

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-

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

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

041
042
043
044
045
046
047
048
049
050
051
052
053
054
055
056
057
058
059
060
061
062
063
064
065
066
067
068
069
070
071
072
073
074
075
076
077
078
079
080

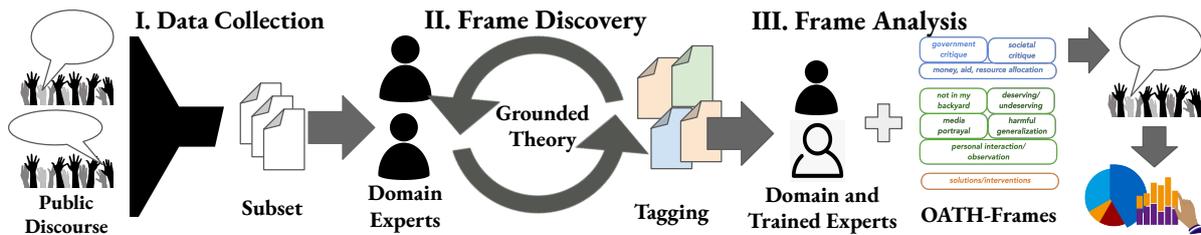


Figure 1: **Data Collection and Framing:** We collect 3.1M posts on the topic of homelessness. The process of framing consists of two main components: *Frame Discovery* and *Frame Analysis*. Given a subset of our posts, domain experts apply grounded theory to iteratively extract the main themes in our 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.

081 approach results in a $6.5\times$ speedup in annotation, 082 while incurring only a 2 point F1 reduction in an- 083 notation accuracy—an attractive tradeoff given the 084 complexity of the multilabel annotation task. Using 085 this strategy, we further annotate 4K posts; overall 086 our corpus is comprised of 8K expert annotated 087 posts collected with and without GPT-4 assistance.

088 Finally, we use our collected data to build a 089 model that can predict OATH-Frames on a large 090 set of 3.1M posts that help us analyse social me- 091 dia conversations on homelessness at scale. More- 092 over, posts from western U.S. states express more 093 **Perceptions**, perhaps due to greater PEH visibil- 094 ity and those from eastern U.S. states express more 095 attitudes that pit PEH against immigrants, refugees, 096 and asylum seekers.

097 Our main contributions include:

- 098 1. **OATH-Frames:** We introduce a framing ty- 099 pology grounded in communications and 100 sociology theory (Entman, 1993; Glaser 101 and Strauss, 1967), encapsulating **Online 102 Attitudes Towards Homelessness**.
- 103 2. **Corpus and Models:** We create a cor- 104 pus of 8K posts annotated by experts, with 105 and without GPT-4 assistance. We train a 106 Flan-T5-Large model on this corpus, which 107 we use to detect OATH-Frames on a large 108 collection of 3.1M posts.
- 109 3. **Analysis of Social Media Discourse on 110 Homelessness:** We find large variations in 111 (predicted) attitudes on homelessness across 112 socio-political dimensions such as regionality, 113 and minority identities.

114 Through our contributions, we hope to aid fast 115 and scalable summarization of public opinion on 116 homelessness expressed in online discourse. We 117 hope this may serve as a tool for advocacy groups 118 and political constituents in tracking effectiveness 119 of policies and targeting relief efforts. We will

120 release our data and models publicly.

121 2 Pragmatic Frames for Attitudes 122 towards PEH: OATH-Frames

123 We propose a novel typology of pragmatic frames 124 for capturing and characterizing societal attitudes 125 towards homelessness and persons experiencing 126 homelessness (PEH). We apply concepts from 127 framing theory (Entman, 1993), which serves to 128 emphasize certain aspects in an issue’s framing 129 that influence broader public opinion. For instance, 130 framing PEH with attributes such as substance use 131 and crime can invoke feelings of fear and hostil- 132 ity amongst the public. This theory has formed 133 the basis of analyzing discourse on immigration, 134 (Mendelsohn et al., 2021), politics in media (Boyd- 135 stun and Gross, 2013), eating disorders (Arseniev- 136 Koehler et al., 2016) and so on. In the context of 137 homelessness discourse on social media, prior work 138 has focused on the framing of stigma towards PEH 139 (Kim et al., 2013). In our work, we aim to charac- 140 terize broader societal attitudes on homelessness, 141 going beyond stigma.

142 Following Boydston and Gross (2013), our fram- 143 ing process follows two steps: frame discovery 144 (§2.1) and frame analysis (§2.2) as detailed in Fig- 145 ure 1. Frame discovery uncovers salient themes 146 in a text corpus on a topic, resulting in a fram- 147 ing typology; in the context of our study, this re- 148 sults in OATH-Frames (Frames for **O**nline **A**ttitudes 149 **T**owards **H**omelessness). Frame analysis entails 150 annotating documents with the discovered frames 151 and post-hoc analysis.

152 Given the diversity (e.g., socio-economic and po- 153 litical) of users on Twitter²(Hu et al., 2021), we use 154 posts from this platform to study societal attitudes 155 towards homelessness. We use a keyword search

²Now, X: <https://x.com>.

<p>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.</p>
<p>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.</p>
<p>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.</p>
<p>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.</p>
<p>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.</p>
<p>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</p>
<p>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.</p>
<p>(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.</p>
<p>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</p>

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.

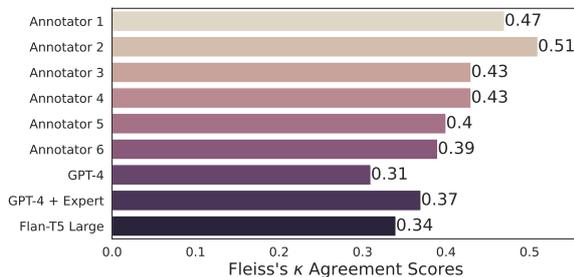


Figure 2: **Inter-annotator Agreement Scores:** We plot the Fleiss's κ agreement scores between all of the annotators, along with GPT-4, GPT-4 + Expert and Flan-T5-Large, w.r.t all of the other annotators.

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.

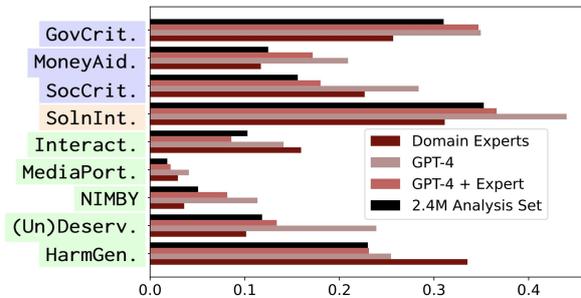


Figure 3: **Proportion of Issue-specific frames:** We show that the proportion of Issue-specific frames in our data splits and observe that the distribution of frames across splits is similar.

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.

247 annotations on their own.

248 3.1 Predicting OATH-Frames with GPT-4

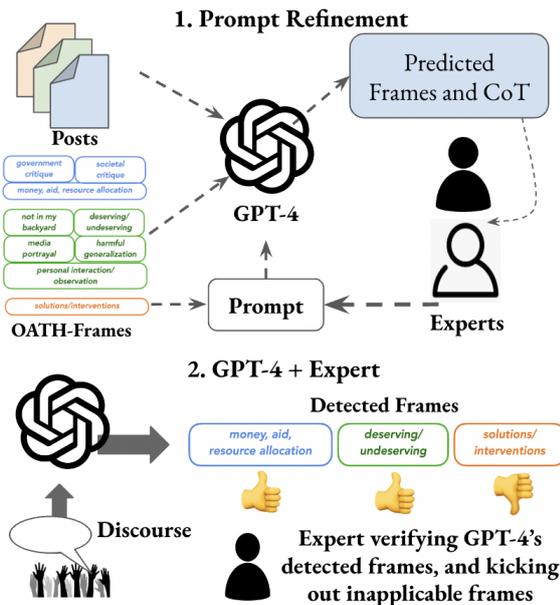


Figure 4: **GPT-4 + Expert Annotation Pipeline:** We prompt GPT-4 with OATH-Frames and their corresponding definitions where we use chains-of-thought to clarify the definitions in our prompt and validate against a set of expert annotated posts following prior works (Pangakis et al., 2023). Using our 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.

249 We prompt GPT-4 in a sequential manner that
 250 emulates our framing typology. Our prompting
 251 pipeline consists of two phases. In the first phase,
 252 we apply the Public Opinion frame to filter out

253 posts that do not contain a public attitude towards
 254 homelessness. Filtering posts out results in a
 255 dataset of 4k posts annotated by GPT-4. The second
 256 phase consists of our Issue-specific frames
 257 detection where we prompt GPT-4 to perform a
 258 multilabel classification task for each post using
 259 our frame definitions. Examples of our prompt are
 260 provided in Appendix B. We evaluate the performance
 261 of GPT-4 as an annotator using the pipeline
 262 outlined in Figure 4. The cycle of prompt refine-
 263 ment involves clarifying definitions and ambiguities
 264 for the incorrect classifications and measuring
 265 performance against our set of domain expert an-
 266 notations until we reach a desirable performance.
 267 We use chain-of-thought (CoT) prompting (Wei
 268 et al., 2022) to better understand the kinds of errors
 269 GPT-4 makes and to guide the definitions of the
 270 frames in our prompt.

271 We find that GPT-4 achieves a precision of 0.39
 272 ± 0.05 and recall of 0.64 ± 0.05 across all 9
 273 Issue-specific frames showing that it may have
 274 reasonable coverage of the predicted labels but is
 275 also producing false positives due to the inherent
 276 ambiguity in the posts that may be easier for a
 277 domain expert to resolve. GPT-4 often confuses
 278 SocCrit. and GovCrit. frames due to the ambi-
 279 guity in the target group of the post. When soci-
 280 ety or the government are not explicitly mentioned
 281 in the post, domain experts are able to use the con-
 282 text of the post to resolve the target group. For
 283 example, ‘@mention it appears the problem
 284 in el paso with illegal aliens were solved
 285 overnight, so that you could have a
 286 photo op without anyone seeing hundreds

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 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.

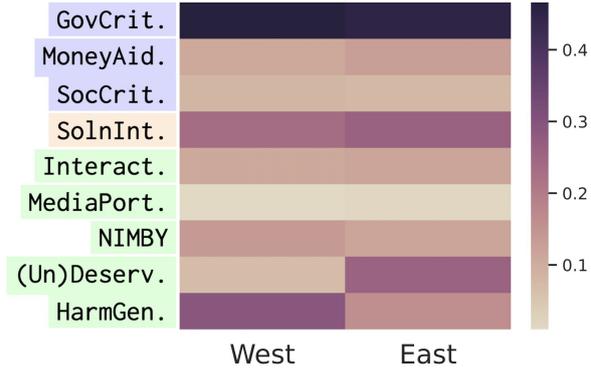


Figure 5: **Proportion of Issue-specific frames per Region in the U.S.:** (Un)Deserv. and HarmGen. are more significantly present in posts mentioning east and west coast states respectively ($p < 0.0001$) where we hypothesize that differing Perceptions may be due to PEH visibility in those regions.

430 plore the proportion of state representation for the
 431 Issue-specific frame of NIMBY where we find
 432 that California is represented proportionally higher
 than other states as illustrated in Figure 6.

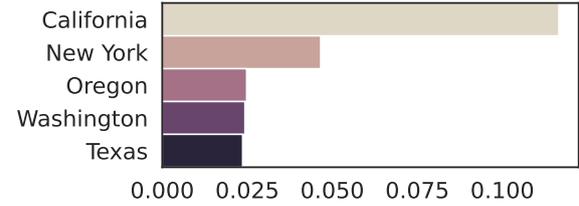


Figure 6: **State-wise proportion of posts tagged with NIMBY frame:** We plot the top 5 states that are represented in the 2.4M Analysis Split, and for each of them, plot the proportion of NIMBY frame occurrences.

433
 434 **PEH Identity Analysis** Discourse on immi-
 435 grants and foreign policy often elicits national-
 436 istic rhetoric where people discuss homelessness
 437 in relation to competing issues that are more or
 438 less deserving of aid. One of the most preva-
 439 lent themes in our analysis is how PEH are pit-
 440 ted against other marginalized groups and put in
 441 competition for resources where their identity is
 442 used as a justification for being more or less de-
 443 serving of aid. We segment our dataset by posts
 444 that contain the identity terms: veteran, immi-
 445 grant, refugee, and asylum seeker. We compare
 446 the prevalence of our Issue-specific frames in
 447 posts that contain the identity term in a subset
 448 of 172k posts compared to a random sample of
 449 posts from our dataset. As illustrated in Figure
 450 7, we find that the Issue-specific frames of
 451 (Un)Deserv., GovCrit., and MoneyAid., are

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

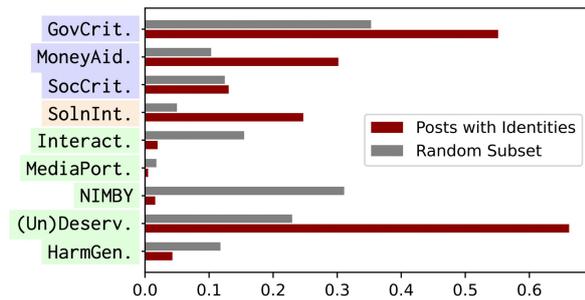


Figure 7: **Proportion of Issue-specific frames w.r.t posts which contain identity groups:** We plot the proportion of Issue-specific frames for posts containing mentions of identity groups (veterans, immigrants, refugees and asylum seekers), where (Un)Deserv., GovCrit. and MoneyAid. ($p < 0.0001$) are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

533	Limitations and Ethical Considerations	in support of alleviating the suffering of PEH.	581
534	Privacy		582
535	Due to the diversity of our dataset and lexical di-	People create social media posts often quickly us-	583
536	versity of social media posts, we anonymized user	ing highly culturally specific and contextual lan-	584
537	handles, and did not obtain or use any priors about	guage. Interpretation of language meaning is not	585
538	the user from their profile or follower network. We	neutral. Words and meanings vary significantly	586
539	only rely on the content of the post for our study.	by culture, informed by age, gender, sexual ori-	587
540	Our dataset includes post IDs and no additional	entation, race, geography, and ethnicity. These	588
541	identifying characteristics as to protect the privacy	understandings and linguistic interpretations are	589
542	of the users.	not fixed and often change over time and context.	590
543	United States Focus	Collaborating with domain experts, advocat-	591
544	Most of our posts are from users with a Western	ing for people experiencing homelessness, bring	592
545	background and as a result, not necessarily repre-	their understanding of institutions and their ever-	593
546	sentative of attitudes towards homelessness outside	changing practices, policies, incidents, and shared	594
547	the U.S. We recognize that this may not be a repre-	discipline-specific vocabulary to speak with depth	595
548	sentative sample of all attitudes towards homeles-	about what we understand is being said in the post.	596
549	ness and this is an important consideration to take	Each annotator brings their own identity and pri-	597
550	into account when reasoning about our results.	ors coupled with professional knowledge into the	598
551	Subjectivity of Attitudes	language interpretation process which forms our	599
552	Due to the nuances and subjectivity of sensitive	positionality as individual researchers.	600
553	social issues such as homelessness, we also un-		
554	derstand that this issue is framed very differently	References	601
555	depending on regionality, ideology, political and	Alina Arseniev-Koehler, Hedwig Lee, Tyler H. Mc-	602
556	economic climate among other socio-political fac-	Cormick, and Megan Andreas Moreno. 2016.	603
557	tors that might not be represented in our dataset.	#proana: Pro-eating disorder socialization on twitter.	604
558	The use of LLMs can further confound these vari-	<i>The Journal of adolescent health : official publication</i>	605
559	ables and also inject societal biases in reasoning	<i>of the Society for Adolescent Medicine</i> , 58 6:659–64.	606
560	about an already marginalized population. Simi-	Arturo Baiocchi and Tyler Arguello. 2019. <i>Street Talk:</i>	607
561	larly, our framing typology may have introduced	<i>Homeless Discourses and the Politics of Service Pro-</i>	608
562	our own prior biases in reasoning about these posts	<i>vision</i> , pages 123–147.	609
563	that may not be in direct alignment with the general	Amber E. Boydston and Justin H. Gross. 2013. <i>Identi-</i>	610
564	public’s attitudes towards homelessness. However,	<i>fying media frames and frame dynamics within and</i>	611
565	we believe this diversity in perspectives is useful	<i>across policy issues.</i>	612
566	in shedding light on the various attitudes towards	Benjamin D Douglas, Patrick J Ewell, and Markus	613
567	homelessness on social media.	Brauer. 2023. Data quality in online human-	614
568	Bad Actors	subjects research: Comparisons between mturk, pro-	615
569	Although our data set was not specifically targeted	lific, cloudresearch, qualtrics, and sona. <i>Plos one</i> ,	616
570	at collecting harmful rhetoric against PEH, we find	18(3):e0279720.	617
571	there are many instances of discriminatory remarks	Robert M. Entman. 1993. <i>Framing: Toward Clarifica-</i>	618
572	and prejudice against marginalized populations. Al-	<i>tion of a Fractured Paradigm.</i> <i>Journal of Communi-</i>	619
573	though we recognize that the characterization of	<i>cation</i> , 43(4):51–58.	620
574	these frames can further exemplify and reinforce	Joseph L Fleiss. 1971. Measuring nominal scale agree-	621
575	harmful rhetoric against PEH and there exists a risk	ment among many raters. <i>Psychological bulletin</i> ,	622
576	of bad actors to further perpetuate stigmas against	76(5):378.	623
577	PEH, we take the stance of supporting PEH by	Barney G. Glaser and Anselm L. Strauss. 1967. <i>Discov-</i>	624
578	further tracking their portrayal on social media at	<i>ery of Grounded Theory Strategies for Qualitative</i>	625
579	scale to benefit advocacy groups and political con-	<i>Research.</i> AldineTransaction, London.	626
580	stituents to better inform their efforts and policies	T. Gowan. 2010. <i>Hobos, Hustlers, and Backsliders:</i>	627
		<i>Homeless in San Francisco.</i> University of Minnesota	628
		Press.	629

630 Minda Hu, Ashwin Rao, Mayank Kejriwal, and
631 Kristina Lerman. 2021. Socioeconomic correlates
632 of anti-science attitudes in the us. *Future Internet*,
633 13(6):160.

634 Nathan Kim, Colette Auerswald, and Jessica Lin. 2013.
635 Perceptions regarding homelessness: Analyzing us
636 tweets.

637 JB Kingree and Walter F Daves. 1997. Preliminary vali-
638 dation of the attitudes toward homelessness inventory.
639 *Journal of community psychology*, 25(3):265–288.

640 Andrew Klein Jessica Richards Fernanda Jahn Verri
641 Marques Vestal Kirsten Moore Sheeley, Alisa Be-
642 linkoff Katz and Zev Yaroslavsky. 2021. *The making
643 of a crisis: A history of homelessness in los angeles*.

644 Minzhi Li, Taiwei Shi, Caleb Ziems, Min-Yen Kan,
645 Nancy F Chen, Zhengyuan Liu, and Diyi Yang. 2023.
646 Coannotating: Uncertainty-guided work allocation
647 between human and large language models for data
648 annotation. *arXiv preprint arXiv:2310.15638*.

649 Julia Mendelsohn, Ceren Budak, and David Jurgens.
650 2021. Modeling framing in immigration discourse
651 on social media. *arXiv preprint arXiv:2104.06443*.

652 Nicholas Pangakis, Samuel Wolken, and Neil Fasching.
653 2023. Automated annotation with generative ai re-
654 quires validation. *arXiv preprint arXiv:2306.00176*.

655 Christopher Michael Rytting, Taylor Sorensen, Lisa Ar-
656 gyle, Ethan Busby, Nancy Fulda, Joshua Gubler, and
657 David Wingate. 2023. Towards coding social sci-
658 ence datasets with language models. *arXiv preprint
659 arXiv:2306.02177*.

660 Tanya Sousa, Alyssa Andrichik, Ed Prestera, Katherine
661 Rush, Colette Tano, and Micaiah Wheele. 2023. *The
662 2023 annual homelessness assessment report (ahar)
663 to congress*.

664 Manfred Stede, Yannic Bracke, Luka Borec, Neele Char-
665 lotte Kinkel, and Maria Skeppstedt. 2023. *Framing
666 climate change in nature and science editorials: Ap-
667 plications of supervised and unsupervised text cate-
668 gorization*. *Journal of Computational Social Science*,
669 6(2):485–513.

670 Jack Tsai, Crystal YS Lee, Jianxun Shen, Steven M
671 Southwick, and Robert H Pietrzak. 2019. Public
672 exposure and attitudes about homelessness. *Journal
673 of Community Psychology*, 47(1):76–92.

674 Lindsey L. Upton. 2016. *Care, control, or criminal-
675 ization?: Discourses on homelessness and social
676 responses*. Ph.D. thesis.

677 Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten
678 Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou,
679 et al. 2022. Chain-of-thought prompting elicits rea-
680 soning in large language models. *Advances in Neural
681 Information Processing Systems*, 35:24824–24837.

A Frame Prediction: Dataset Statistics and Model Training Details 682 683

Combining domain-expert annotated 684
Issue-specific frames in §2, along with 685
posts that are labelled using GPT-4 + Expert in 686
§3, we obtain a total of 10400 posts. We then split 687
these posts into a train-validation-test split using 688
the following criteria: 689

- We ensure roughly an 80 : 10 : 10 proportion 690
for the above mentioned split. 691
- The test set contains *only* domain-expert anno- 692
tated posts. We pick 280 posts that were used 693
in our agreement set in §2, and add another 694
1000 instances to the pool. This overall set of 695
1280 posts is used to pick the best perform- 696
ing Flan-T5 variation for final inference on a 697
larger set of posts. 698
- The remaining set of domain-expert annotated 699
and GPT-4 + Expert annotated posts are then 700
combined and split into a train and validation 701
split. 702

We use the pre-trained Flan-T5 models for the 703
conditional generation task, pooled from the 704
HuggingFace⁶ repository. The task is structured as 705
follows: 706

- **Input format:** The input format consists of 707
the following instruction, along with the post 708
to be labeled – Classify the given post 709
into one or more of the following 10 710
labels: {list of labels} post: {} 711
- **Output format:** If a post needs to be fil- 712
tered out using the Public Opinion filter, we 713
generate a ‘0’ label, otherwise, we generate 714
a comma-separated list of Issue-specific 715
frames. 716

We demonstrate the data splits, training details 717
and results of different Flan-T5 variations in **Table 3**, **Table 4**, **Table 5** and **Table 6**. 718 719

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
	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	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 at least 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