

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

Anonymous ACL submission

Abstract

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

Homelessness in the U.S. is widespread, eliciting complex attitudes among individuals (e.g. critical as well as sympathetic), often expressed on social media. These attitudes are challenging to summarize at scale, further obfuscating the broader public opinion which is imperative for guiding 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, 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, 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 F1 difference in annotation performance. Our effort results in a collection of 8K social media posts labeled by domain 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 3.1M posts on homelessness. We find marked differences in perceptions towards homelessness between the east and west coast of the U.S. We also find that posts often pit people experiencing homelessness, specifically veterans, against immigrants and asylum seekers as being either more- or less-deserving of resources and 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 experienced homelessness in 2023.¹ As a result, home-

lessness intersects with U.S. socio-economic life in a myriad complex ways, eliciting a diverse spectrum of attitudes among the populace; these attitudes differ across socio-political factors such as location, ideology and economic climate. Social media serves as a vehicle for the general public to express these attitudes and opinions on the crisis of homelessness. However, these attitudes and opinions are hard to characterize given both the complexity as well as the scale of such conversations in social media. Prior work from the fields of sociology and social work have studied discourse about homelessness grounded in ethnographic studies and qualitative surveys, conducted with few individuals (Kim et al., 2013; Tsai et al., 2019; Kingree and Daves, 1997). 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 can be broadly categorized into Critiques, Perceptions and Responses to people experiencing homelessness (PEH), and further subdivided into nine finer-grained, Issue-specific frames, developed using framing theory (Entman, 1993) and grounded theory (Glaser and Strauss, 1967) from social science. We summarize our typology and corresponding definitions in Table 1.

Using our OATH-Frames, we annotate a corpus of 4K posts from X with the help of domain and trained experts; we find that most posts contain multiple frames across the three broad categories. Further, building on the generative successes of large language models, we employ a GPT-4 annotation assistant to ease the annotation process by performing a first round of annotations to be verified by the experts. This GPT-4 + Expert

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

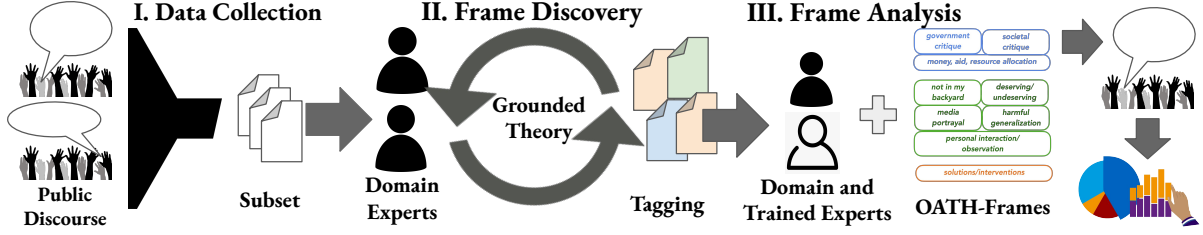


Figure 1: **Data Collection and Framing:** We collect 3.1M posts on the topic of homelessness. The process of framing consists of two main components: *Frame Discovery* and *Frame Analysis*. Given a subset of our posts, domain experts apply grounded theory to iteratively extract the main themes in our data, and develop a set of frames that describe public attitudes towards homelessness: OATH-Frames. During, *Frame Analysis* we train two additional experts to annotate our posts with our frames and we annotate a corpus of 4k posts with our team of experts.

approach results in a $6.5\times$ speedup in annotation, while incurring only a 2 point F1 reduction in annotation accuracy—an attractive tradeoff given the complexity of the multilabel annotation task. Using this strategy, we further annotate 4K posts; overall our corpus is comprised of 8K expert annotated posts collected with and without GPT-4 assistance.

Finally, we use our collected data to build a model that can predict OATH-Frames on a large set of 3.1M posts that help us analyse social media conversations on homelessness at scale. Moreover, posts from western U.S. states express more **Perceptions**, perhaps due to greater PEH visibility and those from eastern U.S. states express more attitudes that pit PEH against immigrants, refugees, and asylum seekers.

Our main contributions include:

1. **OATH-Frames:** We introduce a framing typology grounded in communications and sociology theory (Entman, 1993; Glaser and Strauss, 1967), encapsulating **Online Attitudes Towards Homelessness**.
2. **Corpus and Models:** We create a corpus of 8K posts annotated by experts, with and without GPT-4 assistance. We train a Flan-T5-Large model on this corpus, which we use to detect OATH-Frames on a large collection of 3.1M posts.
3. **Analysis of Social Media Discourse on Homelessness:** We find large variations in (predicted) attitudes on homelessness across socio-political dimensions such as regionality, and minority identities.

Through our contributions, we hope to aid fast and scalable summarization of public opinion on homelessness expressed in online discourse. We hope this may serve as a tool for advocacy groups and political constituents in tracking effectiveness of policies and targeting relief efforts. We will

release our data and models publicly.

2 Pragmatic Frames for Attitudes towards PEH: OATH-Frames

We propose a novel typology of pragmatic frames for capturing and characterizing societal attitudes towards homelessness and persons experiencing homelessness (PEH). We apply concepts from framing theory (Entman, 1993), which serves to emphasize certain aspects in an issue’s framing that influence broader public opinion. For instance, framing PEH with attributes such as 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. In the context of homelessness discourse on social media, prior work has focused on the framing of stigma towards PEH (Kim et al., 2013). In our work, we aim to characterize broader societal attitudes on homelessness, going beyond stigma.

Following Boydston and Gross (2013), our framing process follows two steps: frame discovery (§2.1) and frame analysis (§2.2) as detailed 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**). Frame analysis entails annotating documents with the discovered frames and post-hoc analysis.

Given the diversity (e.g., socio-economic and political) of users on Twitter²(Hu et al., 2021), we use posts from this platform to study societal attitudes towards homelessness. We use a keyword search

²Now, X: <https://x.com>.

GovCrit.	Government Critique: Criticism about government body, policies and laws including discussion of homelessness through the lens of political parties and values. Example: 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 Also labeled with: MoneyAid. , (Un)Deserv.
MoneyAid.	Money Aid Resource 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. Example: 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? Also labeled with: 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. Example: 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. Also labeled with: Interact. , SolnInt.
SolnInt.	Solutions and Interventions: Discussion of solutions, interventions, charitable acts and remedies to address the homelessness crisis. Example: 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 Also labeled with: HarmGen.
Interact.	Personal Interaction: Anecdote describing a direct personal exchange with PEH. Example: 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. Also labeled with: 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). Example: is the national news reporting on the homeless on the streets and encampments as well as random crime exaggerated? beautiful city and location. Also labeled with: NIMBY
NIMBY	Not in My Backyard: Characterization of opposition by residents to proposed developments in their local area, as well as support for strict land use regulations, specifically against wanting to see PEH in their local area. Example: funny you guys cleaned everything up and moved all of the illegal aliens that were homeless on the streets of el paso before you visited . . . Also labeled with: GovCrit. , HarmGen.
(Un)Deserv.	Deserving and Undeserving of Resources: Perpetuating a hierarchy of which PEH with other marginalized communities that are more or less deserving of aid. Includes nationalistic rhetoric. Example: 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 Also labeled with: 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. Example: 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

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 . Each post can be associated with not one, but multiple frames, across different overarching themes.

on the X API to randomly sample posts containing the seed token “homeless”; this results in a collection 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

Our process of discovering frames involved four trained domain experts independently annotating a random sample of 100 posts each from our 3.1M collection followed by a discussion until consen-

sus was reached. This process is repeated on three different subsets of the posts until reaching thematic saturation at 9 Issue-specific frames with 3 overarching themes, loosely following prior work (Baiocchi and Arguello, 2019; Kim et al., 2013; Gowan, 2010), we find three overarching themes. First, we found 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 found posts reflecting different Perceptions of

PEH, such as posts which attribute undesirable characteristics to PEH to make harmful generalizations. Third, we found 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 categories, 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 filtered out ambiguous posts as well as 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. 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 frame analysis (Boydston and Gross, 2013), we build a team of four domain experts and two trained annotators. Our team labels 5K posts, of which 1.2K posts are filtered out. In this set, 280 posts are multiply annotated where we achieve an inter-annotator Fleiss κ agreement of 0.47 across the 9 Issue-specific frames (Fleiss, 1971).³ We detail our inter-annotator agreement for each coder in Figure 2 on a set of 280 multiply-annotated posts.

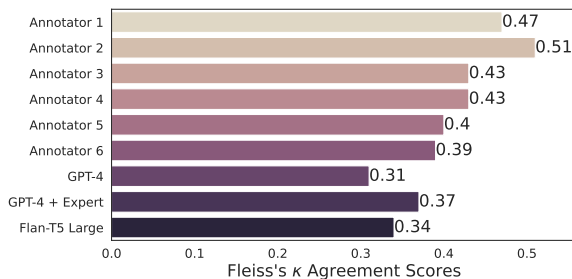


Figure 2: **Inter-annotator Agreement Scores:** We plot the Fleiss' κ agreement scores between all of the annotators, along with GPT-4, GPT-4 + Expert and Flan-T5-Large, w.r.t all of the other annotators.

We find that each post may contain multiple

³Our 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.

Issue-specific frames, even across our three broad categories where frames co-occurring with each other can represent a multitude of attitudes in a single post.

Figure 3 details the distribution of Issue-specific frames in our set of 4k expert-annotated posts where we find that GovCrit., SolnInt. and HarmGen. to be the most common Issue-specific frames and the most salient theme is Critiques.

Even in the team of trained and domain experts, this task was extremely laborious and time consuming as our posts contain rhetoric which was challenging to process. Given the expense and quality issues associated with crowdsourcing (Douglas et al., 2023), we next explore employing LLMs as annotation assistants to alleviate the pressures of manual annotation.

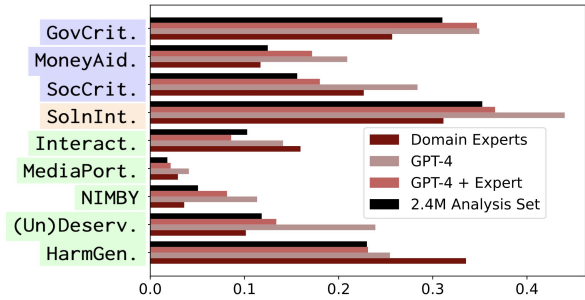


Figure 3: **Proportion of Issue-specific frames:** We show that the proportion of Issue-specific frames in our data splits and observe that the distribution of frames across splits is similar.

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). However, highly subjective tasks necessitate a greater diversity in annotations and LLMs may not always be able to reliably reason about socially complex tasks. Pangakis et al. (2023) propose a pipeline to use LLMs for annotating large amounts of data and they posit that LLM generated annotations should always be validated with a set of high quality expert annotations as performance is dependent on the nature of the task. Following this prior work, 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

Annotators →	1st Author	GPT-4	GPT-4 + Expert	Flan-T5-Large
Issue-specific ↓	F1. Score (↑)			
GovCrit.	0.68 ± 0.04	0.70 ± 0.08	0.72 ± 0.07	0.65 ± 0.07
MoneyAid.	0.57 ± 0.17	0.58 ± 0.14	0.61 ± 0.15	0.59 ± 0.18
SocCrit.	0.37 ± 0.12	0.46 ± 0.06	0.50 ± 0.05	0.42 ± 0.09
SolnInt.	0.67 ± 0.06	0.62 ± 0.06	0.65 ± 0.06	0.68 ± 0.03
Interact.	0.36 ± 0.08	0.45 ± 0.07	0.52 ± 0.07	0.54 ± 0.11
MediaPort.	0.70 ± 0.07	0.50 ± 0.08	0.63 ± 0.11	0.69 ± 0.10
NIMBY	0.28 ± 0.20	0.30 ± 0.16	0.39 ± 0.21	0.34 ± 0.26
(Un)Deserv.	0.44 ± 0.07	0.46 ± 0.06	0.63 ± 0.08	0.67 ± 0.01
HarmGen.	0.60 ± 0.04	0.49 ± 0.06	0.51 ± 0.06	0.55 ± 0.05
Aggregated	0.54 ± 0.01	0.46 ± 0.04	0.52 ± 0.04	0.51 ± 0.03

Table 2: We report F1. scores for each Issue-specific frame, and macro F1. scores across all 9 Issue-specific frames (Aggregated) where we compute F1 scores with all 6 annotators in our team of experts. 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 labels in comparison to domain experts.

annotations on their own.

3.1 Predicting OATH-Frames with GPT-4

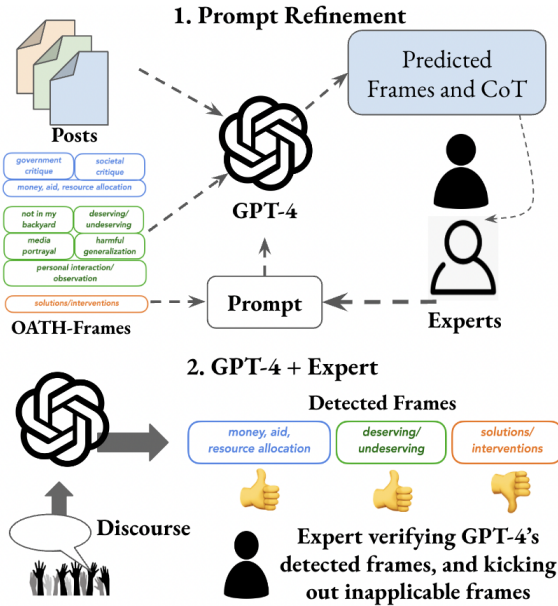


Figure 4: **GPT-4 + Expert Annotation Pipeline:** We prompt GPT-4 with OATH-Frames and their corresponding definitions where we use chains-of-thought to clarify the definitions in our prompt and validate against a set of expert annotated posts following prior works (Pangakis et al., 2023). Using our refined prompt, we generate predictions of OATH-Frames on a set of 4k posts and experts verify the predictions while kicking 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. Filtering posts out results in a dataset of 4k posts annotated by GPT-4. The second phase consists of our Issue-specific frames detection where we prompt GPT-4 to perform a multilabel classification task for each post using our frame definitions. Examples of our prompt are provided in Appendix B. We evaluate the performance of GPT-4 as an annotator using the pipeline outlined in Figure 4. The cycle of prompt refinement involves clarifying definitions and ambiguities for the incorrect classifications and measuring performance against our set of domain expert annotations until we reach a desirable performance. We use chain-of-thought (CoT) prompting (Wei et al., 2022) to better understand the kinds of errors GPT-4 makes and to guide the definitions of the frames in our prompt.

We find that GPT-4 achieves a precision of 0.39 ± 0.05 and recall of 0.64 ± 0.05 across all 9 Issue-specific frames 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. GPT-4 often confuses 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, it is important to resolve such ambiguities with post hoc human validation. Due to the interpretive nature of our frames, it is difficult for GPT-4 to capture all the nuances of the task and it’s necessary for a human to verify the predictions post generation.

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. During all our annotation tasks, we record the time taken by the annotators to annotate each post, both individually and with the GPT-4 + Expert approach. After removing outliers, we note that by themselves, annotators take 187.49 (± 30.71) seconds, and 28.80 (± 8.55) seconds with GPT-4 + Expert, which is a $6.5\times$ boost in annotation speed and results in a 6 point increase in F1 score over GPT-4 as an annotator.

GPT as an Annotator: Error Analysis Our GPT-4 + Expert experiments are designed such that domain experts kick out labels that GPT-4 misclassified. We include Table 8 in the Appendix which outlines frames that have the highest false discovery rate where (i) GPT-4 is not able to reflect the attitude in our frames in alignment with our frame definitions, and (ii) inherent subjectivity and interpretation of the post warrants misclassifications. A salient attitude in discourse on homelessness is pitting PEH against immigrants and putting them in competition for aid and resources. However, we find that GPT-4 is not able to capture this juxtaposition between PEH and other groups as noted in the examples in Table 8. Furthermore, the shorter length of posts 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.

4 Large-scale Frame Analysis

4.1 Training Frame Prediction Models

Combining our expert-annotated posts with those obtained using GPT-4 + Expert (§3), we obtain a total of 10.4K labeled 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 9 Issue-specific frames. Further details of the dataset split, model training and ablations with Flan-T5 sizes are provided in Appendix A. Results in Table 2 show that Flan-T5-Large performs better than GPT-4, and is at par with the GPT-4 + Expert setting on our multiply-annotated agreement set. Using our trained Flan-T5-Large model, we then generate Issue-specific frame predictions for our corpus of 3.1M posts. After using the Public Opinion filter, we obtain a total of 2.4M posts annotated with our Issue-specific frame predictions.

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

Understanding public opinion via online discourse can serve as a useful tool for advocacy and policy groups, as attitudes concerning policy and relief efforts can inform services and interventions. One potential use for our framework is for advocacy organizations who straddle the often-competing interests of community stakeholders, neighborhood residents, and PEH. Synthesizing public attitudes can inform more strategic advocacy efforts for bringing homeless services and more equitable housing policies to local communities. Our framing typology helps investigate how attitudes towards homelessness vary across socio-cultural dimensions. We demonstrate the utility of our frames in characterizing attitudes in discourse on social media across geographical locations and the sub-populations of veterans, immigrants and refugees using our dataset of 3.1M posts annotated with OATH-Frames.

Visibility of Homelessness and Regionality. We investigate the saliency of our Issue-specific frames in posts that mention states in the west and the east.⁴ Discourse on homelessness is closely tied

⁴Our 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.

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 the New York which has the second highest rate of overall homelessness but 95% of the population is sheltered (Sousa et al., 2023). We hypothesize that greater visibility of PEH in the west coast due to a higher unsheltered population and as a result, public attitudes towards homelessness to be more centered on **Perceptions**. New York specifically experienced an influx of asylum seekers in 2023 (Sousa et al., 2023) and thus posts mentioning east coast states are more likely to discuss competition for jobs and resources between PEH and undocumented immigrants. We segment our dataset by posts mentioning these east and west coast states using NER⁵. We find that **HarmGen.** is the most salient Issue-specific frame in posts mentioning mentioning west coast states and **(Un)Deserv.** is the most prevalent 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 be more prone to express perceptions of PEH, whereas posts mentioning east coast states reflect discourse that pits PEH against asylum seekers for resources. 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. 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 and PEH being in competition with each other for shelter.

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 (Kirsten Moore Sheeley and Yaroslavsky, 2021). To this end, we ex-

⁵hf-reddit-ner-place-names: https://huggingface.co/cjber/reddit-ner-place_names.

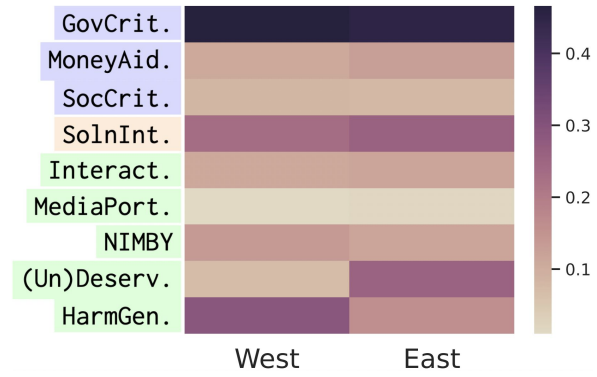


Figure 5: **Proportion of Issue-specific frames per Region in the U.S.:** **(Un)Deserv.** and **HarmGen.** are more significantly present in posts mentioning east and west coast states respectively ($p < 0.0001$) where we hypothesize that differing **Perceptions** may be due to PEH visibility in those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

plore 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.

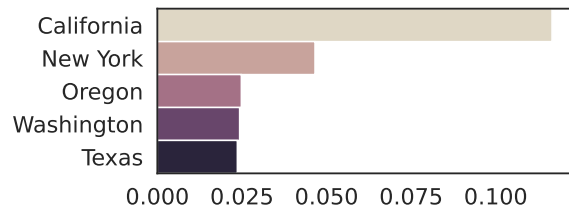


Figure 6: **State-wise proportion of posts tagged with NIMBY frame:** We plot the top 5 states that are represented in the 2.4M Analysis Split, and for each of them, plot the proportion of **NIMBY** frame occurrences.

PEH Identity Analysis Discourse on immigrants and foreign policy often elicits nationalistic rhetoric where people discuss homelessness in relation to competing issues that are more or less deserving of aid. One of the most prevalent themes in our analysis is how PEH are pitted against other marginalized groups and put in competition for resources where their identity is used as a justification for being more or less deserving of aid. We segment our dataset by posts that contain the identity terms: veteran, immigrant, refugee, and asylum seeker. We compare the prevalence of our Issue-specific frames in posts that contain the identity term in a subset of 172k posts compared to a random sample of posts from our dataset. As illustrated in Figure 7, we find that the Issue-specific frames of **(Un)Deserv.**, **GovCrit.**, and **MoneyAid.**, are

statistically more prevalent in our sample of posts that contain the identity terms.

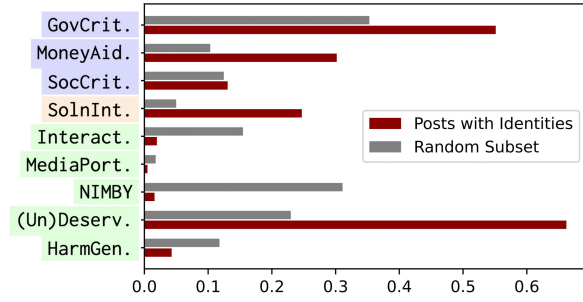


Figure 7: **Proportion of Issue-specific frames w.r.t posts which contain identity groups:** We plot the proportion of Issue-specific frames for posts containing mentions of identity groups (veterans, immigrants, refugees and asylum seekers), where (Un)Deserv., GovCrit. and MoneyAid. ($p < 0.0001$) are the most salient attitudes.

5 Related Work

Characterizing Attitudes Towards Homelessness Prior work characterizes public opinion and attitudes towards homelessness through ethnographic studies, surveys and psychometric analyses on smaller scale samples (Gowan, 2010; Tsai et al., 2019; Kingree and Daves, 1997; Kim et al., 2013). The Attitudes Towards Homelessness Inventory (ATHI) inventory categorizes attitudes around personal characteristics, societal causes, affiliation and solvable problems (Kingree and Daves, 1997) while a survey of U.S. adults assesses attitudes and beliefs about causes of homelessness where responses are collected from a sample of 541 adults from 47 US states (Tsai et al., 2019). (Baiocchi and Arguello, 2019) identifies a basic topology from ethnographic studies to identify key narratives in discourse about homelessness: sin-talk, sick-talk, and system-talk. Kim et al. (2013) characterize stigma against PEH using social media posts by manually annotating 6400 posts. It is difficult to understand the complexities of attitudes towards homelessness at scale from qualitative studies alone. To this end, we plan to expand these studies by performing large-scale analyses on social media posts annotated with OATH-Frames.

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 (Upton, 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). We build on prior works 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 works 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.

6 Conclusion

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 identities. We expect our framework to be the first step towards characterizing, summarizing and understanding complex societal attitudes on intricate socio-economic issues *at scale*.

Limitations and Ethical Considerations

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.

United States Focus

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.

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.

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.

Collaborating with domain experts, advocating for people experiencing homelessness, bring their understanding of institutions and their ever-changing practices, policies, incidents, and shared discipline-specific vocabulary to speak with depth about what we understand is being said in the post. Each annotator brings their own identity and priors coupled with professional knowledge into the language interpretation process which forms our positionality as individual researchers.

References

- 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.
- Amber E. Boydston and Justin H. Gross. 2013. [Identifying media frames and frame dynamics within and across policy issues](#).
- Benjamin D Douglas, Patrick J Ewell, and Markus Brauer. 2023. Data quality in online human-subjects research: Comparisons between mturk, prolific, cloudresearch, qualtrics, and sona. *Plos one*, 18(3):e0279720.
- 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.
- T. Gowan. 2010. *Hobos, Hustlers, and Backsliders: Homeless in San Francisco*. University of Minnesota Press.

Minda Hu, Ashwin Rao, Mayank Kejriwal, and Kristina Lerman. 2021. Socioeconomic correlates of anti-science attitudes in the us. *Future Internet*, 13(6):160.

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.

Andrew Klein Jessica Richards Fernanda Jahn Verri Marques Vestal Kirsten Moore Sheeley, Alisa Belinkoff Katz and Zev Yaroslavsky. 2021. [The making of a crisis: A history of homelessness in los angeles](#).

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*.

Julia Mendelsohn, Ceren Budak, and David Jurgens. 2021. Modeling framing in immigration discourse on social media. *arXiv preprint arXiv:2104.06443*.

Nicholas Pangakis, Samuel Wolken, and Neil Fasching. 2023. Automated annotation with generative ai requires validation. *arXiv preprint arXiv:2306.00176*.

Christopher Michael Rytting, Taylor Sorensen, Lisa Argyile, Ethan Busby, Nancy Fulda, Joshua Gubler, and David Wingate. 2023. Towards coding social science datasets with language models. *arXiv preprint arXiv:2306.02177*.

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.

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 criminalization?: Discourses on homelessness and social responses*. Ph.D. thesis.

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.

A Frame Prediction: Dataset Statistics and Model Training Details

Combining domain-expert annotated Issue-specific frames in §2, along with posts that are labelled 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 HuggingFace⁶ 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.

We demonstrate the data splits, training details and results of different Flan-T5 variations in [Table 3](#), [Table 4](#), [Table 5](#) and [Table 6](#).

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

Table 3: Dataset Splits for training Flan-T5-Large

⁶https://huggingface.co/docs/transformers/model_doc/flan-t5

Config	Assignment
models	Flan-T5-Large Number of parameters: 770M
	Flan-T5-Base Number of parameters: 250M
	Flan-T5-Small Number of parameters: 80M
train batch size	32
eval batch size	16
seed	0
max epochs	25
learning rate	3e-5
learning scheduler	fixed
GPU	Quadro RTX 8000
# of GPUs	8 (training), 20 (inference)
Training time	8 hours
Stopping Criteria	Early Stopping on Validation Loss

Table 4: Model Training Details

Frame →	Issue-specific			
Setting (vs. Annotators) ↓	Prec.	Rec.	F1	Agg.
1st Author	0.6 ± 0.06	0.53 ± 0.02	0.54 ± 0.01	0.47 ± 0.02
GPT-4	0.39 ± 0.05	0.64 ± 0.05	0.46 ± 0.04	0.31 ± 0.04
GPT-4 + Expert	0.51 ± 0.07	0.6 ± 0.05	0.52 ± 0.04	0.37 ± 0.04
Flan-T5-Large	0.56	0.48	0.51	0.34 ± 0.03
Flan-T5-Base	0.54	0.49	0.51	0.31 ± 0.02
Flan-T5-Small	0.53	0.4	0.42	0.24 ± 0.01

Table 5: Comparing variations of Flan-T5 with other settings

B GPT-4 Prompt

We list the final prompt that we achieve after iterating through the OATH-Frames instructions, for obtaining GPT-4 annotations (Table 7).

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 frames for each post.

Frame →	Issue-specific		
Setting (vs. Annotators) ↓	Prec.	Rec.	F1
Flan-T5-Large	0.50	0.48	0.50
Flan-T5-Base	0.48	0.47	0.47
Flan-T5-Small	0.46	0.41	0.42

Table 6: Comparing variations of Flan-T5 on a larger test set

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 8: **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.

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

category	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> <hr/> <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> <hr/> <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> <hr/> <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> <hr/> <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 8: **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.

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: Macro-averaged Precision, Recall and F1. for each of the Issue-specific frames