad themes, represent-

¹<https://www.huduser.gov/portal/sites/default/files/pdf/2023-AHAR-Part-1.pdf>

²Formerly, Twitter: <https://x.com>.

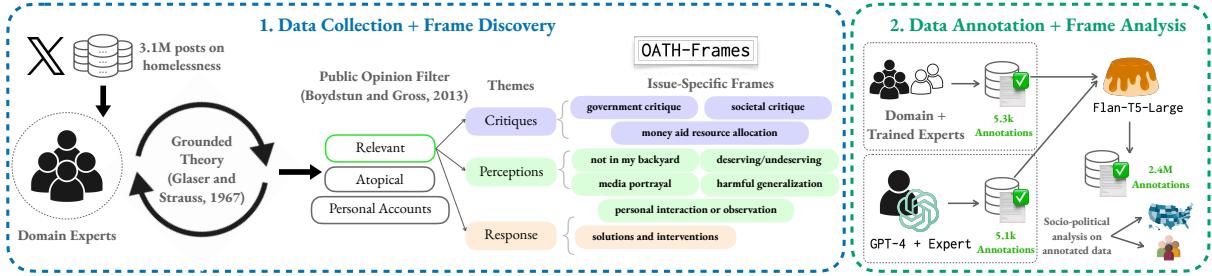


Figure 1: **OATH-Frames discovery, annotation and analysis:** Given a subset of our 3.1M posts on homelessness, domain experts apply grounded theory (Glaser and Strauss, 1967) to develop frames that describe Online Attitudes Towards Homelessness: OATH-Frames. We collect data annotations using domain and trained experts and GPT-4 + Expert to train a Flan-T5-Large to generate predictions on a set of 2.4M posts. During *Frame Analysis*, we use annotated posts to analyze variations in attitudes across social and political dimensions that affect public opinion towards homelessness.

ing complex and nuanced attitudes towards homelessness. The most salient theme in our annotated dataset is **Critiques** which contains posts that critique the government and normative structures in regards to policies and aid distribution. Overall, we find that the annotation of each X post takes our experts 3 minutes on average, highlighting the labor intensity of our task. To address this, we employ a GPT-4 annotation assistant building on the generative successes of large language models, to perform a first round of annotations, which are then verified by the experts. This GPT-4 + Expert approach results in a $6.5 \times$ speedup in annotation, while incurring only a 2 point F_1 reduction in annotation accuracy compared to the manual annotation—an attractive trade-off given the complexity of the multi-label annotation. Using this strategy, we further annotate 4K posts; overall our corpus is comprised of 8K expert annotated posts collected with or without GPT-4 assistance.

Finally, we use our collected data to fine-tune a Flan-T5-Large model that can predict OATH-Frames on a larger set of 2.4M X posts on homelessness to analyze these conversations at scale. We find large variations in (predicted) attitudes on homelessness across socio-political factors such as regionality, and amongst subpopulations of PEH and vulnerable populations (i.e. immigrants).

Public opinion serves as a primary driving force for advocacy groups and political constituents to garner more support for policy formulation and reform efforts (Burstein, 2020; Applebaum, 2001). We hope our contributions serve as a tool to provide fast and scalable summarization of public opinion for advocacy groups and political constituents in

tracking the most salient attitudes towards homelessness as expressed in online discourse. Our data and models will be publicly available³.

2 Pragmatic Frames for Attitudes towards PEH: OATH-Frames

We introduce a novel typology of pragmatic frames for capturing and characterizing societal attitudes towards homelessness and people experiencing homelessness (PEH). We apply concepts from framing theory (Entman, 1993), to identify how certain aspects in an issue’s portrayal are emphasized to influence broader public opinion. For instance, framing homelessness with substance use and crime can invoke feelings of fear and hostility amongst the public. This theory has formed the basis of analyzing discourse on immigration, (Mendelsohn et al., 2021), politics in media (Boydston and Gross, 2013), eating disorders (Arseniev-Koehler et al., 2016) and so on. Prior survey studies and longitudinal content analyses have provided valuable insights into how perceptions towards PEH have changed over time (Tsai et al., 2017; Tompsett et al., 2006). In addition, due to the many complex factors that contribute to homelessness, prior work has also surveyed the public on their perceived causes of homelessness (Lee et al., 1990; Dittmeier et al., 2018; Agans and Liu, 2015). In the context of homelessness discourse on social media, prior work has focused on the framing of stigma towards PEH (Kim et al., 2013) and surveys that study attitudes and perceptions towards PEH (Toro and McDonnell, 1992; Agans and Liu, 2015). In our work, we take inspiration from Kim et al. (2013) and their application of grounded theory

³<https://jr4fs.github.io/OATH-Frames/>

Issue-specific Frame	Definition	Example	Additional Labels
GovCrit. Government Critique	Criticism about government body, policies and laws including discussion of homelessness through the lens of political parties and values.	'these people are already housed possibly over crowded but at least housed .there are up to 130000 homeless living rough on our streets every night women and children in the back of vans and cars and the gov gives billions away to help third world countries.? time he looked here'	MoneyAid. , (Un)Deserv.
MoneyAid. Money Aid Re-source Allocation	Discussion of money, aid or resource disbursement on addressing the homelessness issue. Examples include the allocation of necessities such as essential items for emergency relief supplies, or government budgeting with respect to competing priorities.	'you propose atlanta/fulton county fund a homeless city. i'm asking what return are we getting for our tax dollars? police officers and firefighters provide a service which can justify us building a training for them. what justification do you have for your city for the homeless?'	GovCrit. , SolnInt. , (Un)Deserv. , MoneyAid.
SocCrit. Societal Critique	Criticism of social norms, systems and society at large in how homelessness is being addressed and perceived. Also includes pointing out hypocrisy and performative activism.	'i see it the most with addicts/homeless. the lack of connection w/a community is killing the sensitive souls of our world. we're hurting each other, but not on purpose. we've lost true community where no one is left alone/neglected. be mindful of the quiet ones; draw them close.'	Interact. , SolnInt.
SolnInt. Solutions and Interventions	Discussion of solutions, interventions, charitable acts and remedies to address the homelessness crisis.	'Ever think that instead of jail, A MENTALLY ILL HOMELESS PERSON needs actual help? Yes, I have compassion for his victims. But I also have compassion for this guy. He needs help. He doesn't need to be thrown in jail where he'll rot and probably be left to die. You pointed out'	HarmGen.
Interact. Personal Interaction	Anecdote describing a direct personal exchange with PEH.	'i'm legit sad af. i saw a homeless woman and her daughter. i really wanted to help but i had no cash. i hope they are still there when i circle back around.'	SolnInt.
MediaPort. Media Portrayal	Portrayal of (fictional or real) PEH as described in the media (e.g. discussing PEH in a TV show or in the news).	'is the national news reporting on the homeless on the streets and encampments as well as random crime exaggerated? beautiful city and location.'	NIMBY
(Un)Deserv. Deserving and Undeserving of Resources	Perpetuating a hierarchy of PEH with other marginalized communities or the use of harmful generalizations such as substance use and mental illness to justify that PEH that are more or less deserving of aid. Includes nationalistic rhetoric.	'these people are already housed possibly over crowded but at least housed .there are up to 130000 homeless living rough on our streets every night women and children in the back of vans and cars and the gov gives billions away to help third world countries.? time he looked here'	GovCrit. , MoneyAid.
HarmGen. Harmful Generalization	Blanket statements that ascribe an undesirable characteristic to PEH that include but are not limited to generalizing all PEH as having an unkempt appearance, or being violent, racist, thieves, or sexual predators.	'gna look go to work looking pretty tomorrow bc i know they've forgotten. this homeless person who's been turning up to the office for the last month is not me, idk her x'	N/A

Table 1: **OATH-Frames**: Definitions with examples for each of our 9 Issue-specific frames that describe attitudes towards homelessness, grouped across three overarching themes: (i) Critiques , (ii) Perceptions , and (iii) Responses .

(Glaser and Strauss, 1967) to characterize broader societal attitudes on homelessness at scale, going beyond stigma.

Following Boydston and Gross (2013), our framing process consists of two main parts: (i) data collection and frame discovery (§2.1) and (ii) data annotation and frame analysis (§2.2) as shown in Figure 1. Frame discovery uncovers salient themes in a text corpus on a topic, resulting in a framing typology; in the context of our study, this results in OATH-Frames (Frames for Online Attitudes Towards Homelessness). Data annotation and frame analysis entails annotating posts with the discovered frames and post-hoc analysis.

Given the diversity (e.g., socio-economic and political) of users on X (Hu et al., 2021), we use posts from this platform to study societal attitudes towards homelessness. We use a keyword search on the X API⁴ to randomly sample posts containing the seed token “homeless”; this results in a collec-

tion of 5M posts, which we further deduplicate and clean to retain 3.1M posts spanning from 2021 to 2023.

2.1 Discovery of OATH-Frames

The development of OATH-Frames consisted of an extensive iterative process that involved four domain experts. First, each domain expert independently annotated a random sample of the same 100 posts with their perceived labels for attitudes. This was followed by a consensus discussion which consisted of a review of all the posts grouping similar labels together and finally a consensus on the labels for the perceived attitudes in a post. This process was repeated on three different subsets of the posts until reaching thematic saturation at nine Issue-specific frames with 3 overarching themes. We provide details on our intermediate iterations of our Issue-specific frames in Figure 8 in Appendix D along with guidelines for annotating our posts in Appendix B.

Loosely following prior work (Baiocchi and Ar-guello, 2019; Gowen, 2010), we find three over-

⁴<https://developer.X.com/en/docs/X-api> accessed in February 2023

arching themes. First, we find **Critiques** of systems and institutions enabling homelessness and also the use of homelessness as a vehicle to critique others (e.g. actors, programs or spending). Second, we find posts reflecting different **Perceptions** of PEH, such as posts which attribute undesirable characteristics to PEH to make harmful generalizations. Third, we find many **Responses** to the issue, either posed as solutions, interventions, calls to action and charitable acts. Inspired by grounded theory (Glaser and Strauss, 1967) from social science, we surface nine **Issue-specific** frames under these overarching themes, detailed in **Table 1**. Issue-specific frames provide finer-grained characterizations of broader topics (e.g. Public Opinion from Boydston and Gross, 2013). This hierarchical categorization results in our OATH-Frames, as illustrated in **Figure 1**.

Additionally, we filter out ambiguous posts including those not discussing attitudes towards homelessness, or unrelated topics (e.g. political homelessness). Finally, we do not consider any posts discussing lived experiences of PEH from a first person point of view. To this end, we reference the Public Opinion frame to only retain posts that contain public opinions and attitudes towards homelessness (Boydston and Gross, 2013).

2.2 Annotating Posts with OATH-Frames

To perform data annotation and frame analysis (Boydston and Gross, 2013), we build a team of four domain experts and two trained experts. Amongst our team of six experts, we collect annotations on an Amazon Mechanical Turk Sandbox environment as shown in **Figure 16** in Appendix F. Each annotator first determines whether the post should be filtered out according to the Public Opinion filter. For relevant posts, we choose one or more of the Issue-specific frames to annotate the post. The trained experts are trained by the domain experts to do the annotation task. The training process consists of three main steps. First, the trainee familiarizes themselves with the definitions and examples of the frames as shown in **Table 1**. Second, the trainee performs a small exercise labeling 15 posts with OATH-Frames and works with the domain expert to clarify any ambiguities. Third, the trainee performs an independent annotation exercise on a set of 280 posts that are multiply annotated by the domain experts. If the trainee achieves moderate agreement with the do-

main experts, this concludes the training process. Otherwise, we repeat this process by going back to the first step.

Our team of domain and trained experts label 5.3K posts, of which 1.2K posts are filtered out leaving 4.1k posts that we consider in our analysis of expert annotated posts. In this set, 280 posts are multiply annotated across 6 annotators where we achieve an average inter-annotator Fleiss κ agreement of 0.47 across nine Issue-specific frames (Fleiss, 1971) with the first author.⁵ We detail our inter-annotator agreement for each annotator in **Figure 2** on this set of 280 multiply-annotated posts; we call this set the Agreement Test Set (280).

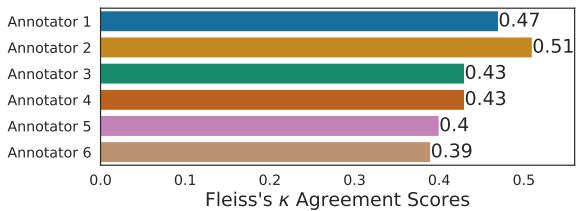


Figure 2: Inter-annotator Agreement Scores: We plot the average Fleiss' κ agreement scores for each domain and trained expert (Annotators 1...6) where each Annotator's Fleiss' κ is computed by comparing their annotations to every other annotator.

We find that each post may contain multiple Issue-specific frames, even across our three broad themes where frames co-occurring with each other can represent multiple attitudes in a single post. We include a plot of the co-occurrence of our frames in **Figure 11** in Appendix E and examples of our expert annotations in **Table 13** in Appendix F. **Figure 3** details the distribution of Issue-specific frames in our subsets of data obtained from different annotation strategies where we find that in our set of 4.1k Domain + Trained expert annotated posts, **GovCrit.**, **SolnInt.** and **HarmGen.** are the most frequently occurring Issue-specific frames and the most salient theme is **Critiques**.

Even in the team of trained and domain experts, collecting 4.1k annotated posts was extremely laborious and time consuming as our posts contain rhetoric which was challenging to process. Given that domain expertise is essential in our annotation process to ensure the quality of annotations for sensitive topics such as homelessness, we next

⁵Our agreement across 9 frames is ‘moderate’; prior work has reported similar agreement of 0.3-0.4 across 8 frames (Stede et al., 2023) in a multilabel classification task.

explore employing LLMs as annotation assistants to alleviate the pressures of manual annotation.

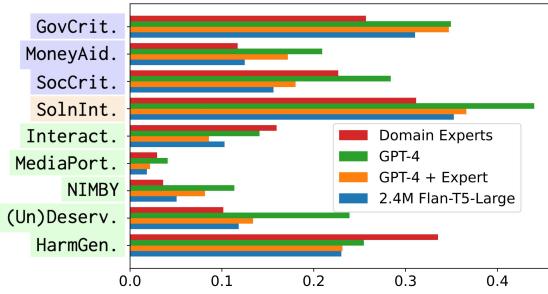


Figure 3: **Proportion of Issue-specific frames:** We show the proportion of Issue-specific frames in our data sources across different annotation methods (experts, GPT-4, GPT-4 + Expert, and Flan-T5-Large predictions) and observe that the distribution of frames across sources is similar. Details about our annotation sources are included in Table 5 in the Appendix A.

3 Expanding Data Collection via GPT-4 + Expert Annotations

LLM-assisted annotations are emerging as a promising application of LLMs, making annotation more time- and labor-efficient (Rytting et al., 2023; Pangakis et al., 2023; Li et al., 2023). Pangakis et al. (2023) propose a pipeline to use LLMs for annotating large amounts of data. They posit that LLM generated annotations should always be evaluated with a set of high quality expert annotations during prompt tuning since performance is highly contingent on the nature of the task. Furthermore, LLMs on their own may not always be able to reliably reason about socially complex tasks. More specifically, harmful rhetoric against PEH is masked by humor and sarcasm making the task more interpretive and challenging (Kiritchenko et al., 2023; Nejadgholi et al., 2022; Wiegand et al., 2019). In such cases, it's also important to validate LLM generated annotations post hoc with experts. We use GPT-4 as an assistant in the human annotation process: GPT-4 + Expert, to reduce the amount of time it takes for humans to complete annotations on their own. We evaluate the performance of GPT-4 and GPT-4 + Expert in Table 2 on our Agreement Test Set (280) as this set contains multiply-annotated posts, which allows us to compute performance within a margin of error across all 6 annotators. Due to compute budget restrictions, we do not generate GPT-4 annotations on our test set comprising of 1280 (singly) human-annotated posts and as a result, the metrics reported are on the multiply-

annotated Agreement Test Set (280). We include details of our dataset splits for Domain + Trained experts, GPT-4, and GPT-4 + Expert annotated posts in Table 5 in Appendix A.

3.1 Predicting OATH-Frames with GPT-4

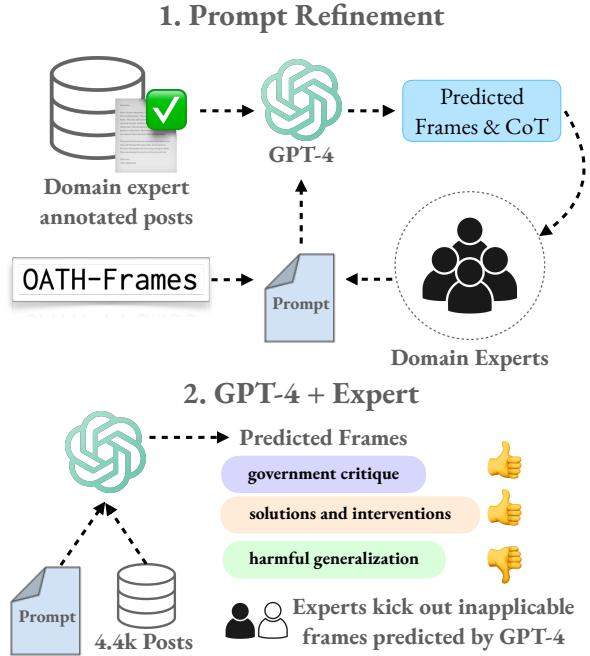


Figure 4: **GPT-4 + Expert Annotation Pipeline:** We prompt GPT-4 with OATH-Frames and their corresponding definitions where we use generated chains-of-thought to refine our prompt and validate against a set of expert annotated posts following prior work (Pangakis et al., 2023). Using our refined prompt, we generate predictions of OATH-Frames on a set of 4.1k posts and experts verify the predictions to kick out frames that do not belong.

We prompt GPT-4 in a sequential manner that emulates our framing typology. Our prompting pipeline consists of two phases. In the first phase, we apply the Public Opinion frame to filter out posts that do not contain a public attitude towards homelessness. We generate GPT-4 predictions on a set of 10k posts where filtering posts out results in a set of 7k posts. The second phase consists of our Issue-specific frames detection where we prompt GPT-4 to perform a multilabel classification task for each post. Due to compute restrictions, we take a subset of 4.4k posts out of the 7k posts to generate Issue-specific frame predictions. Examples of our prompt are provided in Table 10 in Appendix C. The cycle of prompt refinement involves clarifying definitions and ambiguities for the incorrect classifications and measuring perfor-

Annotation →	1st Author	GPT-4	GPT-4 + Expert	Flan-T5-Large	Flan-T5-Large
Test Set →	Agreement (280)			Test (1280)	
Issue-specific ↓	F ₁ Score (↑)				
GovCrit.	0.68 ± 0.04	0.70 ± 0.08	0.72 ± 0.07	0.65 ± 0.07	0.63
MoneyAid.	0.57 ± 0.17	0.58 ± 0.14	0.61 ± 0.15	0.59 ± 0.18	0.61
SocCrit.	0.37 ± 0.12	0.46 ± 0.06	0.50 ± 0.05	0.42 ± 0.09	0.49
SolnInt.	0.67 ± 0.06	0.62 ± 0.06	0.65 ± 0.06	0.68 ± 0.03	0.57
Interact.	0.36 ± 0.08	0.45 ± 0.07	0.52 ± 0.07	0.54 ± 0.11	0.7
MediaPort.	0.70 ± 0.07	0.50 ± 0.08	0.63 ± 0.11	0.69 ± 0.10	0.54
NIMBY	0.28 ± 0.20	0.30 ± 0.16	0.39 ± 0.21	0.34 ± 0.26	0.26
(Un)Deserv.	0.44 ± 0.07	0.46 ± 0.06	0.63 ± 0.08	0.67 ± 0.01	0.61
HarmGen.	0.60 ± 0.04	0.49 ± 0.06	0.51 ± 0.06	0.55 ± 0.05	0.58
Aggregated	0.54 ± 0.01	0.46 ± 0.04	0.52 ± 0.04	0.51 ± 0.03	0.5

Table 2: **F₁ scores for each Issue-specific frame, and macro F₁ scores across all 9 Issue-specific frames (Aggregated):** The standard deviation represents the difference in F₁ scores obtained from comparing across different annotator’s labels. 1st Author refers to average F₁ scores computed across 5 annotators as compared with the 1st author. We find that GPT-4 + Expert and Flan-T5-Large consistently outperform GPT-4 where GPT-4 + Expert incurs a 2 point Macro F₁ difference over the aggregated frames in comparison to domain experts (1st Author). We also include Flan-T5-Large F₁ on the larger test set of 1280 singly annotated instances. We include the precision, recall and F₁ score for our Issue-specific frames on the Agreement Test Set (280) in Table 8 and Table 9 and results on the larger test set (1280) for our ablations for Flan-T5 in Table 4 in Appendix D.

mance against our set of expert annotations until we reach a desirable performance (Pangakis et al., 2023). We use chain-of-thought (CoT) prompting (Wei et al., 2022) to better understand the kinds of errors GPT-4 makes and to clarify the definitions of the frames in our prompt. More precisely, GPT-4 was more likely to make errors if the generated chains-of-thought were not aligned with the frame definition. The domain experts examined the misclassifications and used GPT-4’s misinterpretations to clarify the definitions in the prompt until desirable performance was reached.

We find that GPT-4 achieves a precision of 0.39 ± 0.05 and recall of 0.64 ± 0.05 across all nine Issue-specific frames on the Agreement Test Set (280) showing that it may have reasonable coverage of the predicted labels but is also producing false positives due to the inherent ambiguity in the posts that may be easier for a domain expert to resolve.

3.2 GPT-4 as an Annotation Assistant

Although LLMs serve as a useful tool to efficiently generate annotations, the inherent subjectivity of our task poses a challenge to using GPT-4 as an annotator on its own. In order to leverage GPT-4’s capabilities as an efficient annotator while ensuring high quality annotations, we propose a GPT-4 + Expert approach as outlined in Figure 4 to verify

and remove incorrectly classified examples from GPT-4 generations. Given the 7k posts annotated by GPT-4, we take the subset of 4.4k posts that have been annotated with Issue-specific frames and a team of experts validates the predictions on this set. Upon validation, we find that experts filter out an additional 300 posts according to the Public Opinion filter, which results in a set of 4.1k GPT-4 + Expert validated posts. During our annotations, we record the time taken by the annotators to annotate each post, both individually and with the GPT-4 + Expert approach. We note that by themselves, Domain + Trained experts take 187.49 (± 30.71) seconds, whereas with GPT-4 + Expert, they take 28.80 (± 8.55) seconds which is a $6.5\times$ boost in annotation speed and results in a 6 point increase in macro F₁ score (aggregated over nine Issue-specific frames) over GPT-4 on its own as shown in Table 2.

GPT as an Annotator: Error Analysis Our GPT-4 + Expert experiments are designed such that experts kick out labels that GPT-4 misclassified. By only removing falsely positive frame predictions, we’re expecting a boost in precision in GPT-4 + Expert verified posts compared to GPT-4 generations as observed in Table 8. We also see a slight drop in recall due to the 300 posts that get additionally filtered out by experts. We include Table 11 in the Appendix which outlines

frames that have the highest false discovery rate where inherent subjectivity and interpretation of the post warrants misclassifications. For example, the shorter length of posts and ambiguous target group of the post leaves more room for interpretation. The **Interact.** frame warrants such ambiguity in interpretation where GPT-4 makes generalizations beyond the content provided in the post. In these cases, it's important to have a human resolve the ambiguities. Furthermore, GPT-4 often confuses the **SocCrit.** and **GovCrit.** frames due to the ambiguity in the target group of the post. When society or the government are not explicitly mentioned in the post, domain experts are able to use the context of the post to resolve the target group. For example, '@mention it appears the problem in el paso with illegal aliens were solved overnight, so that you could have a photo op without anyone seeing hundreds of homeless illegals on the streets.' In this post, the topics of immigration policy, displacement sweeps and photo-ops are often discussed in the context of the government and directed at politicians. In these cases, GPT-4 is not able to recognize this post as **GovCrit.** and thus, it is important to resolve such ambiguities with post hoc human validation. We also find that GPT-4 has the highest precision for frames where the CoT reasoning is most closely aligned with our frame definitions for the **HarmGen.** and **GovCrit.** frames. The generated chains-of-thought were helpful in clarifying misinterpretations of our frame definitions and could be useful for practitioners who use GPT-4 as an assistant for annotating socially sensitive data.

4 Large-scale Frame Analysis

4.1 Training Frame Prediction Models

Combining our 5.3k expert-annotated posts with 4.1k obtained using GPT-4 + Expert (§3), and an additional 1k posts that were filtered out by the Public Opinion frame, we obtain a total of 10.4k annotated posts. We split these into a train-val-test split containing 8217, 913 and 1280 posts respectively and train a Flan-T5-Large model to filter out posts according to the Public Opinion frame, and predict our nine Issue-specific frames. Further details of the dataset split, model training and ablations with Flan-T5 sizes are provided in Table 6 and Table 7 in Appendix A. We test three variations of Flan-T5 and find that Flan-T5-Large performs the best compared to Flan-T5-Base and

Flan-T5-Small showing that larger models are able to achieve a higher performance on our task. Results in Table 2 show that Flan-T5-Large performs better than GPT-4 with a Macro F₁ of 0.51 ± 0.05, and is at par with the GPT-4 + Expert setting on our Agreement Test Set (280). We also find that Flan-T5-Large obtains a Macro F₁ score of 0.5 on a larger test set of 1280 posts. Using our trained Flan-T5-Large model, we then generate Issue-specific frame predictions for our corpus of 3.1M posts. After filtering posts out, we obtain a total of 2.4M posts annotated with our Issue-specific frames which we use for our socio-political analysis.

4.2 Analysis of Socio-political trends using OATH-Frames.

Public opinion serves as a key determinant in informing public policy (Burstein, 2020). Synthesizing public opinion via online discourse can serve as a useful tool for advocacy groups and policymakers to inform more strategic efforts for bringing homeless services and more equitable housing policies to local communities. Our framing typology provides an avenue to provide such insights in how attitudes towards homelessness vary across socio-political dimensions. We demonstrate the utility of our frames in characterizing differences in attitudes on social media as they pertain to different geographical locations and sub-populations of PEH (i.e. veterans) and other vulnerable populations (i.e. immigrants). We perform our analysis on the set of 2.4M posts that are retained after filtering out posts according to Public Opinion filter. Details of our annotation strategy and dataset splits, along with the number of posts that have been filtered out are included in Table 5 in Appendix A.

Attitudes across State Mentions We investigate the saliency of our Issue-specific frames in posts that mention states in the west and the east coast of the United States.⁶ Discourse on homelessness is closely tied to the region's ideological standing and the prevalence of PEH in each region (Tsai et al., 2019). California has the highest rate of overall homelessness where 68% are unsheltered in comparison to New York which has the second highest rate of overall homelessness but 4.9% of the population is unsheltered (Sousa

⁶Our analysis includes the following states for the west: California, Washington, Oregon and Nevada, and for the east: New York, Massachusetts, District of Columbia across 147k posts that contain these mentions in the set of 3.1M posts.

et al., 2023). We hypothesize that greater visibility of PEH in the west coast due to a larger unsheltered population results in public attitudes centered around **Perceptions** such as **HarmGen.**. Comparatively, New York specifically experienced an influx of asylum seekers and immigrants in 2023 (Sousa et al., 2023) and thus we hypothesize that posts about east coast states are more likely to discuss competition for jobs and resources between PEH and other vulnerable populations. We segment our dataset by posts mentioning east and west coast states using NER⁷. We find that **HarmGen.** is the most salient Issue-specific frame in posts mentioning west coast states and **(Un)Deserv.** is the most significant frame in the posts mentioning east coast states as shown in Figure 5. This trend reflects our hypothesis that due to a larger visibility of PEH in west coast states, people may express more stereotypes and harmful rhetoric against PEH, whereas posts about east coast states reflect discourse that compares PEH with vulnerable populations. For example, the post: '@mention so true! there's more and more places now allowing homeless to shit on the sidewalks and hand out drug needles that will end up in the streets and in playgrounds. california is so great.' discusses **HarmGen.** of PEH with respect to substance use in the state of California. Meanwhile, the post '@mention this is against your interest and all americans. in nyc, homeless americans are on the street, while illegal immigrants are in lavish hotels being served gourmet meals! how ridiculously unfair, all because democrats want voters who will support their absolutely crazy policies!' discusses immigrants being compared to PEH for accessibility to shelter. We provide a breakdown of attitudes per state for the most frequently occurring states in our data in the Appendix in Figure 15.

California also contains a large number of encampments, and a scarcity of affordable housing resulting in a greater prevalence of public attitudes towards land use regulations (Sheeley et al., 2021). To this end, we explore the proportion of state representation for the Issue-specific frame of **NIMBY** where we find that California is represented proportionally higher than other states as illustrated in Figure 6.

Differences in Attitudes Across Vulnerable Populations and PEH

In 1834, the English Poor

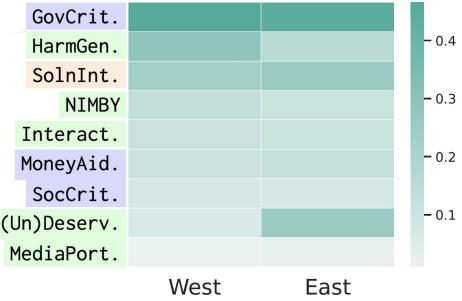


Figure 5: Proportion of Issue-specific frames per Region in the U.S.: (Un)Deserv. and HarmGen. are more significantly present in posts mentioning east and west coast states respectively ($p < 0.0001$) where we hypothesize that differing Perceptions may be due to PEH visibility in those regions.

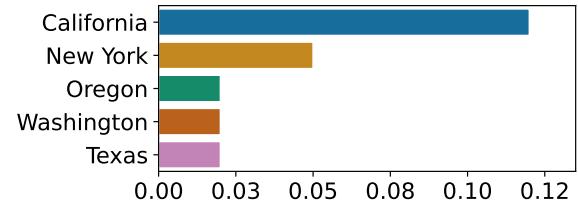


Figure 6: State-wise proportion of posts tagged with NIMBY frame: We plot the top 5 states mentioned in our corpus (in 147K of 2.4M posts). Among these, posts mentioning California have the highest occurrence of NIMBY frames.

Law distinguished between the "deserving" and "undeserving" poor where those who are perceived as victims of their circumstances are considered as more deserving of aid and resources compared to those who are perceived as lazy and a burden to society (Solas, 2018). We observe this attitude of (Un)Deserv. in the context of homelessness where public attitudes reflect a juxtaposition between vulnerable populations (i.e. immigrants, asylum seekers, refugees, Ukrainians) and PEH for funding and aid. During frame discovery, we observe that (Un)Deserv. attitudes aggravate intersectional biases by comparing populations that belong to multiple disadvantaged demographic identities (Kiritchenko et al., 2023; Hoffmann, 2019). For example, we observe that veterans who are also experiencing homelessness are used as a comparison against other vulnerable populations such as Ukrainians and immigrants. Such attitudes of (Un)Deserv. put both vulnerable populations and PEH at risk for being further ostracized from garnering public support for receiving aid and resources (Lalor et al., 2022). We want to fur-

⁷hf-reddit-ner-place-names: https://huggingface.co/cjber/reddit-ner-place_names.

PEH	Vulnerable Population	Significant Issue-specific Frames	Example
'Homeless' Veteran	Ukrainian	GovCrit., MoneyAid.	@mention they have been given not 1 billion dollars, not 80 billion dollars (russian military budgets, but 180 billion dollars. do you really think that money is being used to help the ukrainian people-or is it something else? go help a homeless vet today! Post labeled with: MoneyAid., GovCrit., (Un)Deserv.
	Immigrant	NIMBY	@mention oh please they are dropping off bus loads of immigrants to hotels 6 blocks from my house that we the tax payers get to pay for. we have homeless vets what about them !! wake up Post labeled with: NIMBY, GovCrit., (Un)Deserv.
'Homeless' People	Ukrainian	GovCrit., MoneyAid.	ukrainian refugees can get ssi, housing benefits, and free health care. meanwhile, we have a lot of homeless people who have no access to such benefits - we have the highest number of the unemployed black population. Post labeled with: MoneyAid., GovCrit., (Un)Deserv.
	Immigrant	Interact., NIMBY, HarmGen.	@mention new york city's pulling the rug out from underneath the illegal immigrants kicking them out of the \$700 a day hotel rooms. was it the bedbugs or the thousands of empty bottles of beer and liquor. we used to have a problem with homeless people now we have a catastrophe of homeless Post labeled with: NIMBY, GovCrit., (Un)Deserv.
'Homeless' American	Ukrainian	GovCrit., MoneyAid.	@mention grifter journalist got to grift . how about reporting on how close we are from being in a nuclearwar with russia. or how corrupted our politicians have become. write about how homeless americans live across this country and the us thinks sending money to ukraine is saving anyone Post labeled with: MoneyAid., GovCrit., (Un)Deserv.
	Immigrant	SocCrit., SolnInt., Interact., MediaPort., NIMBY, HarmGen.	@mention help homeless americans that are being discriminated against by the huge number of foreign aliens in america or give money to another county. i say americans are f***ed until we start saying no to everything immigrant. more foreign aliens = more poor americans = communism grows Post labeled with: SocCrit., SolnInt., (Un)Deserv.

Table 3: **Statistically significant frames for posts containing most frequent (PEH, Vulnerable Population) co-occurrence pairs.** For each subset of posts containing a (PEH, Vulnerable Population) co-occurrence pair, we find the most statistically significant frames ($p < 0.0001$). We observe that posts containing comparisons between PEH and Ukrainians elicit more Critiques whereas posts comparing PEH to immigrants elicit more NIMBY and HarmGen. attitudes. We provide a complete set of our results in Table 12 in the Appendix.

ther investigate the finer-grained differences in attitudes as characterized by OATH-Frames across the vulnerable populations represented in our subset of posts labeled with (Un)Deserv. (Rosenheck et al., 1999; Chishti and Gelatt, 2022).

To investigate the representation of vulnerable populations and PEH in our dataset, we search for the most frequently occurring bigrams in our subset of annotated 2.4M posts that are labeled with (Un)Deserv.. Amongst the most frequently bigrams, we identify the following subpopulations of PEH and vulnerable populations represented in our data: *homeless veteran*, *homeless people*, *homeless immigrant*, *homeless refugee*, *homeless american*, *asylum seeker*, *immigrant*, *refugee*, *illegal alien*, *ukrainian*. We extract a subset of 153k posts that contain mentions of these vulnerable populations and PEH. By definition, the (Un)Deserv. attitude compares PEH to other vulnerable populations so we make the distinction between vulnerable populations: *asylum seeker*, *immigrant*, *refugee*, *illegal alien*, *ukrainian* and PEH: *homeless veteran*, *homeless people*, *homeless immigrant*, *home-*

less refugee, *homeless american*⁸. Posts labeled with (Un)Deserv. contain mentions of both vulnerable populations and PEH. To investigate the frequency of these co-occurrences, we plot the proportion of posts containing co-occurrences of each vulnerable population and subpopulation of PEH in Figure 7. We find that the following six pairs co-occur most frequently: ('Homeless Veteran', 'Ukrainian'), ('Homeless Veteran', 'Immigrant'), ('Homeless People', 'Ukrainian'), ('Homeless People', 'Immigrant'), ('Homeless American', 'Ukrainian'), ('Homeless American', 'Immigrant'). For each co-occurring (PEH, Vulnerable Population) pair, we examine the most statistically significant attitudes as detailed in Table 3. We find marked differences in attitudes when PEH are compared to Ukrainians versus when they are compared to immigrants. Public attitudes in posts that

⁸There exists a lot of variation in ways that PEH are referred to (i.e. hobo, vagrant) and there are many subpopulations of PEH (i.e. youth, people of color, native/indigenous) that are not captured here. We choose to explore those subpopulations of PEH and vulnerable populations that are most salient in our data.

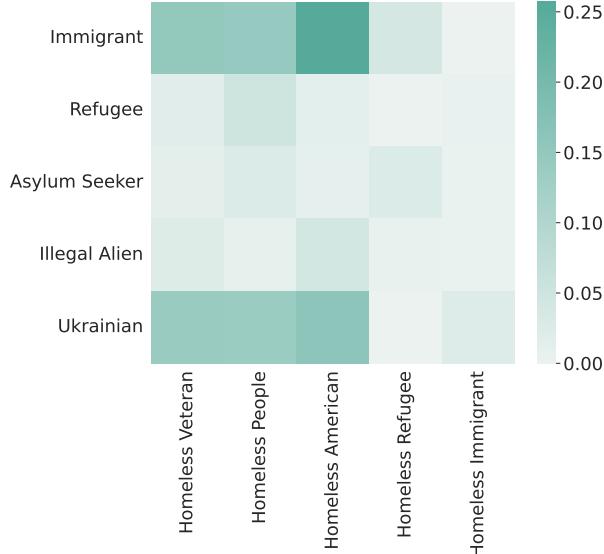


Figure 7: Co-occurrence of PEH and vulnerable populations: Proportion of posts containing mentions of both PEH ('homeless veteran', 'homeless people', 'homeless american', 'homeless refugee', 'homeless immigrant') and vulnerable populations ('immigrant', 'refugee', 'asylum seeker', 'illegal alien', 'Ukrainian').

compare PEH with Ukrainians are more directed in the form of critique against the government (`GovCrit.`) for foreign aid (`MoneyAid.`). For example: '@mention they have been given not 1 billion dollars, not 80 billion dollars (russian military budgets, but 180 billion dollars. do you really think that money is being used to help the ukrainian people-or is it something else? go help a homeless vet today'. Additionally, we observe that attitudes in posts comparing PEH to immigrants elicit more nationalistic sentiment and are more oriented around relocation (`NIMBY`) and can contain dehumanizing and harmful rhetoric (i.e. '@mention help homeless americans that are being discriminated against by the huge number of foreign aliens in america or give money to another county. i say americans are f***ed until we start saying no to everything immigrant. more foreign aliens = more poor americans = communism grows'). We provide a complete set of results comparing differences in attitudes across all co-occurrences in the Appendix in Table 12.

Characterizing differences in attitudes can help shed light on the ways in which marginalized groups are compared to each other. We show that OATH-Frames can characterize these finer grained differences amongst vulnerable populations as they are compared to PEH. Characterizing this kind

of discourse has important implications for understanding how homelessness is conflated with additional issues such as immigration giving us more insights into how public opinion is shaped by geopolitical issues and political factors. For example, substance use and mental illness have been identified as conflating factors that may influence the public's opinion about PEH, where such issues are often used as more socially acceptable registers to express `Critiques` against PEH and make harmful generalizations (Baiocchi and Arguello, 2019; Goffman, 1991). Identifying such factors using OATH-Frames can further help us understand the unique ways in which vulnerable populations experience subtle stereotyping and harms online and identify where existing sentiment and toxicity classifiers may fall short. We plot the distribution of the Perspective API toxicity score⁹ in Figure 13 and the probability of the negative sentiment score in Figure 14 for the twitter-XLM-roBERTa-base model (Rosenthal et al., 2017) for sentiment analysis over our set of 4.1k human-annotated posts. For the posts labeled with `HarmGen.`, we expect to see a left skew in the toxicity score distribution but find that a majority of the posts are labeled with a toxicity of less than 0.5 indicating that the classifier may be missing out instances that are considered harmful against PEH. We find the following examples are tagged with a toxicity score of less than 0.5: 'some people wear clothes at home i wear for an outing and here i dress up like a homeless person inside my home whos a drug addict and has been sleeping since forever', and 'i'm more concerned with clearing out the homeless people with committing genocide. prosecute them for possession and offer treatment in lieu of jail. any violation is violation of parole. guards are graduates of the program who are also on probation. any violation means prison.' where both examples contain instances of harmful stereotypes and violences against PEH. In addition, we find that the sentiment classifier does not distinguish `HarmGen.` from posts that do not contain `HarmGen.` where the distributions are close to overlapping. There exist limited data resources that characterize the unique and complex ways in which marginalized groups experience harms online (Kiritchenko et al., 2023; Hutchinson et al., 2020). OATH-Frames provides a bespoke framework to study the particular ways in which PEH

⁹<https://perspectiveapi.com/>

may experience harms and subtle stereotyping and we hope our framework can serve as a resource to improve existing classifiers in detecting such harms.

5 Related Work

Characterizing Attitudes Towards Homelessness Prior work characterizes public opinion and attitudes towards homelessness through ethnographic studies, surveys, content analyses and psychometric analyses on small samples of human participants (Gowan, 2010; Barnett et al., 1997; Toro and McDonnell, 1992; Agans and Liu, 2015; Kingree and Daves, 1997). Prior studies have found that diversity in public attitudes towards homelessness are heavily influenced by personal experiences and socio-demographic backgrounds. More specifically, Tsai et al. (2019) conducted a survey of 541 U.S. adults on public attitudes about homelessness with the goal of determining socio-demographic characteristics of survey respondents that are associated with particular attitudes. They find that a majority of the participants who identified as female, lower income, Democrat and were personally exposed to homelessness expressed attitudes of compassion towards PEH and believed that the federal government should take more constructive measures in funding and reform policies for PEH. Similarly, Smith et al. (2023) and Mullenbach et al. (2023) study people’s prior personal experiences with homelessness and its effect on attitudes and perceptions towards PEH. Prior studies also analyze how sociodemographic backgrounds of participants influence an individual’s belief behind the cause of homelessness through surveys (Lee et al., 1990; Dittmeier et al., 2018; Agans and Liu, 2015; Hocking and Lawrence, 2000). Such studies on attitudes and perceptions towards PEH play an important role in influencing public policies that address homelessness and restrictive housing policies (Mullenbach et al., 2023; Link et al., 1995). Public opinion is heavily influenced by the social and political climate. To this end, prior work have also looked at changing attitudes towards homelessness from 1990 to 2016 (Tsai et al., 2017) and 1993 to 2001 (Tompsett et al., 2006) by comparing survey responses conducted in different years and they found increasingly more liberal and compassionate responses towards homelessness. Beyond analyzing survey responses, Baiocchi and Arguello (2019) identifies a basic topology from ethnographic stud-

ies to identify key narratives in discourse about homelessness: sin-talk, sick-talk, and system-talk which characterize the causes of homelessness enlisting criminal behavior and mental illness as primary contributing factors. Although such studies provide invaluable insights into public opinion, it is extremely challenging to collect and analyze survey responses at a larger scale. Furthermore, prior work has also identified shortcomings in existing datasets and toxicity classifiers in recognizing degrading rhetoric against poor and disadvantaged populations. This is because harmful language against marginalized groups is often masked by humor and sarcasm and includes more subtle accounts of stereotyping that may be specific to a particular vulnerable population and thus harder to recognize (Kiritchenko et al., 2023; Curto et al., 2023). As a result, there exist few specialized datasets to study language targeted specifically towards PEH (Kiritchenko et al., 2023). To this end, we plan to expand on these studies by performing large-scale analyses on social media posts annotated with OATH-Frames and providing a data resource that analyzes complex, and at times, harmful attitudes against PEH.

Quantifying Framing in Social Media Discourse. Framing (Entman, 1993) offers one mechanism to understand the communicative power of language by characterizing what people find most important in a text. Framing has been used to analyze the politics of homelessness and crime in New York City to characterize responses to homelessness (Upston, 2016). Furthermore, framing has also been used to study discourse on social media for topics such as immigration and politics (Mendelsohn et al., 2021; Boydston and Gross, 2013). Our line of work is most similar to prior work that studies stigma and stereotypes against PEH in the media (Kim et al., 2013; Link et al., 1991; Leibowitz and Krueger, 2005; Belcher and DeForge, 2012). More specifically, Kim et al. (2013) characterize stigma against PEH using social media posts by manually annotating 6.4k posts and defining a set of codes that characterize different stigmatizing traits attributed to PEH. We build on prior work that use framing theory to surface Issue-specific frames that characterize attitudes towards homelessness on social media and apply OATH-Frames to study attitudes at scale.

LLMs as an Annotation Assistant. Collecting high quality annotations for socially complex, and

sensitive tasks is not only a laborious and time consuming process, but also mentally taxing. Prior work have explored using LLMs as an assistant in the annotation process to resolve ambiguities and reduce the burden off of human annotators (Li et al., 2023; Rytting et al., 2023). However, Pangakis et al. (2023) test the performance of GPT-4 across 27 annotation tasks and find that LLM performance varies widely across tasks and is largely dependent on the difficulty and subjectivity of the annotation task. They've suggested that LLMs should be used as an annotator in the loop with human validation. For socially complex tasks such as homelessness, we use GPT-4 as an assistant in the annotation process while validating its predictions with human experts to leverage both the efficiency while also maintaining of quality of LLM generated annotations. It's important for us to validate GPT-4 predictions as LLMs have been shown to misrepresent identity groups which can cause downstream harms when reasoning about problems with sensitive social implications such as homelessness (Wang et al., 2024).

6 Conclusion and Future Work

We introduce a framing typology: OATH-Frames characterizing complex attitudes on social media towards people experiencing homelessness (PEH) with three overarching themes: Critiques, Perceptions, and Responses, and nine Issue-specific frames. Our annotations are carried out with a combination of domain experts and a GPT-4 assistant, resulting in a corpus of 8K posts from X. We use models trained on this data to predict OATH-Frames on a large scale collection of posts on homelessness. Our analysis reveals several trends in attitudes towards homelessness, with variations across regionality and PEH as they are compared to vulnerable populations. We hope to expand upon OATH-Frames in future works to capture a greater degree of contextual information including correlated factors such as substance use and mental illness, and the target group and perceived tone of the post. Public opinion is largely influenced by how social, political and economic factors contribute to the causes of homelessness. Characterizing such factors may help better explain the statistical prevalence of our frames in reaction to news cycles, and geopolitical events that spark reactive and inflammatory comments on social media. We include examples of these additive labels in

Figure 10 in the Appendix D. OATH-Frames serves as a tool for social work practitioners in analyzing large text corpora for social trends and attitudes at a larger scale. Furthermore, our GPT-4 + Expert annotation scheme serves as a promising use case of human validated LLM annotations for practitioners who are otherwise bottlenecked by laborious and challenging manual data annotation tasks. We expect our framework to be the first step towards characterizing, summarizing and understanding complex societal attitudes on intricate social issues *at scale*.

Limitations and Ethical Considerations

United States Focus and Data Collection

Most of our posts are from users with a Western background and as a result, not necessarily representative of attitudes towards homelessness outside the U.S. We recognize that this may not be a representative sample of all attitudes towards homelessness and this is an important consideration to take into account when reasoning about our results to avoid making generalizations about homelessness. Furthermore, The demographic representation on X may be biased to a certain socio-economic status, containing mostly western views. Our collection of X data was restricted to using a keyword match with ‘homeless’ as a seed word but there are many other words and variations that are used to reference PEH that are perhaps more subtle that were not captured in our dataset. Initially, we started with a much larger set of seed words including: “Homeless, Homelessness, Unhoused, Unsheltered, Encampment, Tent, Houseless, Living on the streets, Hobo, Doss, Bag Lady, Migrant, Couchsurf, Housing Benefit, Street People, soup kitchen, roofless, vagabond, down-and-out, destitute”, but we found that this returned a lot of content not related to homelessness, however, we leave it up to future work to capture variations in references to PEH.

Privacy

Due to the diversity of our dataset and lexical diversity of social media posts, we anonymized user handles, and did not obtain or use any priors about the user from their profile or follower network. We only rely on the content of the post for our study. Our dataset includes post IDs and no additional identifying characteristics as to protect the privacy of the users. We don't have the demographic back-

grounds of X users to respect anonymity, however, this makes it challenging to better understand why some attitudes are more salient than others since demographic background has been shown to be correlated with divergent perspectives on homelessness (Tsai et al., 2019).

Subjectivity of Attitudes

Due to the nuances and subjectivity of sensitive social issues such as homelessness, we also understand that this issue is framed very differently depending on regionality, ideology, political and economic climate among other socio-political factors that might not be represented in our dataset. The use of LLMs can further confound these variables and also inject societal biases in reasoning about an already marginalized population. Similarly, our framing typology may have introduced our own prior biases in reasoning about these posts that may not be in direct alignment with the general public's attitudes towards homelessness. However, we believe this diversity in perspectives is useful in shedding light on the various attitudes towards homelessness on social media.

It's also important to consider additional social, political and economic factors that can point to why some attitudes are more salient than others. For example, substance use and mental illness have been identified as conflating factors that may influence the public's opinion about PEH, where such issues are often used as more socially acceptable registers to critique PEH and make harmful generalizations (Baiocchi and Arguello, 2019). OATH-Frames can serve as the foundation for identifying such factors which can further corroborate the statistical prevalence of certain attitudes over others. Furthermore, taking into consideration the target group of the post may provide more finer grained signals as to who the attitude is directed at and help better classify harmful rhetoric towards PEH (Kiritchenko et al., 2023). Initially, we considered the characterization of these additional factors which are shown in Figure 10, but it's extremely difficult and laborious to label posts at scale with more than 50+ frames and labels, as a result, we leave it up to future work to characterize additional social and political factors that can provide important and necessary context in elaborating the statistical prevalence of the frames in our dataset.

Bad Actors

Although our data set was not specifically targeted at collecting harmful rhetoric against PEH, we find there are many instances of discriminatory remarks and prejudice against marginalized populations. Although we recognize that the characterization of these frames can further exemplify and reinforce harmful rhetoric against PEH and there exists a risk of bad actors to further perpetuate stigmas against PEH, we take the stance of supporting PEH by further tracking their portrayal on social media at scale to benefit advocacy groups and political constituents to better inform their efforts and policies in support of alleviating the suffering of PEH.

Positionality Statement

People create social media posts often quickly using highly culturally specific and contextual language. Interpretation of language meaning is not neutral. Words and meanings vary significantly by culture, informed by age, gender, sexual orientation, race, geography, and ethnicity. These understandings and linguistic interpretations are not fixed and often change over time and context. Each annotator brings their own identity and priors coupled with professional knowledge into the language interpretation process which forms our positionality as individual researchers. Our task could benefit from additional multiply annotated posts and a greater diversity in the annotator pool. Our annotation team consisted of domain and trained experts from social work and computer science from a mix of backgrounds including 4 from the U.S., 1 from India, and 1 from China. All annotators have different engagement levels with the community and advocacy efforts in homelessness but all take the stance of advocating for PEH. Our stance and backgrounds may have contributed to certain biases in our framework that may not represent the full range of divergent perspectives on the topic.

Acknowledgements

We would like to thank Julia Mendelsohn and Vinodkumar Prabhakaran for their valuable insights on our framing typology and annotation pipeline. We would also like to thank the USC NLP faculty and students for their continuous feedback in the formulation of our project and the anonymous reviewers in the ARR OpenReview cycle for their helpful suggestions. This research was supported

by an award from the Allen Institute for AI and from Intel.

References

- Jihan Abdalla. 2022. Why republican governors are relocating migrants across us.
- P. Robert Agans and Guangya Liu. 2015. Public attitudes toward the homeless. *Global Journal of Science Frontier Research*, 15(E3):1–8.
- Lauren D Applebaum. 2001. The influence of perceived deservingness on policy decisions regarding aid to the poor. *Political psychology*, 22(3):419–442.
- Alina Arseniev-Koehler, Hedwig Lee, Tyler H. McCormick, and Megan Andreas Moreno. 2016. #proana: Pro-eating disorder socialization on twitter. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine*, 58 6:659–64.
- Arturo Baiocchi and Tyler Arguello. 2019. *Street Talk: Homeless Discourses and the Politics of Service Provision*, pages 123–147. Springer US.
- M.A. Barnett, S.W. Quackenbush, and L.K. Pierce. 1997. Perceptions of and reactions to the homeless: A survey of fourth-grade, high school, and college students in a small midwestern community. *Journal of Soc Distress Homeless*, 6:283–302.
- John R Belcher and Bruce R DeForge. 2012. Social stigma and homelessness: The limits of social change. *Journal of Human Behavior in the Social Environment*, 22(8):929–946.
- Amber E. Boydston and Justin H. Gross. 2013. Identifying media frames and frame dynamics within and across policy issues.
- Paul Burstein. 2020. The determinants of public policy: What matters and how much. *Policy Studies Journal*, 48(1):87–110.
- Muzaffar Chishti and Julia Gelatt. 2022. Busing and flights of migrants by gop governors mark a new twist in state intervention on immigration.
- Courtney Cronley. 2010. Unraveling the social construction of homelessness. *Journal of Human Behavior in the Social Environment*, 20(2):319–333.
- Georgina Curto, Svetlana Kiritchenko, Isar Nejadgholi, and Kathleen C Fraser. 2023. The crime of being poor. *arXiv preprint arXiv:2303.14128*.
- Kerry Dittmeier, Sharon H Thompson, Emma Kroger, and Nancy C Phillips. 2018. Perceptions of homelessness: Do generational age groups and gender matter? *College student journal*, 52:441.
- Robert M. Entman. 1993. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4):51–58.
- Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.
- Barney G. Glaser and Anselm L. Strauss. 1967. *Discovery of Grounded Theory Strategies for Qualitative Research*. AldineTransaction, London.
- Erving Goffman. 1991. *Frame analysis an essay on the organization of experience*. Northeastern Univ. Pr.
- T. Gowan. 2010. *Hobos, Hustlers, and Backsliders: Homeless in San Francisco*. University of Minnesota Press.
- John E Hocking and Samuel G. Lawrence. 2000. Changing attitudes toward the homeless: The effects of prosocial communication with the homeless. *Journal of Social Distress and the Homeless*, 9:91–110.
- Anna Lauren Hoffmann. 2019. Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society*, 22:900 – 915.
- Minda Hu, Ashwin Rao, Mayank Kejriwal, and Kristina Lerman. 2021. Socioeconomic correlates of anti-science attitudes in the us. *Future Internet*, 13(6):160.
- Ben Hutchinson, Vinodkumar Prabhakaran, Emily Denton, Kellie Webster, Yu Zhong, and Stephen Denuyil. 2020. Social biases in nlp models as barriers for persons with disabilities. *arXiv preprint arXiv:2005.00813*.
- Nathan Kim, Colette Auerswald, and Jessica Lin. 2013. Perceptions regarding homelessness: Analyzing us tweets.
- JB Kingree and Walter F Daves. 1997. Preliminary validation of the attitudes toward homelessness inventory. *Journal of community psychology*, 25(3):265–288.
- Svetlana Kiritchenko, Georgina Curto Rex, Isar Nejadgholi, and Kathleen C. Fraser. 2023. Aporophobia: An overlooked type of toxic language targeting the poor. In *The 7th Workshop on Online Abuse and Harms (WOAH)*, pages 113–125, Toronto, Canada. Association for Computational Linguistics.
- John Lalor, Yi Yang, Kendall Smith, Nicole Forsgren, and Ahmed Abbasi. 2022. Benchmarking intersectional biases in NLP. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3598–3609, Seattle, United States. Association for Computational Linguistics.
- Barrett A Lee, Sue Hinze Jones, and David W Lewis. 1990. Public beliefs about the causes of homelessness. *Social Forces*, 9:253–65.
- Lori Leibowitz and Joachim Krueger. 2005. Attitudes and stereotypes about the homeless: A study on self-persuasion and stereotype accuracy. *Journal of Social Distress and the Homeless*, 14:125–150.

- Minzhi Li, Taiwei Shi, Caleb Ziems, Min-Yen Kan, Nancy F Chen, Zhengyuan Liu, and Diyi Yang. 2023. Coannotating: Uncertainty-guided work allocation between human and large language models for data annotation. *arXiv preprint arXiv:2310.15638*.
- Bruce G Link, Barrett A Lee, and Paul A Toro. 1991. *Images of the homeless: Public views and media messages*. *Housing Policy Debate*, 2(3):649–682.
- Bruce G Link, Sharon Schwartz, Robert Moore, Jo Phe-
ian, Elmer Struening, Ann Stueve, and Mary Ellen Colten. 1995. *Public knowledge, attitudes, and be-
liefs about homeless people: evidence for com-
passion fatigue*. *American journal of community psy-
chology*, pages 533–55.
- Julia Mendelsohn, Ceren Budak, and David Jurgens. 2021. Modeling framing in immigration discourse on social media. *arXiv preprint arXiv:2104.06443*.
- Charles H. Moore, David W. Sink, and Patricia Hoban-
Moore. 1988. *The politics of homelessness*. *PS:
Political Science* 38; *Politics*, 21(1):57–63.
- Lauren E Mullenbach, Nick A. D Pitas, and Benjamin Hickerson. 2023. *The role of social contact in public perceptions of homelessness in parks and public spaces*. *International Journal of Public Administra-
tion*, 0(0):1–14.
- Deepa Narayan and Patti Petesch. 2002. *Voices of the poor from many lands*. 1.
- Isar Nejadgholi, Kathleen Fraser, and Svetlana Kir-
itchenko. 2022. *Improving generalizability in im-
plicitly abusive language detection with concept ac-
tivation vectors*. In *Proceedings of the 60th Annual
Meeting of the Association for Computational Lin-
guistics (Volume 1: Long Papers)*, pages 5517–5529, Dublin, Ireland. Association for Computational Lin-
guistics.
- Nicholas Pangakis, Samuel Wolken, and Neil Fasching. 2023. Automated annotation with generative ai re-
quires validation. *arXiv preprint arXiv:2306.00176*.
- Marjorie J. Robertson and Milton Greenblatt. 1992. *Homelessness*, pages 339–349. Springer US, Boston, MA.
- Robert Rosenheck, Ellen Bassuk, and Amy Salomon. 1999. Special populations of homeless americans. In *Practical lessons from the 1998 national symposium on homeless research*, pages 1–27.
- Sara Rosenthal, Noura Farra, and Preslav Nakov. 2017. Semeval-2017 task 4: Sentiment analysis in twitter. In *Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017)*, pages 502–518.
- Christopher Michael Rytting, Taylor Sorensen, Lisa Argyle, Ethan Busby, Nancy Fulda, Joshua Gubler, and David Wingate. 2023. Towards coding social science datasets with language models. *arXiv preprint arXiv:2306.02177*.
- Kirsten Moore Sheeley, Alisa Belinkoff Katz, Andrew Klein, Jessica Richards, Fernanda Jahn Verri, Marques Vestal, and Zev Yaroslavsky. 2021. *The making of a crisis: A history of homelessness in los angeles*.
- Megan Smith, Thomas Byrne, and Ann Elizabeth Mont-
gomery. 2023. *Examining the relationships between exposure to homelessness and housing insecurity and perceptions and attitudes toward homelessness among low- and moderate-income people*. *Journal of Social Distress and Homelessness*, 0(0):1–12.
- John Solas. 2018. Deserving to deserve: Challenging discrimination between the deserving and undeserv-
ing in social work. *Journal of Social Work Values and Ethics*, 15(2):62–70.
- Tanya Sousa, Alyssa Andrichik, Ed Prestera, Katherine Rush, Colette Tano, and Micaiah Wheele. 2023. *The 2023 annual homelessness assessment report (ahar) to congress*.
- Manfred Stede, Yannic Bracke, Luka Borec, Neele Charlotte Kinkel, and Maria Skeppstedt. 2023. *Framing climate change in nature and science editorials: Applications of supervised and unsupervised text categorization*. *Journal of Computational Social Science*, 6(2):485–513.
- Carolyn J Tompsett, Paul A Toro, Melissa Guzicki, Manuel Manrique, and Zatakia Jigna. 2006. *Home-
lessness in the united states: assessing changes in prevalence and public opinion, 1993-2001*. *American Journal of Community Psychology*, 37:47–61.
- Paul Toro and D McDonnell. 1992. *Beliefs, attitudes, and knowledge about homelessness: A survey of the general public*. *American journal of community psychology*, 20:53–80.
- Jack Tsai, Crystal Lee, Thomas Byrne, Robert Pietrzak, and Steven Southwick. 2017. *Changes in public attitudes and perceptions about homelessness between 1990 and 2016*. *American Journal of Community Psychology*, 60.
- Jack Tsai, Crystal YS Lee, Jianxun Shen, Steven M Southwick, and Robert H Pietrzak. 2019. Public exposure and attitudes about homelessness. *Journal of Community Psychology*, 47(1):76–92.
- Lindsey L. Upton. 2016. *Care, control, or criminaliza-
tion?: Discourses on homelessness and social responses*. Ph.D. thesis, Old Dominion University.
- Angelina Wang, Jamie Morgenstern, and John P Dick-
erson. 2024. *Large language models cannot replace human participants because they cannot portray identity groups*. ArXiv, abs/2402.01908.
- Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837.

Michael Wiegand, Josef Ruppenhofer, and Thomas Kleinbauer. 2019. [Detection of Abusive Language: the Problem of Biased Datasets](#). 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 602–608, Minneapolis, Minnesota. Association for Computational Linguistics.

A Frame Prediction: Dataset Statistics and Model Training Details

Combining expert annotated posts along with posts that are annotated by GPT-4 and expert validated using GPT-4 + Expert in §3, we obtain a total of 10400 posts. We then split these posts into a train-validation-test split using the following criteria:

- We ensure roughly an 80:10:10 proportion for the above mentioned split.
- The test set contains *only* domain-expert annotated posts. We pick 280 posts that were used in our agreement set in §2, and add another 1000 instances to the pool. This overall set of 1280 posts is used to pick the best performing Flan-T5 variation for final inference on a larger set of posts.
- The remaining set of domain-expert annotated and GPT-4 + Expert annotated posts are then combined and split into a train and validation split.

We use the pre-trained Flan-T5 models for the conditional generation task, pooled from the HuggingFace¹⁰ repository. The task is structured as follows:

- **Input format:** The input format consists of the following instruction, along with the post to be labeled – Classify the given post into one or more of the following 10 labels: {list of labels} post: {}
- **Output format:** If a post needs to be filtered out using the Public Opinion filter, we generate a ‘0’ label, otherwise, we generate a comma-separated list of Issue-specific frames. Note, this is slightly different from how GPT-4 predictions are generated where GPT-4 is prompted in a sequential manner to first filter out the posts according to the Public Opinion filter and then predict the Issue-specific frames on the remaining posts whereas we use Flan-T5 to predict the Public Opinion filter in conjunction with the 9 Issue-specific frames all at once.

We demonstrate the data splits, training details and results of different Flan-T5 variations in [Table 5](#), [Table 6](#), [Table 7](#), [Table 8](#) and [Table 4](#).

¹⁰https://huggingface.co/docs/transformers/model_doc/flan-t5

B Annotation Reminders and Common Questions

B.1 Common Questions

1. References to homelessness as lowest point, or linguistic structure of “listing”: ‘the country is in shambles, immigration sucks, inflation is up, and homeless crisis is up’ → are **GovCrit**.
2. Posts that fall under **HarmGen**. are references to homelessness as lowest point, usually these posts read “i’d rather be homeless than blah blah blah”
3. **SolnInt.** can be expressed in many different ways: volunteering, advocacy etc. If someone is defending PEH from harmful stereotypes, tag with both **SolnInt.** and **HarmGen**.
4. If someone is advocating against a specific harmful stereotype – label as **HarmGen**. and **SolnInt.**, otherwise, neutral sentiments and positive comments about PEH can be labeled as **Interact**.
5. **HarmGen**. coupled with **SolnInt.** means defending against a specific harmful stereotype
6. **HarmGen**. on its own or coupled with **Interact**. means person writing the post is making a harmful stereotype
7. If no harmful stereotype is mentioned and person is just making a positive comment about PEH: **Interact**., possibly **SolnInt**.
8. ‘Homeless edition’, ‘anti-homeless furniture’ and other metaphors/similes ascribing homelessness with negative qualities should be tagged with **HarmGen**.
9. If comparing homelessness to other issues and discussing prioritization of issues: **(Un)Deserv**., potentially **SolnInt**.
10. ‘looking homeless’ implying PEH have a dirty, unkempt appearance is **HarmGen**.
11. There is a difference between one experience of PEH doing something harmful vs. generalizing this attribute to PEH as a whole. **HarmGen**. referring to generalizations made

- about PEH population as a whole as opposed to one incident.
12. If specifically discussing how money, aid or resources should be allocated across PEH vs. other groups such as immigrants or refugees, tag with **MoneyAid**.
 13. **GovCrit**. also include posts which point out political parties by name explicitly → you're looking for a juxtaposition between parties, for e.g using democratic ideals specifically to critique the party's response to homelessness in comparison to republican ideals
 14. **SolnInt**. includes calls to action which are more directed and pointed and often more declarative in nature, For ex. 'Hurry up govs!' or includes critiquing or discussing specific solutions
 15. Statements like: "the city is so great, why do you have homelessness" should be tagged with **GovCrit**.
 16. When the post is critiquing a solution or is questioning a solution of intervention, it should be augmented with additional labels from the **Critiques** theme along with **SolnInt**.
 17. If there is critique of solutions funded by the government or government sponsored, then label as **GovCrit**.

B.2 Reminders

1. Google names of politicians if they show up (**GovCrit**. frame often names politicians, if you suspect that's what they're doing and don't recognize the name, then google to confirm)
2. Don't open links to external sources
3. If the post mentions policing of homelessness, consider tagging with **GovCrit**. since this is government mandated
4. Critiquing of how the ultra-rich are donating money is **MoneyAid**.
5. discussion/debate around contributing factors to homelessness: **SolnInt**. potentially

6. If someone is defending PEH from harmful stereotypes, tag with both **SolnInt**. and **HarmGen**.
7. **HarmGen**. on its own or coupled with **Interact**. → person is making a harmful stereotype
8. If author of the post is expressing prejudice against PEH → tag with **HarmGen**. If the author of the post is calling out prejudice, tag with **HarmGen**. and **SolnInt**.

C GPT-4 Prompt and Error Analysis

We list the final prompt that we achieve after iterating through the OATH-Frames instructions, for obtaining GPT-4 annotations (Table 10).

Our prompt consists of a two-step process: (1) Using the Public Opinion prompt to filter relevant posts, followed by (2) Using OATH-Frames to get multi-label Issue-specific frames for each post. We also included examples of frames with the highest false discovery rates and highest precision as predicted by GPT-4 in Table 11.

D OATH-Frames Iterations

In Figure 8, we provide an overview of an earlier iteration of OATH-Frames where we had close to 50 Issue-specific frames across six different themes. We found that there was a lot of overlap between frames, and our consensus sessions involved choosing the most salient themes in our analysis, and reaching consensus on the frames that should be kept. Our final iteration of OATH-Frames is shown in Figure 9. In our primary iterations, we also considered annotating factors that are often conflated with homelessness. Examples of such factors include substance use and mental illness. Such factors provide important context for how public opinion can be influenced by the social and economic climate of a particular region. We provide details of these factors in Figure 10 and leave this up to future works as an extension of OATH-Frames.

E Frame Co-Occurrence

Figure 11 is a plot of how frequently the Issue-specific frames co-occur with each other in our set of 4.1k expert annotated posts. We find that **(Un)Deserv**. and **MoneyAid**. frequently co-occur. The time period of our posts is from 2021-2023, and in this time, a main agenda for

the republic party was sending asylum seekers across the country, predominantly to blue states; this resulted in a lot of immigrants becoming homeless and also resulted in a news cycle that pitted immigrants against PEH in competition for jobs, housing and funding (Abdalla, 2022). Furthermore, GovCrit. and MoneyAid. frequently co-occur as critique against government institutions and politicians is often in the context of funding and resource allocation. HarmGen. co-occurs frequently with NIMBY and Interact.. NIMBY rhetoric can be expressed in harmful and exclusionary ways towards PEH, and furthermore, personal interactions with PEH can elicit more harmful stereotyping.

F Domain + Trained Expert Annotations

We provide an example of our annotation interface in Figure 16 where we create a custom interface in the Amazon Mechanical Turk Sandbox environment. The expert first determines whether the post should be filtered out if it is not relevant to public attitudes towards homelessness (Boydston and Gross, 2013). If the post is not filtered out, the expert must select at least one or more Issue-specific frames to label the post. We provide examples of our expert annotated posts in Table 13.

Aggregated (Across 9 Issue-specific frames)			
Metrics	Flan-T5-Large	Flan-T5-Base	Flan-T5-Small
Precision	0.50	0.48	0.46
Recall	0.48	0.47	0.41
F ₁	0.50	0.47	0.42

Table 4: **Macro averaged precision, recall and F₁ on a larger test set (1280):** Comparing variations of finetuned Flan-T5 on a larger test set (1280)

Annotation Method	Posts	Filtered Out	Remaining
Domain + Trained Experts	5.3k	1.2k	4.1k
GPT-4	10k	3k	7k
GPT-4 + Expert (expert validated)	4.4k \in 7k GPT-4	300	4.1k
Flan-T5-Large (model predictions)	3.1M	700k	2.4M

Table 5: **Annotation Sources:** We include the number of posts originally annotated, the number of posts filtered out after applying the Public Opinion filter to filter out posts not related to public attitudes towards homelessness and the number of posts remaining for our analysis. We do our socio-political analysis on our set of posts annotated with OATH-Frames after filtering.

Split	Posts
Train	8217
Dev	913
Test	1280
Agreement Test Set	280 (also incl. in Test)

Table 6: **Dataset Splits for training Flan-T5:** The Train split contains both GPT-4 + Expert and singly-annotated (domain+trained expert) posts. The Dev split contains singly annotated (domain+trained expert) posts. The Test split contains 280 multiply annotated (domain+trained expert) posts and 1k singly annotated (domain+trained expert) posts. The Agreement Test Set contains 280 multiply annotated posts and was used to calculate our evaluation metrics in Table 2.

models	Config	Assignment		
	Flan-T5-Large			
	Number of parameters: 770M			
Flan-T5-Base				
models	Number of parameters: 250M			
	Flan-T5-Small			
	Number of parameters: 80M			
train batch size				
eval batch size				
seed				
max epochs				
learning rate				
learning scheduler				
GPU				
# of GPUs				
Training time				
Stopping Criteria				

Table 7: **Model Training Details:** We provide details about our models that were finetuned on our Domain + Expert data and GPT-4 + Expert data where Flan-T5-Large achieved the highest Macro F₁ across all 9 Issue-specific frames.

Aggregated (Across 9 Issue-specific frames)						
Metrics	1st Author	GPT-4	GPT-4 + Expert	Flan-T5-Large	Flan-T5-Base	Flan-T5-Small
Precision	0.6 ± 0.06	0.39 ± 0.05	0.51 ± 0.07	0.54 ± 0.07	0.47 ± 0.06	0.45 ± 0.08
Recall	0.53 ± 0.02	0.64 ± 0.05	0.6 ± 0.05	0.52 ± 0.05	0.47 ± 0.05	0.39 ± 0.03
F ₁	0.54 ± 0.01	0.46 ± 0.04	0.52 ± 0.04	0.51 ± 0.03	0.45 ± 0.04	0.38 ± 0.04
Fleiss' κ	0.47 ± 0.02	0.31 ± 0.04	0.37 ± 0.04	0.34 ± 0.03	0.31 ± 0.02	0.24 ± 0.01

Table 8: **Macro-averaged precision, recall and F1 aggregated across all 9 Issue-specific frames:** Comparing variations of Flan-T5 with other annotation strategies evaluated on the multiply annotated Agreement Test Set (280).

Issue-specific	Metrics	1st Author	GPT-4	GPT-4 + Expert	Flan-T5-Large
GovCrit.	Prec.	0.7 ± 0.1	0.64 ± 0.16	0.68 ± 0.16	0.61 ± 0.13
	Rec.	0.69 ± 0.14	0.82 ± 0.09	0.81 ± 0.09	0.74 ± 0.09
	F1	0.68 ± 0.04	0.7 ± 0.08	0.72 ± 0.07	0.65 ± 0.07
MoneyAid.	Prec.	0.64 ± 0.12	0.45 ± 0.14	0.5 ± 0.16	0.60 ± 0.23
	Rec.	0.54 ± 0.2	0.87 ± 0.06	0.86 ± 0.07	0.61 ± 0.07
	F1	0.57 ± 0.17	0.58 ± 0.14	0.61 ± 0.15	0.59 ± 0.18
SocCrit.	Prec.	0.33 ± 0.1	0.36 ± 0.09	0.47 ± 0.09	0.44 ± 0.15
	Rec.	0.45 ± 0.17	0.71 ± 0.12	0.56 ± 0.13	0.43 ± 0.11
	F1	0.37 ± 0.12	0.46 ± 0.06	0.5 ± 0.05	0.42 ± 0.09
SolnInt.	Prec.	0.79 ± 0.11	0.54 ± 0.07	0.65 ± 0.07	0.67 ± 0.04
	Rec.	0.58 ± 0.04	0.72 ± 0.09	0.66 ± 0.1	0.70 ± 0.06
	F1	0.67 ± 0.06	0.62 ± 0.06	0.65 ± 0.06	0.68 ± 0.03
Interact.	Prec.	0.49 ± 0.1	0.34 ± 0.07	0.45 ± 0.1	0.53 ± 0.11
	Rec.	0.29 ± 0.08	0.73 ± 0.13	0.66 ± 0.15	0.56 ± 0.14
	F1	0.36 ± 0.08	0.45 ± 0.07	0.52 ± 0.07	0.54 ± 0.11
MediaPort.	Prec.	0.72 ± 0.12	0.41 ± 0.09	0.6 ± 0.11	0.75 ± 0.16
	Rec.	0.71 ± 0.11	0.69 ± 0.14	0.67 ± 0.15	0.64 ± 0.08
	F1	0.7 ± 0.07	0.5 ± 0.08	0.63 ± 0.11	0.69 ± 0.10
NIMBY	Prec.	0.37 ± 0.19	0.27 ± 0.19	0.48 ± 0.39	0.50 ± 0.42
	Rec.	0.32 ± 0.29	0.47 ± 0.29	0.47 ± 0.29	0.28 ± 0.21
	F1	0.28 ± 0.2	0.3 ± 0.16	0.39 ± 0.21	0.34 ± 0.26
(Un)Deserv.	Prec.	0.65 ± 0.11	0.32 ± 0.06	0.56 ± 0.07	0.64 ± 0.14
	Rec.	0.34 ± 0.08	0.87 ± 0.11	0.75 ± 0.17	0.73 ± 0.17
	F1	0.44 ± 0.07	0.46 ± 0.06	0.63 ± 0.08	0.67 ± 0.10
HarmGen.	Prec.	0.64 ± 0.08	0.53 ± 0.13	0.58 ± 0.15	0.61 ± 0.13
	Rec.	0.57 ± 0.09	0.48 ± 0.07	0.48 ± 0.07	0.51 ± 0.05
	F1	0.6 ± 0.04	0.49 ± 0.06	0.51 ± 0.06	0.55 ± 0.05

Table 9: Precision, Recall and F1. for each of the Issue-specific frames on Agreement Test Set (280).

Public Opinion prompt

System Prompt: You are an AI model trained to classify tweets related to homelessness into 2 different labels. The labels are `<attitude_towards_homelessness>` and `<other>`.

Label Descriptions: - `<attitude_towards_homelessness>`: Includes tweets about homelessness to talk about another topic or are generally about the social issue of homelessness. They cover a range of topics related to homelessness, including criticism of government bodies, institutions, or political parties, discussions about the allocation and disbursement of money, aid, and resources, criticism of societal attitudes towards homelessness, debates about who is more deserving of resources, harmful generalizations or stereotypes about homeless people, opposition to having homeless people in local areas or neighborhoods, references to media portrayals of homelessness, anecdotes about interactions with homeless people, and suggestions or ideas for solving the homelessness crisis.

- `<other>`: Includes personal anecdotes from people experiencing homelessness who are sharing their personal experience while being homeless or asking for assistance and aid. This category does NOT include tweets about fictional characters and personal interactions with other homeless people. Includes statements that are nonsensical or difficult to decipher and require access to additional resources like links, media, images, etc in order to properly interpret the tweet or references to homeless animals or being politically homeless.

Prompt: Classify the following tweet into one of the provided labels:

"post"

In concise points, please provide the relevant label that best characterizes the content of the tweet. Do not "read into" the text with interpretations, stick to the definitions of the categories strictly. The format should be the predicted label, followed by "because", followed by reason. Do not add any additional text.

Feel free to reference the label descriptions to support your classification. Provide any relevant context that influenced your classification.

OATH-Frames prompt

System Prompt: You are an AI model trained to classify tweets related to homelessness into 9 different labels. The labels include `<government_critique>`, `<money_aid_resource>`, `<public_critique>`, `<deserving undeserving_of_resources>`, `<harmful_statements_against_homelessness>`, `<not_in_my_backyard>`, `<media_portrayal>`, `<interaction_with_homeless_person>`, and `<solutions_interventions>`.

Label Descriptions:

- `<government_critique>`: criticism about the government body, government institutions or political parties including critique of specific politicians, policies about homelessness, critique of programs that are being funded or considered by the government such as welfare programs, and the policing of homelessness. Also includes statements where homelessness is used as a vehicle or stand-in to talk about a broader issue portraying homelessness amongst other negative social and government problems in a list-like manner in a tweet like "murder rates, homelessness, immigration and inflation. all suck". Also includes statements that mention names of politicians.

- `<money_aid_resource>`: Primarily includes discussion of money, for long term relief of homelessness. Includes aid or resource disbursement and allocation by government, institutions, organizations or wealthy individuals (not regular public) and also includes discussion or critique and suggestions on how the government decides to spend money and resources. Also includes discussions of giving or providing money, aid and resources to homeless people.

- `<public_critique>`: Criticism of society in general or social norms that includes discussion of society at large instead of specific people, often pointing out hypocrisy and critiquing society's general attitudes towards homelessness. Also includes critiquing someone helping homelessness in order to gain some personal benefit where someone is being explicitly called out for doing charitable acts while filming a video or for recognition.

- `<deserving undeserving_of_resources>`: Discussion of competing priorities where homelessness is compared to other issues that more or less deserve aid and resources. Includes statements that express anti-immigration and support for policies, political initiatives and actions that restrict immigration often comparing and prioritizing aid to people experiencing homelessness over immigrants. Also includes nationalistic statements that prioritize one's own nation over others including discussion about prioritizing aid and relief for veterans and the nation's citizens over non-citizens.

- `<harmful_statements_against_homelessness>`: Blanket statements that generalize a negative, harmful or undesirable attribute to all people experiencing homelessness and invoke stereotypes and make assumptions about people experiencing homelessness as a whole. Examples include statements that say all people experiencing homelessness are violent, addicts, thieves, mentally ill, unkempt, dirty, and poor at managing finances and also comparing dirty, disheveled clothing to 'looking homeless'. Includes statements that express prejudice against homelessness such as sexism, homophobia, racism, anti-semitism and transphobia or dehumanize people experiencing homelessness depriving them of positive human qualities and viewing them as sub-human or as trash. Includes statements that portray homelessness as the lowest point in one's life where homelessness is used as an example of something wrong or bad. This also includes metaphors to describe objects like anti-homeless. Could also include statements that express the desire to be violent strictly against people experiencing homelessness including threats against homelessness. Includes statements that portray homelessness as the lowest point in one's life where homelessness is used as an example of something wrong or bad. This includes listing homelessness in conjunction with other issues that are viewed as problematic or negative.

- `<not_in_my_backyard>`: Opposition by residents to proposed developments in their local area, as well as support for strict land use regulations against wanting to see homelessness in their local area and neighborhood. Also includes displacement sweeps to remove PEH from certain areas and neighborhoods.

- `<media_portrayal>`: Reference to a fictional character that is portraying homelessness and includes tweets and links about local news media.

- `<interaction_with_homeless_person>`: Only includes anecdotes describing a real-life interaction with a homeless person.

- `<solutions_interventions>`: Suggestions, remedies, problem solving and ideas for alleviating the homelessness crisis including support for policy reform, existing policies and welfare programs. Includes individual people giving money, food and help for immediate relief of homelessness. Also includes charitable acts, non-profit work, providing help and emergency aid relief, and defending people experiencing homelessness from harmful stereotypes and generalizations and advocating for positive qualities for people experiencing homelessness. Also includes call to action statements that invoke a sense of urgency in taking action towards helping the homelessness crisis.

Prompt: Classify the following tweet into one or more of the provided labels:

"post"

In concise points, carefully assess the relevant label(s) that best characterize the content of the tweet; try to list all the labels that are applicable for the tweet. Do not "read into" the text with interpretations or indications or make any assumptions, and stick to the definitions of the labels strictly. Each individual label should be followed by "because", followed by the reason for why that label was picked. Do not add any additional text. You have to select atleast one label, you cannot leave it out.

Feel free to reference the label descriptions to support your classification. Provide any relevant context that influenced your classification.

Table 10: **GPT-4 prompts:** We use the above Public Opinion and OATH-Frames prompt in sequence, to obtain GPT-4 labels.

Issue-specific Frame	Examples
(Un)Deserv. (Highest False Discovery)	<p>Post: @mention we have 600,000 homeless and you sound self-centered. typical capitalist mindset.</p> <p>Predicted CoT: Because the post mentions the number of homeless people, implying that resources should be allocated to them, and criticizes the individual for not prioritizing this issue.</p> <p>Observation: The post uses homelessness as an example to critique an individual, but the CoT makes implications of PEH being in competition for resources when there is not enough information in the post to make that claim.</p> <p>Post: @mention furthermore this city has at least three homeless shelters available for these people - they just have to stay sober.</p> <p>Predicted CoT: Because the post implies that homeless people are undeserving of shelter if they can't stay sober, suggesting a condition for them to receive resources.</p> <p>Observation: The post does not imply that PEH are in competition for resources, rather it is making a harmful generalization about substance use.</p>
Interact. (Highest False Discovery)	<p>Post: @mention are you helping homeless british people including veterans?</p> <p>Predicted CoT: Because the post is addressing someone (presumably a person or organization) about their actions towards homeless people, which can be seen as a form of interaction.</p> <p>Observation: It is not clear if this post is talking about a one on one interaction with PEH, the CoT is generalizing and reading into the post.</p> <p>Post: @mention wonder if any one here has offered up a room to the many thousands of irish homeless</p> <p>Predicted CoT: Because the post is suggesting that the individuals addressed could have a direct interaction with homeless people by offering them a room</p> <p>Observation: The post does not mention a direct interaction with PEH and the CoT implies that GPT-4 is reading into the post.</p>
HarmGen. (Highest Precision)	<p>Post: @mention how are they a nuisance? i'd rather have that than the homeless; their garbage littering public spaces.</p> <p>Predicted CoT: Because the post generalizes all homeless people as a nuisance and associates them with littering public spaces, which is a negative stereotype.</p> <p>Observation: The CoT correctly identifies that PEH are all being portrayed as a nuisance which can be considered a negative stereotype.</p> <p>Post: @mention furthermore this city has at least three homeless shelters available for these people - they just have to stay sober. they're out there by their f*cken choice</p> <p>Predicted CoT: Because the post generalizes that all homeless people are choosing to stay on the streets due to their inability to stay sober, which is a harmful stereotype.</p> <p>Observation: The CoT reasoning correctly identifies the harmful generalization about PEH and substance use, and that all PEH are responsible for their own plight.</p>
GovCrit. (Highest Precision)	<p>Post: @mention @mention @mention @mention ur obviously not from the city cause bratton is a notorious bigot who hates the homeless, blacks and the poor. he said what he said and he meant what he meant... and what he meant was "homeless people are a nuisance." his record speaks for itself, but go on...</p> <p>Predicted CoT: because the post criticizes a specific government figure, Bratton, for his alleged negative attitudes towards the homeless, blacks, and the poor. The user is critiquing his policies and statements about homelessness.</p> <p>Observation: The CoT reasoning correctly identifies critique against William Bratton, the NYC law enforcement officer.</p> <p>Post: @mention @mention big ron is the one who destroyed the institutions that held our mentally ill. he put them on the streets that added thousands of homeless and mentally ill people to live like animals. and people think the republicans are there to help all the conservatives, wrong! just the rich.</p> <p>Predicted CoT: because the post criticizes a specific politician ("big ron") for his policies that allegedly led to an increase in homelessness and mental illness on the streets. The post also criticizes the Republican party, implying they only help the rich.</p> <p>Observation: The CoT reasoning correctly identifies critique against a politician for his policies and political party values.</p>

Table 11: **GPT-4 Error Analysis:** We compare instances that are annotated using GPT-4 and provide the generated Chain-of-Thought explanations, along with pointers of how GPT-4 is getting these instances incorrect/correct.

PEH	Vulnerable Population	Significant Issue-specific Frames	Example
'Homeless' Veteran	Ukrainian	GovCrit., MoneyAid.	@mention they have been given not 1 billion dollars, not 80 billion dollars (russian military budgets, but 180 billion dollars. do you really think that money is being used to help the ukrainian people-or is it something else? go help a homeless vet today! Post labeled with: MoneyAid., GovCrit., (Un)Deserv.
	Illegal Alien	GovCrit., Interact., MediaPort., NIMBY	and this is being done for your homeless, especially homeless veterans, on your city streets? no? well? i didn't think so. send the illegal aliens back to where they came from! vote red! Post labeled with: GovCrit., NIMBY, (Un)Deserv.
	Asylum Seeker	SocCrit., SolnInt., Interact., MediaPort., NIMBY, HarmGen.	@mention better still, if you're so concerned for homeless veterans get off your soapbox and go volunteer to help them. they won't be blaming asylum seekers for their situation. and maybe while you're at it, try meeting an asylum seeker... #spreadinghatethroughignorance Post labeled with: SocCrit., SolnInt., (Un)Deserv.
	Refugee	SocCrit., SolnInt., HarmGen.	@mention typical american who never travels unless on a cruise ship or american! asia is amazing & for someone who is the daughter of a disabled refugee/ don't lecture me on not respecting our amazing veterans. at least we take care of them here - how many millions of homeless vets in us? Post labeled with: SocCrit., (Un)Deserv.
	Immigrant	NIMBY	@mention oh please they are dropping off bus loads of immigrants to hotels 6 blocks from my house that we the tax payers get to pay for. we have homeless vets what about them !! wake up Post labeled with: NIMBY, GovCrit., (Un)Deserv.
	Ukrainian	GovCrit., MoneyAid.	ukrainian refugees can get ssi, housing benefits, and free health care. meanwhile, we have a lot of homeless people who have no access to such benefits - we have the highest number of the unemployed black population. Post labeled with: MoneyAid., GovCrit., (Un)Deserv.
	Illegal Alien	GovCrit., NIMBY, HarmGen.	@mention when adams was sworn into office, there were already an estimated 1/2 million illegal aliens in nyc. it hasn't been reported yet, whether or not adams has continued the deblasio & bloomberg policy of shipping homeless people outside of nyc, as a cheaper option than sheltering. Post labeled with: NIMBY, GovCrit., (Un)Deserv.
	Asylum Seeker	SocCrit., SolnInt., MediaPort., HarmGen.	@mention i take it you'll be taking your colleagues to task then if they slam those on welfare ? maybe; just maybe , you could support both homeless people and asylum seekers . i know that's a hard one for you to understand. Post labeled with: SolnInt., SocCrit., (Un)Deserv.
	Refugee	SocCrit., SolnInt., MediaPort.	@mention refugee camps offer a safe haven for refugees and meet their most basic needs such as food water shelter medical treatment and other basic services there are homeless people all over america that don't even get that..yeah sorry but hater @mention bad sad story been played out #btc Post labeled with: SocCrit., SolnInt., (Un)Deserv.
	Immigrant	Interact., NIMBY, HarmGen.	@mention new york city's pulling the rug out from underneath the illegal immigrants kicking them out of the \$700 a day hotel rooms. was it the bedbugs or the thousands of empty bottles of beer and liquor. we used to have a problem with homeless people now we have a catastrophe of homeless Post labeled with: NIMBY, GovCrit., (Un)Deserv.
'Homeless' American	Ukrainian	GovCrit., MoneyAid.	@mention grifter journalist got to grift . how about reporting on how close we are from being in a nuclear war with russia. or how corrupted our politicians have become. write about how homeless americans live across this country and the us thinks sending money to ukraine is saving anyone Post labeled with: MoneyAid., GovCrit., (Un)Deserv.
	Illegal Alien	GovCrit., NIMBY, HarmGen.	@mention i say bus the illegal aliens to the places where the homeless americans and veterans are freezing outside. that way our homeless citizens might get some of their tax resources back by being mistaken for illegal aliens or (lol) migrants. Post labeled with: NIMBY, GovCrit., (Un)Deserv.
	Asylum Seeker	SocCrit., SolnInt., HarmGen.	@mention hey there are plenty of homeless americans where is the welcome committee for them?! you don't care about being a " humanitarian " or you would start by those who have birth rights in the usa not false claims to asylum... you people are pathetic..#clownsagenda Post labeled with: SocCrit., SolnInt., (Un)Deserv.
	Refugee	SocCrit., SolnInt., Interact., MediaPort.	@mention see there you go avoiding the question, can't ask me to play semantics with you if you won't reciprocate. you got one spot left, a homeless american or fresh new refugee, who gets helped first. if you say at the same time then you are just avoiding your own hard truth Post labeled with: SocCrit., SolnInt., (Un)Deserv.
	Immigrant	SocCrit., SolnInt., Interact., MediaPort., NIMBY, HarmGen.	@mention well... if we didn't have all the illegals here. maybe we could actually get a handle on homeless americans. problem solved. what's your solution? just empty platitudes. Post labeled with: SocCrit., SolnInt., (Un)Deserv.

Table 12: **Statistically significant frames for posts containing (PEH, Vulnerable Population) co-occurrence pairs.** We observe fine-grained differences in attitudes when PEH are compared to different vulnerable populations.

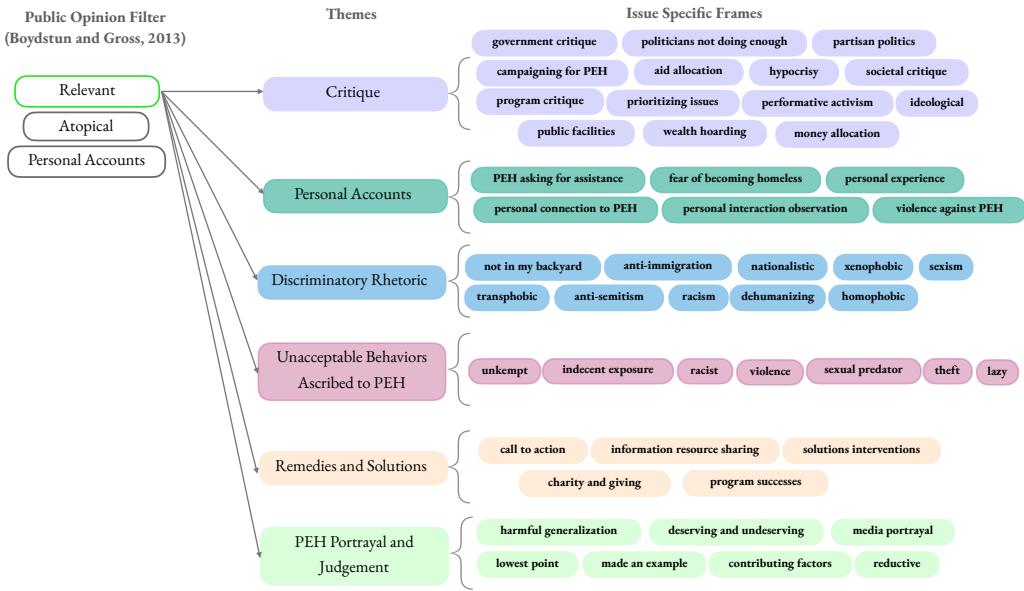


Figure 8: **OATH-Frames v1:** Our first iteration of OATH-Frames included more themes and Issue-specific frames. Through our coding consensus iterations in our team of domain experts, we condensed Issue-specific frames based on saliency in our pilot studies, and relevancy to our topic of attitudes towards homelessness. We went through many intermediate iterations of OATH-Frames the using principles from grounding theory, we reduced our number Issue-specific frames to more closely capture public opinions and attitudes towards homelessness.

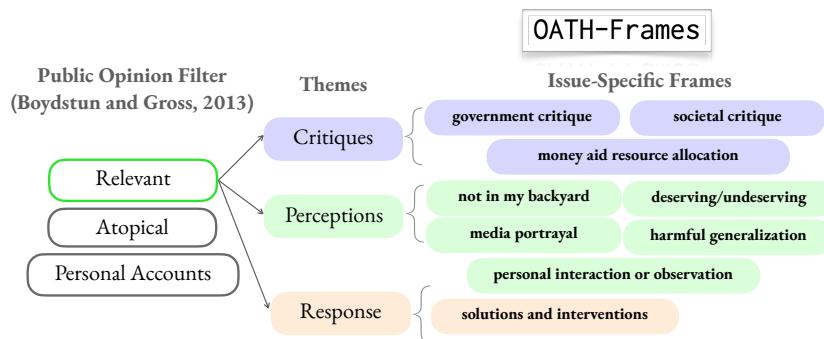


Figure 9: **OATH-Frames Final Iteration:** We converged on a set of 9 Issue-specific frames and 3 broad themes as the final iteration of OATH-Frames that we used to perform our analysis and large scale annotation.

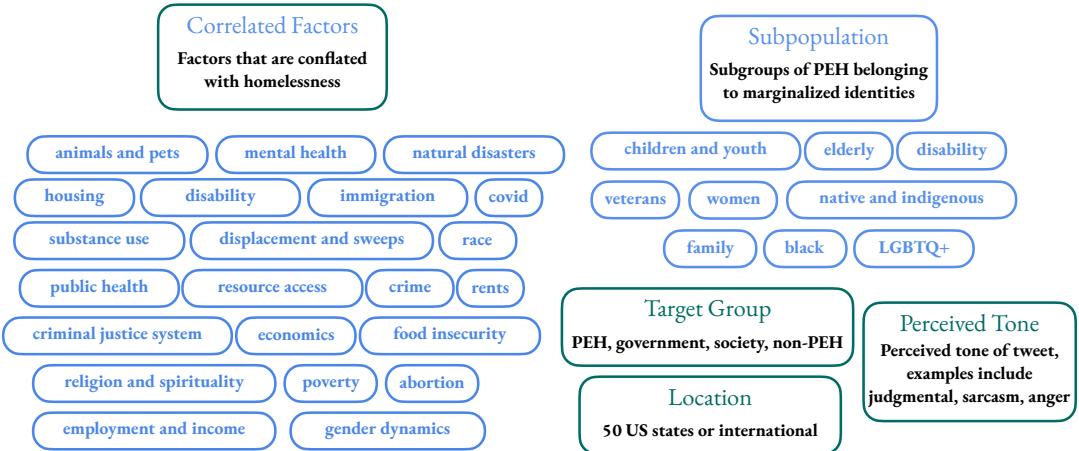


Figure 10: **Additive Labels:** We also considered additive labels that further contextualized the social media post with correlated factors often discussed in conjunction with homelessness, subpopulations of PEH, the target group and perceived tone of the post and any references to location. We hope to incorporate additive labels in future works to further understand the social context of attitudes towards homelessness.

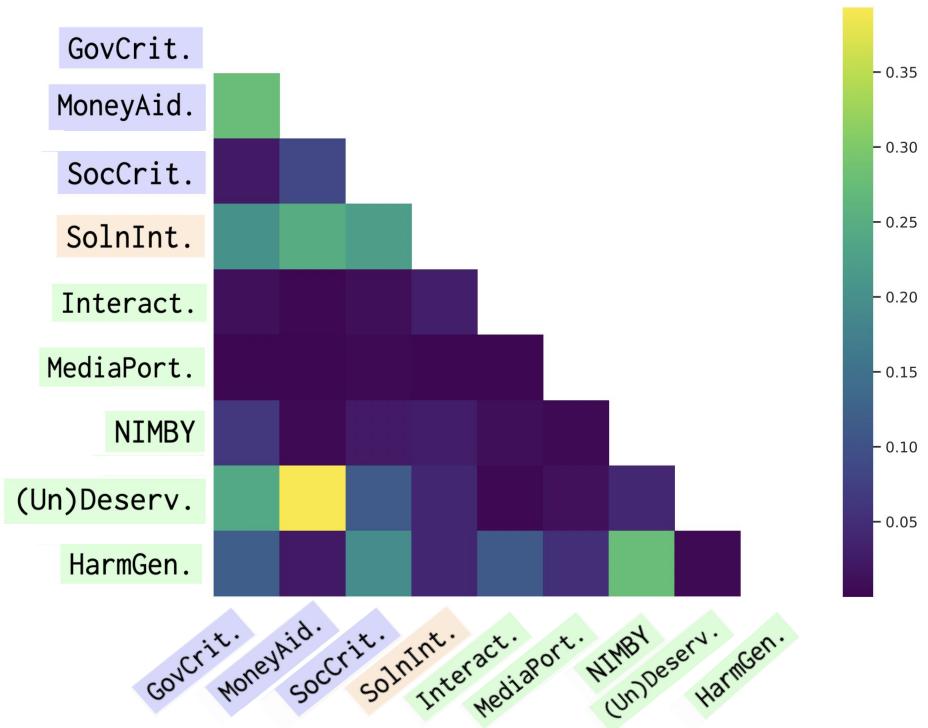


Figure 11: **Co-occurrence between Issue-specific frames in our set of 2.4M annotated posts.**

Post	Expert Annotations
'we used to have a fella come to our late night hockey games, sit in the stands and knit, one night we asked what he was knitting any why he chose a hockey arena, his response was he enjoyed watching the ladies play, and knitting scarfs and hats for the homeless was his goal.'	SolnInt.
'yes, there are rich people who own companies like tesla, who employ thousands of people who are a plus for our economy. they work, they get paid, they buy houses, cars, groceries, become citizens of law abiding communities, etc. they feel good about themselves. no jobs, homeless.'	SolnInt., MoneyAid.
'borders were legally open (only need single lane), business closed cuz they were told to (one place tried to stay open & got harassed by police, homeless were fed & the truckers were the ones harassed, egged & called names.'	(Un)Deserv.
'considerable damage has already been done this isn't just for pain medication too bc so many black, brown, trans, queer poor, disabled (lol), homeless, immigrant, addicts, mentally ill, and beyond (most people in general) don't get believed by doctors and get pawned off'	HarmGen., SocCrit.
'what about our rights to a secure border! those laws were passed by a democratic process. yet you violate your oath of office and give aid and comfort to those that break our laws. use funds to aid them, that could give our citizens m4all! look at the homeless in your state!!!'	GovCrit., (Un)Deserv., SolnInt., MoneyAid.
'now, shall we show the millions of refugees russia created? tens of millions of homeless? yeah. russia staged a photo op where everything is fine while mariupol of half a mil has been destroyed completely by russia. good job on the ethical balance. murderers can be friendly.'	GovCrit.
'your not going to do anything, what makes you think most the homeless want to work, pay rent, pay taxes, most can't maintain long enough to hold or get a job. can't do drugs on the job. they would rather sit in a tent getting high. she loves her meth to much.'	HarmGen.
'i have a couple of ideas for the elderly homeless, and the homeless vets. too radical for most people, but, you might be interested to hear them.'	SolnInt.
'im not saying he is guilt or anything, i hanged around a lot of homeless people when i was still in high school, they are an odd bunch of people, they had public next to their sleeping, however no one used it everyone was shiting everywhere but inside the toilet'	HarmGen., Interact.
'covid; mislead you, talk down the seriousness of the pandemic or letting it spread, gop, top 1%, and their lobbyist theory was; less people more money can go around, believing it would kill most of the homeless or unhealthy people that are a burden to the healthcare sector.'	HarmGen., MoneyAid., GovCrit., (Un)Deserv., SolnInt.
'my impression is that homelessness is highly comorbid with psychiatric illness, so i would *ideally* like to see the homeless being fed as if they were being treated for metabolic syndrome. (i'm actually involved in such an initiative, but we're temporarily on hold.)'	HarmGen., Interact., SolnInt.
'that guy his name is joey. he was part of an organization called center for political innovation, headed by a clown named caleb maupin but apparently lost all of his money, and became homeless, but still defense caleb till the end.'	Interact., GovCrit.
'homeless to the shelters, edp's to the mental hospitals.'	SolnInt.
'why in the world are you giving homeless folks \$12,000 a piece? do denver voters know you are doing this? you are asking for drug addicts and drunks on the streets making the city look like a third-world country. another disgraceful democrat mayor.'	HarmGen., MoneyAid., GovCrit., NIMBY
'man, that's some epic rambling bullsh** right there. what is it that you idolize about that guy? i mean, seriously, you could get the same level of coherence out of a random meth-head from the homeless shelter.'	HarmGen., SocCrit.
'here we are - back to people proving who they are. should homeless people get a chance to vote? people who ran away from home and don't have a birth certificate? why can you buy a gun w/o an id, but you need one to vote? it's because the gop wants guns, but don't want poc voting.'	HarmGen., GovCrit., SolnInt.
'seriously?? he's saying the city is spending \$170 million on bike lanes, while the homeless count soars. \$170m is better spent pretty much on anything else.'	MoneyAid.
'and the people who still have to work, even if they get heatstroke? the homeless? those in poor housing? not everyone has the freedom to defend themselves against extreme heat - or a sociopath government and public discourse.'	SocCrit., GovCrit., SolnInt.
'i literally can't stand when ppl record themselves doing something for someone in need, like giving food to homeless ppl. it isn't ur situation to profit off of my luv'	SocCrit.

Table 13: Examples of expert annotated posts with OATH-Frames



Figure 12: Word Clouds for posts labeled with each Issue-specific frame in subset of 2.4M posts.

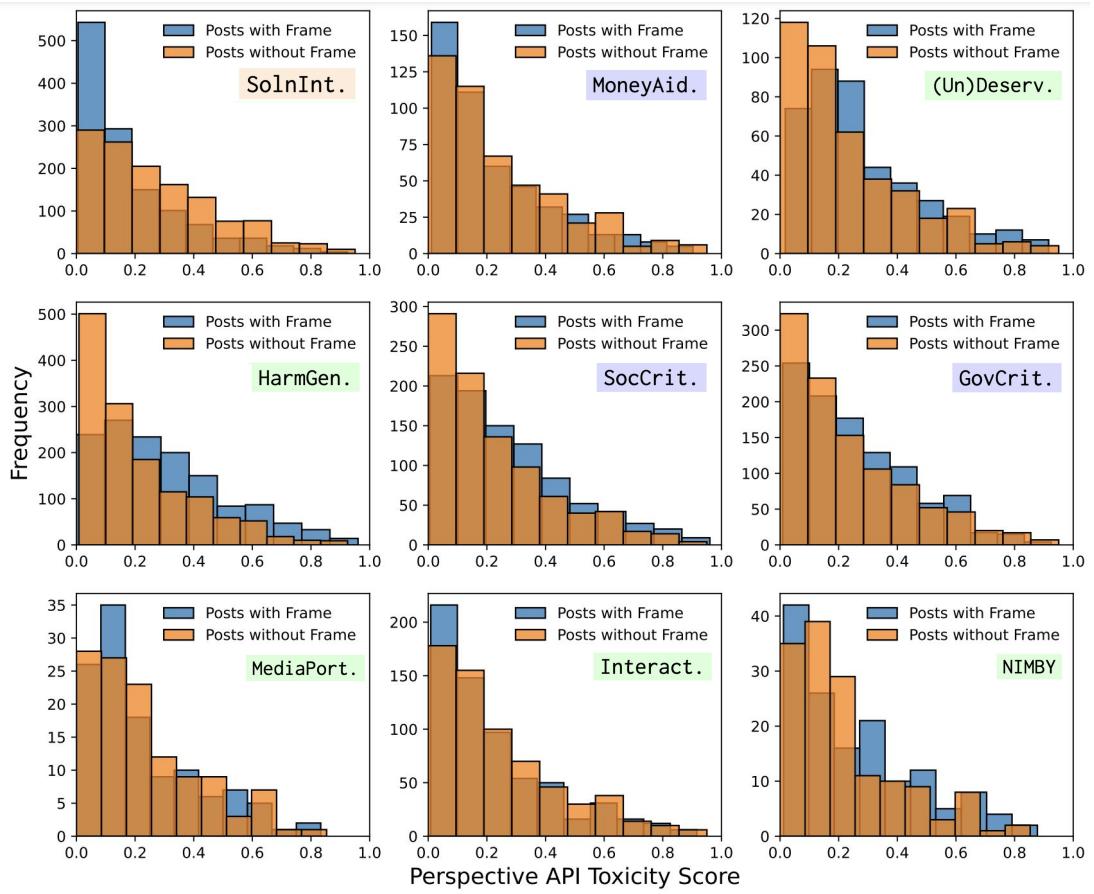


Figure 13: **Distribution of toxicity scores from Perspective API toxicity classifier for posts labeled with each Issue-specific frame in our subset of 4.1k human-annotated posts.**

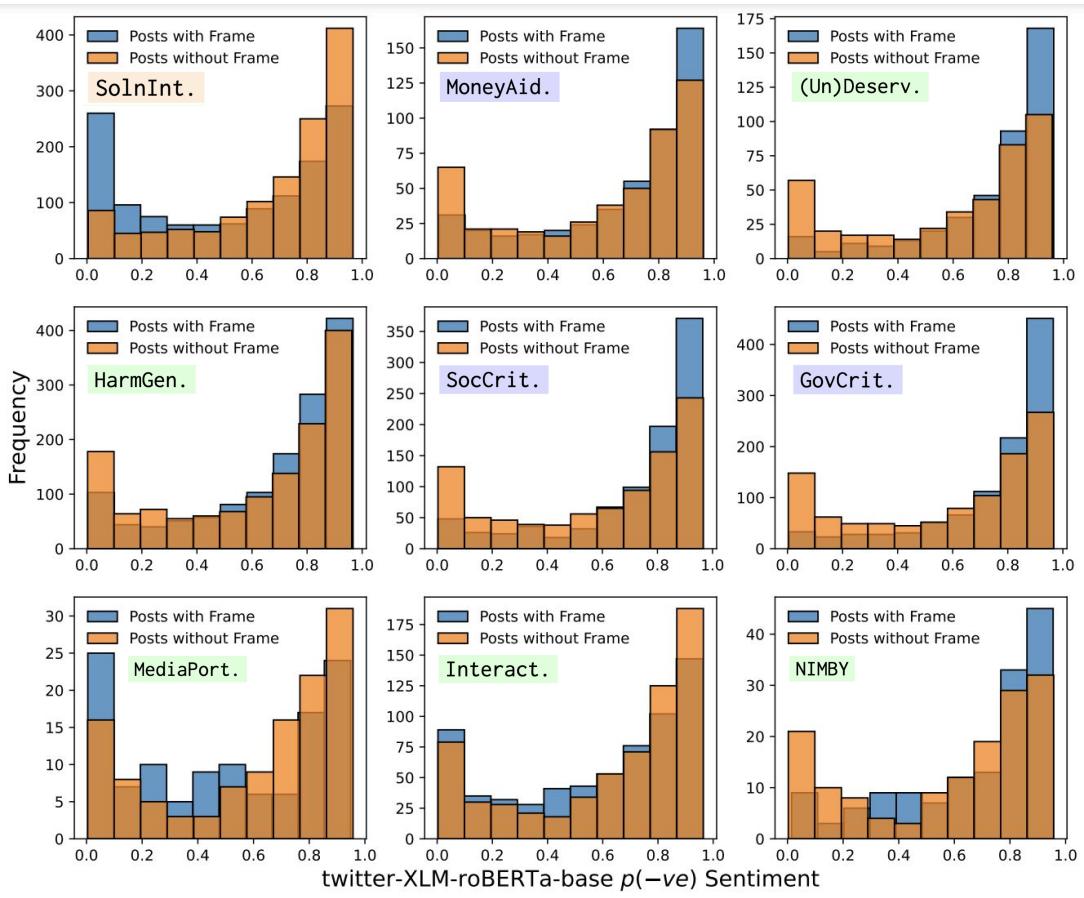


Figure 14: Distribution of negative sentiment probabilities for twitter-XLM-roBERTa-base model for sentiment analysis (Rosenthal et al., 2017) for posts labeled with each Issue-specific frame in our subset of 4.1k human-annotated posts.



Figure 15: Proportion of Issue-specific frames per state for the most frequently occurring state mentions in our posts.

Instructions (click to expand/collapse)

Guidelines (click to expand/collapse)

Examples (click to expand/collapse)

Tweet: "@1893870664441304866 well, he wants to house the homeless at twitter hq. haha"

Select one or more labels for each tweet.

Core Categories

ATOPICAL

needs_context alternative_meaning

PERSONAL ACCOUNTS

person_with_lived_experience

⚠ Did you select a label from one of the above categories ... if so, stop here and submit!

CRITIQUE

government_critique money_aid_resource_allocation partisan_politics made_an_example societal_critique performative_activism

REMEDIES

civic_engagement personal_interaction_observation_of_homelessness call_to_action violence_against_PEH solutions_interventions
 media_portrayal

PERCEPTIONS OF PEH

anti_immigration not_in_my_backyard nationalistic deserving undeserving_of_resources prejudice dehumanizing
 harmful_generalization

PROPOSED CATEGORIES: if you did not select a label from one of the above categories, propose a new label, ⚡ use sparingly!

Target Group (if applicable)

Target Group ▾

Figure 16: **Interface to collect human annotations:** Each annotator is asked to select labels that apply the most to a given post. The time taken for each annotation is also recorded in the backend. This interface details an earlier iteration of OATH-Frames that included a larger number of Issue-specific frames.