The Crime of Being Poor: Associations between Crime and Poverty on Social Media in Eight Countries

Georgina Curto,[1] **Svetlana Kiritchenko,**[2] **Kathleen C. Fraser,** [2] **and Isar Nejadgholi**[2]
[1]University of Notre Dame, Notre Dame, USA
[2]National Research Council Canada, Ottawa, Canada
gcurtore@nd.edu, {svetlana.kiritchenko,kathleen.fraser,isar.nejadgholi}@nrc-cnrc.gc.ca

Abstract

Content Warning: *This paper presents textual examples that may be offensive or upsetting.*

Negative public perceptions of people living in poverty can hamper policies and programs that aim to help the poor. One prominent example of social bias and discrimination against people in need is the persistent *association of poverty with criminality*. The phenomenon has two facets: first, the belief that poor people are more likely to engage in crime (e.g., stealing, mugging, violence) and second, the view that certain behaviors directly resulting from poverty (e.g., living outside, panhandling) warrant criminal punishment. In this paper, we use large language models (LLMs) to identify examples of crime–poverty association (CPA) in English social media texts. We analyze the online discourse on CPA across eight geographically-diverse countries, and find evidence that the CPA rates are higher within the sample obtained from the U.S. and Canada, as compared to the other countries such as South Africa, despite the latter having higher poverty, criminality, and inequality indexes. We further uncover and analyze the most common themes in CPA posts and find more negative and biased attitudes toward people living in poverty in posts from the U.S. and Canada. These results could partially be explained by cultural factors related to the tendency to overestimate the equality of opportunities and social mobility in the U.S. and Canada. These findings have consequences for policy-making and open a new path of research for poverty mitigation with the focus not only on the redistribution of wealth but also on the mitigation of bias and discrimination against people in need.

1 Introduction

Computational methods provide new insights that can trigger innovative interventions towards the UN Sustainable Development Goals (SDGs) (Vinuesa et al., 2020). The "end of poverty in all its forms everywhere" is the #1 UN SDG and constitutes an urgent call for action. However, there are still 685M people living in extreme poverty worldwide (The World Bank, 2023), and COVID-19 has particularly affected the poorest: the number of people living in extreme poverty rose by 11% in 2020 (The World Bank, 2022). Poverty affects not only the population in developing regions but also a significant percentage of those living in thriving economies (Creamer et al., 2022; Eurostat, 2022): in the United States, 11.6% of the population (37.9M people) are in a situation of poverty (Creamer et al., 2022), and within the EU-27, there are 95.3M people (22% of the population) at risk of poverty (Eurostat, 2022). In this context, innovative measures are required to work towards poverty mitigation across the globe. Traditional policies based on the redistribution of wealth may not be enough, as evidenced by a deceleration in the poverty reduction rates in the last decades (Claudia et al., 2018). Interdisciplinary research, incorporating AI-enabled tools into these efforts, offers perspectives previously unavailable with traditional quantitative and qualitative approaches.

Prejudice against people with low socio-economic status can hinder poverty reduction efforts (Arneson, 1997; Everatt, 2009). When the poor are believed to be responsible for their situation and, therefore, unworthy of help ("undeserving poor"), it is difficult for policy makers to approve and implement poverty reduction policies (Nunn and Biressi, 2009). Therefore, the blameworthiness of the poor could have an impact on the actual poverty levels. However, aporophobia—a social bias of rejection and contempt for the poor and the associated institutional discrimination of the group (Cortina, 2022)—has only recently become a topic of research, with initial studies providing empirical evidence of this phenomenon (Curto et al., 2022).

An important facet of aporophobia is a frequent *association of poverty with criminality* in society

(Kiritchenko et al., 2023). It can be seen (1) in public opinion as individuals' beliefs and stereotypes (we refer to this as *crime–poverty bias*), and (2) in discriminatory actions at an institutional level (we refer to it as *institutional criminalization of the poor*). In the first instance, homeless people are stereotyped as threatening, violent, not wanting to work, and mentally ill (Faragó et al., 2022), and poor people, as a group, are believed to be frequently involved in criminal activities, such as theft and illegal drug dealing. In turn, the criminal offenses devised for sleeping rough in many cities of the so-called developed countries are an example of the institutional criminalization of the poor (Barcelona City Council, 2005). Other examples include probation and even incarceration for people who cannot afford to pay minor fines (Geraghty, 2015; Terradillos Basoco, 2020).

The study of the crime–poverty association (CPA) needs to be rooted in cognitive science and the philosophy of discrimination. Allport (1954) explains that human beings interpret information by classifying it into categories based on their previous experience. This process is at the origin of prejudices, which have been described as overgeneralized and, therefore, misleading beliefs that result in systematic and predictable errors in decision making based on available heuristics (Kahneman, 2011). Prejudices can lead to social bias reflected in derogatory speech (Ely, 1980; Greenwald et al., 2003), and even to discrimination through unfair actions towards members of a group (Allport, 1954).

In this paper, we use NLP techniques to analyze the prevalence of social media discourse around CPA in eight geographically-diverse countries where English is an official or majority language. Despite the fact that social networks are not representative of the whole population, they do act as a partial mirror that allow researchers to measure and track societal biases and discriminatory behaviors that might be hard to detect by traditional quantitative analyses, such as surveys. We employ pre-trained large language models (LLMs) to classify 500K social media posts in English from the selected countries, identifying statements discussing CPA. We further conduct a topic-modelling analysis to examine the *content* of the retrieved posts and determine which aspects of CPA are more salient in different regions. Finally, we contextualize our findings by comparing them with published indicators of poverty, criminality, and inequality in each of the countries in our study. The preliminary results offered in this article inform about the differences in the social discourse around CPA within the countries in scope and open lines of research towards the mitigation of poverty by acting on biases and discriminatory actions that affect people in need.

2 Related Work

We briefly review some of the related social science research on the correlation between crime and poverty, as well as computational work on detecting and responding to social biases.

2.1 Correlation between Poverty and Criminality

Previous research provides evidence of the correlation between poverty and criminality (Looney and Turner, 2018; Becker, 1968). For example, economists have examined the correlation between poverty and property crimes, assessing the cost-benefit analysis (Freeman, 1999; Wu and Wu, 2012; Costantini et al., 2018). Research in criminology also correlates poverty and violent crimes, explained by the strain generated by individuals' failure to achieve socially valued objectives (Agnew, 1992, 2001). Within the context of the U.S., studies found that children who grew up in families in the bottom 10% of the income distribution are 20 times more likely to go to prison in their early 30s than children born in top-decile families. Further, one in ten boys born to families with the lowest income decile are in prison at age 30, and they account for 27% of prisoners of that age (Looney and Turner, 2018). Similarly, other multidimensional factors associated with poverty, such as social determinants of health and education, appear to have an impact on incarceration (Miller, 2013; Hinton, 2017). In turn, the mark of a criminal record generates impediments for employment (Pager, 2003; Mueller-Smith, 2015), which constitutes a vicious circle to get out of poverty. Despite the fact that poverty and criminality are correlated, the strength of the correlation can be greatly over-estimated in public opinion, leading to bias and discrimination against people in need. In this paper, we examine online discussions on crime–poverty association (in the form of bias and discrimination) and explore socio-economic factors as well as cultural differences that might affect the prevalence of CPA discussions in certain regions of the world.

It must be highlighted that the existing correlation between poverty and criminality does not justify accepting stereotypes and acts of discrimination that associate people in need, as a group, with crime. As is the case with gender discrimination, racism or xenophobia, this type of shared generalization exacerbates underlying social biases and generates a vicious circle for vulnerable social groups. Further, generalizations that associate people in poverty and crime are detrimental to the dignity of the persons affected, and thwart the efforts towards poverty reduction. Bearing in mind the urgency to alleviate poverty, this type of bias and discrimination needs to be identified, tracked and mitigated.

2.2 Addressing Social Bias with NLP

A significant research effort in NLP has been dedicated to identification and mitigation of social bias in human-written text, particularly in social media. This includes works on stereotype detection (Fokkens et al., 2018; Marzouki et al., 2020; Charlesworth et al., 2021; Fraser et al., 2022) and identifying and countering hate speech and toxic language (Schmidt and Wiegand, 2017; Fortuna and Nunes, 2018; Vidgen et al., 2019; Tekiroğlu et al., 2020; Kiritchenko et al., 2021; Zhu and Bhat, 2021; Ashida and Komachi, 2022). However, previous work mostly focused on identifying sub-types of harmful language that target specific groups, such as sexism (Istaiteh et al., 2020; Chiril et al., 2020; Samory et al., 2021), racism (Istaiteh et al., 2020; Waseem, 2016; Davidson et al., 2017), and anti-immigrant hoaxes (Bourgeade et al., 2023).

Only a handful of studies have investigated bias based on socio-economic status. Curto et al. (2022) provided evidence of bias against poor people in word embeddings, built on Google News, Twitter, and Wikipedia corpora. Perez Almendros et al. (2020) collected and annotated a dataset with patronizing and condescending language that included homeless and poor people among the selected target groups. Kiritchenko et al. (2023) argued that existing toxic language detection models and datasets are inadequate to effectively identify bias against people with low socio-economic status. We continue and extend this line of work to analyze the social media discourse on social bias and institutional criminalization of the poor in different regions of the world.

3 Data and Methods

In this project, we aim to analyze and compare the social media discourse on CPA, expressed both in the form of bias and institutional criminalization of the poor, in various regions of the world. We choose Twitter as our data source, now called X,[1] which provides a rich stream of everyday conversations of ordinary people on a variety of topics. We start by collecting a large set of tweets in English referring to poor people, written by users from eight geographically-diverse countries (Section 3.1).

Next, we automatically retrieve tweet sentences that refer to an association between criminality and poverty. We approach this task as binary text classification with two categories: 'text refers to a crime–poverty association' (or 'CPA') and 'no reference to a crime–poverty association' (or 'not CPA'). This task has not been extensively studied in NLP, and there are no annotated data available. Thus, we turn to pre-trained large language models as zero-shot or few-shot text classifiers (Wang et al., 2023). While text classifiers, in general, benefit from fine-tuning on task-specific labelled datasets, LLMs have shown to be effective classifiers in scenarios with limited resources (Chae and Davidson, 2023). This is specifically relevant to our study, where we analyze various aspects of CPA in social media discourse across different countries at an aggregate level, and individual labels assigned by LLMs are not critical. For model selection and validation, we collect and manually label a small test set (described in Section 3.2), on which we evaluate the performance of several state-of-the-art LLMs, listed in Section 3.3.

After identifying the best performing LLM on the test set, we prompt it to automatically categorize the full set of tweet sentences into the 'CPA' and 'not CPA' categories. Using this approach, we compute and compare the percentages of tweets about poor people that refer to an association between criminality and poverty, in the eight countries. We further examine the CPA posts to discover and analyze common themes on the issues of criminality and poverty around the world using unsupervised topic modeling (Section 3.4). To contextualize our findings, we compare them with the published statistics on various economic and criminality indicators in the studied countries, and

[1] Since the platform was called Twitter when we collected the data, we use that name throughout. Note that we collected the data prior to the introduction of the paywall.

speculate on possible reasons for the observed discrepancies in the Discussion (Section 5). In the following, we describe the dataset and methodology in detail.

3.1 Twitter Corpus

We use the Twitter Research API to collect English tweets pertaining to poor people from 25 August 2022 to 23 November 2022. We first collected a set of query terms from social psychology literature and augmented this set with synonyms and related terms. We then collected a one-week sample of tweets using this set of query terms and manually examined the retrieved tweets to discard the terms that resulted in very small numbers of retrieved tweets or many irrelevant tweets. The final list of query terms includes: *the poor* (used as a noun as opposed to an adjective, as in 'the poor performance'), *poor people*, *poor ppl*, *poor folks*, *poor families*, *homeless*, *on welfare*, *welfare recipients*, *low-income*, *underprivileged*, *disadvantaged*, *lower class*. The single word *poor* is not included as a query term because of its polysemy (it can apply to people, but can also be used to describe other things, e.g., 'poor results'). We exclude explicitly offensive terms that tend to be used in personal insults, such as *trailer trash* or *scrounger*.

Re-tweets, tweets with more than five hashtags, and tweets with URLs are excluded. Tweets written by bots (identified as user accounts with the user or screen name including the word 'bot') are excluded as well. This filtering step helps remove posts from commercial accounts. Since tweets can be up to 280 characters and include several sentences, we split each tweet into individual sentences and keep only sentences that include at least one of the query terms. In total, there are over 1.3 million sentences in the corpus.

We are also interested in the geographical locations from which tweets originated. Unfortunately, only about 2% of tweets include the exact geographical information. Therefore, in addition to the tweet location ('place' field), we rely on user location that users voluntarily provide in their Twitter accounts, which is available for about 60% of posts. The user location is recorded as a free-form text, and tweeters are often very creative in describing their location (e.g., "somewhere on Earth"). We automatically parse user location descriptions to extract country information for the most frequently mentioned countries. In the absence of a country name, we consider the mentions of U.S. states, Canadian provinces, and major cities in the U.S., U.K. and Canada, since these are also commonly used by tweeters. (Major cities from other countries are rarely used without the country name.) Table 1 shows the number of sentences in the corpus per geographical location. In the following analysis, we focus on eight geographically-diverse countries with English as an official or majority language, for which the corpus contains at least 1,000 sentences: the United States of America, the United Kingdom, Canada, India, Nigeria, Australia, South Africa, and Kenya. There are, in total, 486,281 sentences written by tweeters from these eight countries.

Location	# of sentences
United States	326,993
United Kingdom	80,947
Canada	32,978
India	14,029
Nigeria	10,529
Australia	9,698
South Africa	7,729
Kenya	3,378
Total (eight countries)	486,281
Other locations	337,252
No location information	539,365
Total	1,362,898

Table 1: The number of sentences in the tweet corpus per geographical location.

3.2 Manually Labeled Evaluation Datasets

To validate and test automatic LLM-based classifiers, we manually annotate a small portion of the dataset, identifying examples of CPA. We capture both statements that illustrate 'crime-poverty bias' expressed in public opinion and sentences describing 'institutional criminalization of the poor'. In particular, we label as 'CPA':

1. 'Crime-poverty bias': utterances that state or refer to the belief that people living in poverty, as a group, are more likely to be engaged in illegal activities (e.g., stealing, harassing other people, selling illegal drugs, etc.);
2. 'Institutional criminalization of the poor': utterances that state that poor people are more likely to face legal consequences due to their lack of financial resources (e.g., being put and kept in jail for minor offenses or unpaid fees/debts, inability to pay bail or hire lawyers, anti-homelessness laws and actions, etc.).

Both categories characterize CPA and need to be

CPA	***Examples of statements illustrating 'crime-poverty bias':*** Poor people just steal other people's property rather than get a job, earn money and buy what they need. Urban inner-city crime has been real since we had cities filled with poor people. It's been like this since forever so you can say it is normal for poor people to commit crimes. ***Examples of statements illustrating 'institutional criminalization of the poor':*** Law enforcement and prisons are routinely used against poor people not because of safety, but to protect the wealthy. Poor people from poor neighborhoods are easy targets for incarceration, police extortion and police murders. Cities do have bylaws to discard the homeless like trash.
Not CPA	***Examples of statements not expressing an association between poverty and criminality:*** Democratic states are overrun with crime, homelessness, drugs, and debt. They haven't stopped the crime and do little for the homeless. The government is stealing from the poor to give to the rich.

Table 2: Example sentences referring to an association of poverty with criminality (top), and examples that are not labelled as making the association (bottom). All tweet sentences are paraphrased to protect the privacy of the users.

Dataset	Total sent.	CPA sent. (% of total)
Test 'Random'	400	12 (3%)
Test 'Crime'	400	151 (38%)
Development	107	38 (36%)
Total	907	201 (22%)

Table 3: The number of sentences in each manually labeled test set.

Dataset	% agree	Cohen's κ
Test 'Random'	97.0	0.49
Test 'Crime'	88.0	0.74
Development	87.9	0.73

Table 4: Inter-annotator agreement on each test set.

examined side by side. Moreover, each one supports and reinforces the other. Harmful beliefs and stereotypes perpetuated in online communications affect the overall public perception of the group and build a strong foundation for systemic policies. On the other hand, disproportional police surveillance and incarcerations of poor people lead to statistical evidence supporting the stereotypes.

Note that the CPA category does not only include utterances that *perpetuate* the belief of poor people being criminals or *support* the laws and policies discriminating against the poor. In fact, many writers criticize such beliefs and advocate for stronger social support and better policies that would empower people in poverty and help them overcome the adversities. Although such posts do not directly contribute to reinforcing the crime–poverty association, we also label such messages with the CPA category because they are evidence of the existence of the phenomenon.

We first randomly select 50 sentences originating from each of the eight countries of interest (400 sentences in total) and annotate them for CPA. We call this test set 'Random'. Only 12 sentences (3%) in this set are labeled as CPA. To increase the proportion of CPA messages, we create a second test set, called 'Crime', using the following data enrichment procedure. We first select sentences from the tweet corpus that include at least one of the following crime-related words: *crime*, *crimes*, *criminal**[2], *jail**, *prison**, *arrest**, *police*, *cops*, *policing*, *imprison**, *incarcerat**, *prosecut**, *assault**, *harass**, *steal*, *stealing*, *stolen*, *stole*, *theft*. From these sentences, we again randomly select 50 sentences originating from each of the eight countries of interest (400 sentences in total) and annotate them for CPA. In this test set, 151 sentences (38%) are labeled as CPA. Table 2 shows example sentences for both crime–poverty bias in public opinion and institutional criminalization of the poor (labeled as CPA), and sentences that mention crime but are not labelled as CPA since they do not imply bias or discrimination against the group. For the latter, in many cases, crime and poverty are both mentioned as problems that need to be addressed but are not causally related. In other cases, poor people are described as the victims, rather than the perpetrators, of crime.

Finally, a smaller development set was created in a similar way as the test set 'Crime'. Table 3 shows the details of the three datasets.

Two authors of this paper annotated the sentences independently, and then all disagreements were discussed and resolved. Table 4 shows the inter-annotator agreement on the three evaluation

[2]The wildcard * indicates any number of alphabetic characters to cover morphological variants of the word (e.g., criminal, criminals, criminalization, etc.)

datasets. The agreement is measured in two ways: (1) as the percentage of sentences on which the two annotators agree on the label, and (2) as Cohen's κ. Both metrics demonstrate moderate to substantial levels of agreement.

3.3 Automatic Classification of CPA

After manually annotating 907 sentences in the previous section, we identified 201 instances of CPA. Since this is not sufficient to train a classifier to detect CPA from scratch, we instead investigate the possibility of using methods not requiring large amounts of data for training, such as zero- or few-shot learning with LLMs. We use the small annotated dataset to evaluate and compare the models.

We experiment with two open-source LLMs and one commercial model:

1. **Llama 2** (7B), an open-source model released by Meta (Touvron et al., 2023)
2. **Flan-T5** (XL), an open-source model created by Google (Chung et al., 2022)
3. **ChatGPT** (GPT-3.5-turbo-0125), a commercial model produced by OpenAI.

We prompt the LLMs in both zero-shot and few-shot settings. It has been shown that LLMs' performance varies significantly with even minor variations in prompts, and prompts that are optimal for one model might not perform well for another model (Voronov et al., 2024). Therefore, we design zero-shot and few-shot prompts for each of the models separately. The best prompts for each model are reported in Appendix A. Temperature was set to zero for all three LLMs.

3.4 Unsupervised Topic Modeling

Next, we examine the content of CPA posts to discover the most salient themes in the discussions on the association between criminality and poverty in the different regions. We apply unsupervised topic modeling on all the tweet sentences automatically classified as CPA. Topic modeling is a helpful tool to quickly analyze semantic content of large amounts of text. We use BERTopic (Grootendorst, 2022), a flexible state-of-the-art toolkit for unsupervised, semi-supervised, and supervised topic modeling. BERTopic employs a density-based clustering technique HDBSCAN (Campello et al., 2013), which identifies dense regions in the text representation space and leaves texts outside these dense regions as outliers. These dense regions would represent the most commonly discussed topics in tweets associating poor people and criminality.

After a few preliminary experiments, we set the parameters as follows. Texts are converted to numerical vectors using sentence transformers[3] with the *roberta-large-nli-mean-tokens* pretrained embedding model. We use CountVectorizer[4] as the vectorizer model, remove stopwords and terms that appear in less than 5% of the sentences ($min\_df = 0.05$), and set the minimum size of the clusters as $min\_cluster\_size = 100$. For all the other parameters, the default settings of the BERTopic package are used.

We analyze the most common topics discovered by BERTopic and compare their prevalence in Twitter discussions in the eight countries in scope.

4 Results

First, we report the results obtained from the automatic LLM-based classification of a large collection of tweets to identify CPA for the eight countries in scope (Section 4.1). We analyze these results in the context of economic and criminality indicators for each country. Next, we look at the content of CPA posts and examine common themes in the Twitter discourse around CPA in different parts of the world (Section 4.2).

4.1 Crime–Poverty Association in Tweets

We evaluate the performance of the three LLMs with their respective best zero- and few-shot prompts on the development and test sets. Table 5 reports the results. Among the three tested LLMs, on all three datasets, the best performance is achieved by ChatGPT (GPT-3.5-turbo) using a few-shot prompt. Therefore, we proceed with that model in the following analysis of CPA in different regions of the world.

We prompt ChatGPT with the selected few-shot prompt on all 486,281 sentences of the original corpus written by users from the eight countries in scope. In total, 38,034 sentences (8%) were classified by ChatGPT as CPA. Table 6 shows the percentages of CPA sentences per country.

The results indicate that tweeters from the United States and Canada are more likely to re-

[3] https://github.com/UKPLab/sentence-transformers

[4] https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html

Classification method	Dev. set			Test set 'Crime'			Test set 'Random'		
	Acc.	F_{avg}	F_{CPA}	Acc.	F_{avg}	F_{CPA}	Acc.	F_{avg}	F_{CPA}
Llama 2									
zero-shot	0.76	0.75	0.61	0.71	0.70	0.57	0.96	0.96	0.38
few-shot	0.77	0.77	0.71	0.69	0.70	0.61	0.94	0.95	0.40
Flan-T5									
zero-shot	0.74	0.74	0.62	0.70	0.71	0.62	0.95	0.96	0.36
few-shot	0.76	0.72	0.50	0.73	0.69	0.49	**0.97**	0.97	0.44
GPT-3.5-turbo									
zero-shot	0.79	0.78	0.63	0.71	0.70	0.57	**0.97**	0.97	0.40
few-shot	**0.83**	**0.83**	**0.76**	**0.77**	**0.77**	**0.69**	**0.97**	**0.98**	**0.67**

Table 5: Performance of the three LLMs on the evaluation datasets: accuracy (Acc.), support-weighted macro-averaged F1-score (F_{avg}), and F1-score on the CPA category (F_{CPA}). Highest values in each column are in bold.

fer to and discuss crime–poverty associations in their posts than users in the other countries. To contextualize these findings, Figure 1 presents the percentage of sentences labeled as CPA for each of the eight countries together with their economic and crime indicators: overall criminality rate, poverty headcount ratio at $2.5 a day (purchasing power adjusted prices), inequality indicators (Gini Index and 10% income share), and unemployment rate. The results show that the United States and Canada have the highest CPA, a sign of pervasiveness of the phenomenon in these countries, despite having comparable or even lower poverty, criminality, inequality and unemployment rates than the other countries in scope. It is also worth noting that India, which has a higher poverty headcount and similar levels of inequality and unemployment rates to the United States, has the lowest CPA rate, which may indicate that crime and poverty are seen as less associated with each other and discussed as separate negative factors in this society. Also, online posts from South Africa discuss CPA at a substantially lower rate than in the U.S. and Canada, despite South Africa having the highest levels of inequality, criminality, and unemployment among the countries included in the study. We discuss factors that might influence CPA, in addition to socioeconomic indicators, in Section 5.

4.2 Common Themes in CPA Tweets

To examine the content of Twitter discussions related to crime–poverty associations, we apply unsupervised topic modeling using BERTopic on the 38,034 sentences labeled by ChatGPT as CPA. BERTopic extracts 20 topics (ordered by the number of sentences in a topic), and leaves around 55% of sentences unclustered. A few topics are grouped by the targeted subpopulation, e.g., *homeless people* or *people on welfare*, but semantically represent a wide mix of themes pertaining to the group. One such topic to note is Topic 11, which discusses the prevalence of Black people amongst the economically disadvantaged, racial discrimination aggravated by aporophobic attitudes, and other topics on the intersection of race and economic inequality. However, we exclude such topics from our current analysis as we aim to focus on themes relating to criminality in general, and not to specific subgroups.

Of the 20 initial topics, we select ten most interpretable topics that could be mapped to a social theme.[5] Figure 2 shows how often these ten topics are discussed in each of the eight countries (i.e., the shading in cell (i, j) represents the proportion of tweet sentences from Country$_j$ that are clustered in Topic$_i$).

The topic modeling analysis reveals major differences in the social media discourse around poverty in North America as opposed to the other examined countries. The highest proportions of sentences from the U.S. and Canada fall in Topics 1, 3, and 4. Topic 1 comprises sentences expressing negative attitudes towards homeless people, portraying them as criminals and drug addicts (e.g., "*homeless crime is a huge issue*", "*rampant homeless addicts make it an unsafe place to live*", "*filth and petty crimes from homeless*"). Topic 3 includes sentences on homelessness being illegal (e.g., "*if you're homeless you are illegal*"), and Topic 4 consists of calls for authorities to remove homeless people from the streets (e.g., "*get the homeless, druggies off the streets*", "*clear away homeless en-*

[5]Note that the topic interpretation and mapping to social themes are done manually by the authors and, therefore, subjective. Not all sentences in a topic discuss the corresponding theme in the same way, and some topics may contain opposing views. The manual label assigned to a topic is intended to represent the interpretation for the majority of sentences in that topic.

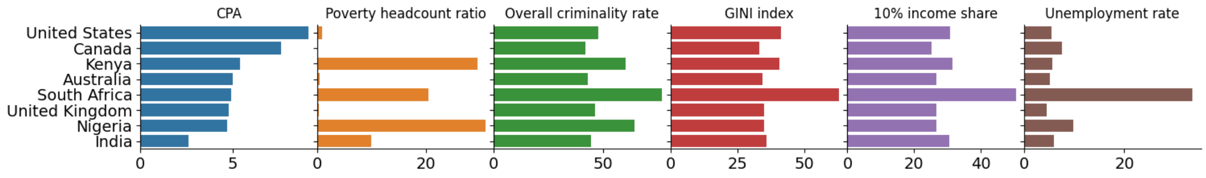

Figure 1: For the countries in scope, percentage of CPA discussions in tweets related to poor people, poverty headcount at $2.5 a day (purchasing power adjusted prices), overall criminality rate, indicators of inequality (Gini index and 10% income share), and unemployment rates. Sources: poverty headcount ratio, Gini index, and 10% income share rates are as reported by the World Bank (2017 or nearest year); unemployment rates are from the World Bank (2021); overall criminality rates are from worldpopulationreview.com.

Country	% CPA
United States	9.1
Canada	7.6
Kenya	5.4
Australia	5.0
South Africa	4.9
United Kingdom	4.8
Nigeria	4.7
India	2.6

Table 6: Percentage of sentences from each country classified as CPA by ChatGPT.

campments"). Further, in Topic 7 tweeters from the U.S. and Canada talk about various incidents of homeless people involved in violent attacks and other crimes (e.g., "*a person was mugged by a homeless dude with an axe*"). Other countries include substantially less posts in these topics.

In contrast, Topics 5 and 6 are among the largest topics for the other countries, especially for Nigeria, Kenya, and India. These topics comprise posts describing the discriminative nature of laws and policies disproportionately affecting people with less resources and power (e.g., "*laws are made for the poor*", '*rules only affect the poor*"). A significant number of posts from the U.K. also appear in Topics 12, 14, and 16, where tweeters criticize government policies that aggravate the hardship of living in poverty or punish poor people for being poor (e.g., "*government continues persecution of the poor*", "*a tool with which to mass murder the poor, and then blame them for it*"). Finally, in Topic 9, tweeters from all countries denounce the current state of social structure where poor and homeless people are frequently arrested and kept in jails for minor offenses or no reason at all (e.g., "*meanwhile, poor people guilty or not guilty of less substantial crimes are going straight to prison*").

5 Discussion

Our results reveal substantial quantitative and thematic differences in online discourse related to the crime–poverty association in the eight studied countries. The proportion of CPA discussions in Twitter in the U.S. is almost twice as high as in the other countries in scope, both in the Global North and Global South. Furthermore, posts from the U.S. and Canada more often express bias against poor people as a group, associating them with different kinds of crime. There are many posts portraying homeless people as thieves, muggers, and drug addicts, and calling for systemic measures to remove them from the streets. People receiving welfare support are often described as scammers and not wanting to work. In contrast, tweeters from the other countries are more concerned with unfair laws and regulations that disproportionately affect the poor and aggravate their hardship.

A potential explanation for these findings could be found in the narrative shared by the United States and Canada of being the "lands of opportunity", where the rich and the poor are thought to have equal chances for success (United Nations, 2018). The poor, therefore, would be blamed and even punished for their inability to get out of poverty (Desmond, 2023). However, the principle of equal opportunity can be considered an oxymoron since every person is exposed to different opportunities in life from the moment of birth (Sandel, 2020), and the job market for individuals with low educational qualifications, disability, and with no assistance to find employment is very limited. The indicators of social mobility and inequality support the claim from the United Nations that the poor in the United States are overwhelmingly those born into poverty (United Nations, 2018): intergenerational social mobility in the United States from the bottom to the top income quantile is as

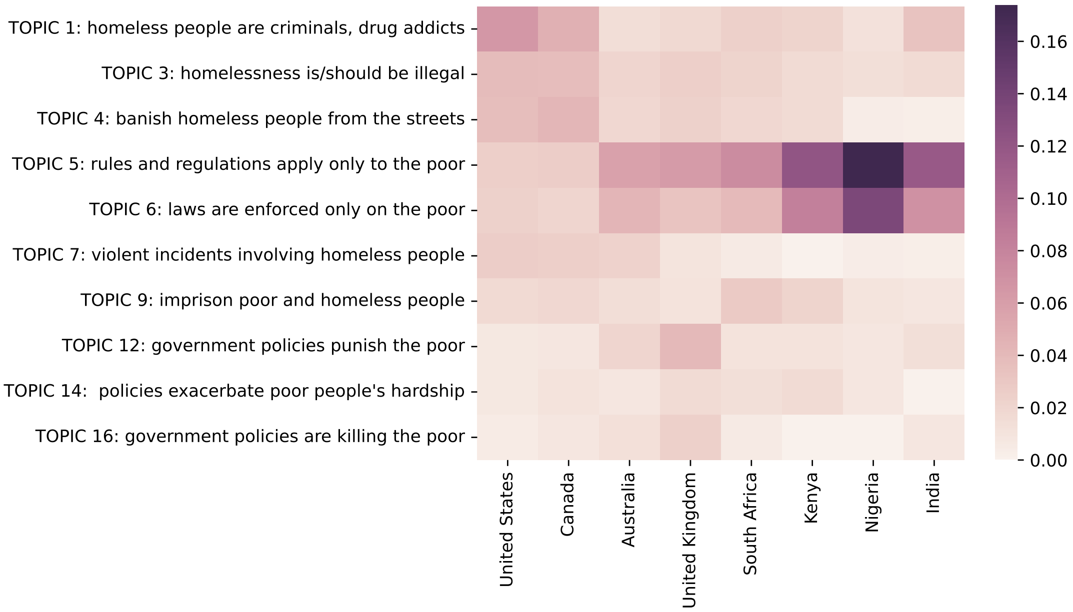

Figure 2: The proportion of tweet sentences from each country in ten most interpretable topics.

low as 7.8%, below European countries such as the U.K., France, Italy, or Sweden (Alesina et al., 2018). In fact, intergenerational mobility has declined substantially over the last 150 years in the United States (Song et al., 2020) and income inequality has been growing since the 1980s (The World Bank, 2023).

6 Conclusions

This paper presents a quantitative and a thematic analyses of the prevalence of online discussions on the association of criminality and poverty in eight geographically-diverse countries. It provides evidence that such discussions more frequently occur within the U.S. and Canada, as compared to the other countries such as South Africa, despite the latter having higher poverty, criminality, and inequality indexes. Moreover, the most prevalent topics in these discussions in the U.S. and Canada demonstrate negative attitudes and social bias against people in need, for example, portraying homeless people as criminals, muggers, and drug addicts. In contrast, in the other countries in scope, such as Nigeria, India or Kenya, the most frequently discussed topics refer to the unfairness of laws that do not provide support or even discriminate against the poor. We speculate that these differences can be partially attributed to the rhetoric of equal opportunities widespread in the U.S. and Canada. The insights obtained from this study shed light towards a new path of research for poverty mitigation, where the focus should be not only on the redistribution of wealth but also on the mitigation of social bias and discrimination against the poor.

Limitations

This article constitutes a preliminary analysis of CPA in social media discourse, including both crime–poverty bias in public opinion and institutional criminalization of the poor. As such, it offers evidence of the phenomenon and aims to open a new line of research that needs to go deeper and explore the different types of stereotypes and acts of discrimination that contribute to the phenomenon.

The presented analysis covers only English-language posts from eight countries representing different regions of the world. While English is an official or majority language in the selected countries, other languages are also widely spoken in some of these countries. The user posts are collected using a pre-specified set of terms in standard English that may exclude related terms in regional dialects. Further, the posts are collected only from one social media platform, Twitter. Similar to any other social network, Twitter represents a non-random sample of the general population in terms of gender, age, ethnicity, and other sociodemographic characteristics (Mislove et al., 2011). In particular, it is predominantly used in the United States (Barbieri et al., 2020). Therefore, the findings from this study may not generalize to population at large. Future work will extend this analysis to cover other languages, geographical locations, and sources of public opinion. In addition, the inclusion of multi-modal data, combining text with images and video, can enrich the analysis of social media discourse on poverty and criminality.

The analysis relies on automatic methods of data collection and categorization. While allowing to process large amounts of data, these methods inevitably introduce errors and the quantitative results might be imprecise. Nevertheless, we believe the overall conclusions of the study, supported by both quantitative and qualitative analyses, should hold.

References

Robert Agnew. 1992. Foundation for a General Strain of Crime and Delinquency. *Criminology (Beverly Hills)*, 30(1):47–88.

Robert Agnew. 2001. Building on the Foundation of General Strain Theory: Specifying the Types of Strain Most Likely to Lead to Crime and Delinquency. *The Journal of Research in Crime and Delinquency*, 38(4):319–361.

Alberto Alesina, Stefanie Stantcheva, and Edoardo Teso. 2018. Intergenerational Mobility and Preferences for Redistribution. *American Economic Review*, 108(2):521–54.

Gordon W Allport. 1954. *The Nature of Prejudice*. Basic Books.

Richard J Arneson. 1997. Egalitarianism and the Undeserving Poor. *The Journal of Political Philosophy*, 5(4):327–350.

Mana Ashida and Mamoru Komachi. 2022. Towards automatic generation of messages countering online hate speech and microaggressions. In *Proceedings of the Sixth Workshop on Online Abuse and Harms (WOAH)*, pages 11–23, Seattle, Washington (Hybrid). Association for Computational Linguistics.

Francesco Barbieri, Jose Camacho-Collados, Luis Espinosa Anke, and Leonardo Neves. 2020. TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification. *Findings of the Association for Computational Linguistics: EMNLP 2020*.

Barcelona City Council. 2005. Ordenança de mesures per fomentar i garantir la convivència ciutadana a l'espai públic de Barcelona Acord del Consell Plenari de 23-12-2005. Article 58.2a.

Gary S. Becker. 1968. Crime and Punishment: An Economic Approach. *The Journal of Political Economy*, 76(2):169–217.

Tom Bourgeade, Alessandra Teresa Cignarella, Simona Frenda, Mario Laurent, Wolfgang Schmeisser-Nieto, Farah Benamara, Cristina Bosco, Véronique Moriceau, Viviana Patti, and Mariona Taulé. 2023. A multilingual dataset of racial stereotypes in social media conversational threads. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 686–696.

Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. 2013. Density-based clustering based on hierarchical density estimates. In *Proceedings of the Pacific–Asia Conference on Knowledge Discovery and Data Mining*, pages 160–172. Springer.

Youngjin Chae and Thomas Davidson. 2023. Large language models for text classification: From zero-shot learning to fine-tuning. *Open Science Foundation*.

Tessa ES Charlesworth, Victor Yang, Thomas C Mann, Benedek Kurdi, and Mahzarin R Banaji. 2021. Gender stereotypes in natural language: Word embeddings show robust consistency across child and adult language corpora of more than 65 million words. *Psychological Science*, 32(2):218–240.

Patricia Chiril, Véronique Moriceau, Farah Benamara, Alda Mari, Gloria Origgi, and Marlène Coulomb-Gully. 2020. An annotated corpus for sexism detection in French tweets. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1397–1403, Marseille, France. European Language Resources Association.

Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, et al. 2022. Scaling instruction-finetuned language models. *arXiv preprint arXiv:2210.11416*.

Gerber Claudia, Alexander D Klemm, Li Liu, and Victor Mylonas. 2018. *Personal Income Tax Progressivity: Trends and Implications*. IMF.

Adela Cortina. 2022. *Aporophobia. Why we reject the poor instead of helping them*. Princeton University Press.

M. Costantini, I. Meco, and A. Paradiso. 2018. Do Inequality, Unemployment and Deterrence Affect Crime over the Long Run? *Regional Studies*, 52(4):558–571.

John Creamer, Emily Shrider, Kalee Burns, and Frances Chen. 2022. Poverty in the US: 2022. Technical report, United States Census Bureau.

G. Curto, M.F Jojoa Acosta, F. Comim, and B. Garcia-Zapirain. 2022. Are the poor being discriminated against on the Internet? A machine learning analysis using Word2vec and GloVe embeddings to identify aporophobia. *AI & Society*.

Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. Automated hate speech detection and the problem of offensive language. In

Proceedings of the international AAAI conference on web and social media, volume 11, pages 512–515.

Matthew Desmond. 2023. *Poverty, by America*. Crown.

John Hart Ely. 1980. *Democracy and distrust : a theory of judicial review*. Harvard University Press.

Eurostat. 2022. Database - Income and living conditions.

David Everatt. 2009. The Undeserving Poor: Poverty and the Politics of Service Delivery in the Poorest Nodes of South Africa. *Politikon*, 35(3):293–319.

Laura Faragó, Dávid Ferenczy-Nyúl, Anna Kende, Péter Krekó, and Zoltán Gurály. 2022. Criminalization as a justification for violence against the homeless in hungary. *The Journal of Social Psychology*, 162(2):216–230.

Antske Fokkens, Nel Ruigrok, Camiel Beukeboom, Gagestein Sarah, and Wouter Van Atteveldt. 2018. Studying Muslim stereotyping through microportrait extraction. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Paula Fortuna and Sérgio Nunes. 2018. A survey on automatic detection of hate speech in text. *ACM Computing Surveys (CSUR)*, 51(4):1–30.

Kathleen C Fraser, Svetlana Kiritchenko, and Isar Nejadgholi. 2022. Computational modeling of stereotype content in text. *Frontiers in Artificial Intelligence*, 5:826207.

R.B. Freeman. 1999. The Economics of Crime. *Handbook of Labor Economics,*, 3:3529–3571.

Sarah Geraghty. 2015. How the criminalization of poverty has become normalized in american culture and why you should care. *Mich. J. Race & L.*, 21:195.

Anthony G Greenwald, Brian A Nosek, and Mahzarin R Banaji. 2003. Understanding and Using the Implicit Association Test. *Journal of Personality and Social Psychology*, 85(2):197–216.

Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

E. Hinton. 2017. *From the war on poverty to the war on crime: the making of mass incarceration in America*. Harvard University Press, Cambridge, Massachusetts.

Othman Istaiteh, Razan Al-Omoush, and Sara Tedmori. 2020. Racist and sexist hate speech detection: Literature review. In *2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, pages 95–99.

Daniel Kahneman. 2011. *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Svetlana Kiritchenko, Georgina Curto Rex, Isar Nejadgholi, and Kathleen C. Fraser. 2023. Aporophobia: An overlooked type of toxic language targeting the poor. In *Proceedings of the 7th Workshop on Online Abuse and Harms (WOAH)*, pages 113–125, Toronto, Canada. Association for Computational Linguistics.

Svetlana Kiritchenko, Isar Nejadgholi, and Kathleen C Fraser. 2021. Confronting abusive language online: A survey from the ethical and human rights perspective. *Journal of Artificial Intelligence Research*, 71:431–478.

Adam Looney and Nicholas Turner. 2018. Work and opportunity before and after incarceration. Technical report, Brookings.

Yousri Marzouki, Eliza Barach, Vidhushini Srinivasan, Samira Shaikh, and Laurie Beth Feldman. 2020. The dynamics of negative stereotypes as revealed by tweeting behavior in the aftermath of the charlie hebdo terrorist attack. *Heliyon*, 6(8):e04311.

R.J. 2013 Miller. 2013. "Race, Hyper-Incarceration, and US Poverty Policy in Historic Perspective". *Sociology Compass,*, 7, no. 7:573.

Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J.Niels Rosenquist. 2011. Understanding the Demographics of Twitter Users. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*.

Michael Mueller-Smith. 2015. The Criminal and Labor Market Impacts of Incarceration. *Working Paper*.

Heather Nunn and Anita Biressi. 2009. The undeserving poor. *Soundings*, 41.

D. Pager. 2003. The mark of a criminal record. *American Journal of Sociology*, 108(5):937–975.

Carla Perez Almendros, Luis Espinosa Anke, and Steven Schockaert. 2020. Don't patronize me! an annotated dataset with patronizing and condescending language towards vulnerable communities. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5891–5902, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Mattia Samory, Indira Sen, Julian Kohne, Fabian Flöck, and Claudia Wagner. 2021. "call me sexist, but...": Revisiting sexism detection using psychological scales and adversarial samples. In *Proceedings of the international AAAI conference on web and social media*, volume 15, pages 573–584.

Michael J. Sandel. 2020. *The tyranny of merit*. Penguin Random House.

Anna Schmidt and Michael Wiegand. 2017. A survey on hate speech detection using natural language processing. In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, pages 1–10.

Xi Song, Catherine G. Massey, Karen A. Rolf, Joseph P. Ferrie, Jonathan L. Rothbaum, and Yu Xie. 2020. Long-term decline in intergenerational mobility in the United States since the 1850s. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 117, pages 251–258. National Academy of Sciences.

Serra Sinem Tekiroğlu, Yi-Ling Chung, and Marco Guerini. 2020. Generating counter narratives against online hate speech: Data and strategies. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1177–1190, Online. Association for Computational Linguistics.

Juan Terradillos Basoco. 2020. Aporofobia y plutofilia : la deriva jánica de la política criminal contemporánea.

The World Bank. 2022. Poverty and Shared Prosperity 2022: Correcting Course. Technical report.

The World Bank. 2023. Poverty and Inequality Platform (PIP).

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

United Nations. 2018. Report of the Special Rapporteur on extreme poverty and human rights on his mission to the United States of America.

Bertie Vidgen, Alex Harris, Dong Nguyen, Rebekah Tromble, Scott Hale, and Helen Margetts. 2019. Challenges and frontiers in abusive content detection. In *Proceedings of the Third Workshop on Abusive Language Online*, pages 80–93, Florence, Italy. Association for Computational Linguistics.

Ricardo Vinuesa, Hossein Azizpour, Iolanda Leite, Madeline Balaam, Virginia Dignum, Sami Domisch, Anna Felländer, Simone Daniela Langhans, Max Tegmark, and Francesco Fuso Nerini. 2020. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications 2020 11:1*, 11(1):1–10.

Anton Voronov, Lena Wolf, and Max Ryabinin. 2024. Mind your format: Towards consistent evaluation of in-context learning improvements. *arXiv preprint arXiv:2401.06766*.

Zhiqiang Wang, Yiran Pang, and Yanbin Lin. 2023. Large language models are zero-shot text classifiers. *arXiv preprint arXiv:2312.01044*.

Zeerak Waseem. 2016. Are you a racist or am I seeing things? Annotator influence on hate speech detection on Twitter. In *Proceedings of the First Workshop on NLP and Computational Social Science*, pages 138–142.

D. Wu and Z. Wu. 2012. Crime, Inequality and Unemployment in England and Wales. *Applied Economics,*, 44(29):3765–3775.

Wanzheng Zhu and Suma Bhat. 2021. Generate, prune, select: A pipeline for counterspeech generation against online hate speech. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 134–149, Online. Association for Computational Linguistics.

A Prompting LLMs

Here, we list the best zero-shot and few-shot prompts for the three LLMs.

GPT-3.5-turbo

Zero-shot prompt: `Your task is to determine whether a given text associates poor people and criminality. Answer ‘yes’ if the text implies that poor or homeless people commit crimes, are arrested, go to jail, or are harassed by police. Answer ‘no’ if the text implies that poor or homeless people are the victims of crimes, if people steal from the poor, or if the text is neutral. Does the following text relate poverty, criminality, and law enforcement? Answer yes or no only. TEXT: <sentence>`

Few-shot prompt: `Your task is to determine whether a given text associates poor people and criminality.Answer ‘yes’ if the text implies that poor or homeless people commit crimes, are arrested, go to jail, or are harassed by police. Answer ‘no’ if the text implies that poor or homeless people are the victims of crimes, if people steal from the poor, or if the text is neutral.`

```
TEXT: Poor people are all criminals.
ANSWER: Yes

TEXT: Jails are only for poor people.
ANSWER: Yes

TEXT: Police arrest only poor people.
ANSWER: Yes

TEXT: Police harass homeless people and
destroy homeless encampments.
ANSWER: Yes
```

TEXT: High crime rates and high homelessness are two main issues in our country.
ANSWER: No

TEXT: Stealing from the poor is unacceptable.
ANSWER: No

Does the following text relate poverty, criminality, and law enforcement? Answer yes or no only.
TEXT: <sentence>
ANSWER:

Llama 2

System Prompt: Your task is to determine whether a given text implies that poor people commit more crimes, or should be punished for being poor. You may only output 'yes' or 'no'

Zero-shot prompt: Does the following text relate poverty, criminality, and law enforcement? Answer yes or no only. TEXT: <sentence>

Few-shot prompt: Your task is to determine whether a given text associates poor people and criminality. Answer 'yes' if the text implies that poor or homeless people commit more crimes, are arrested more, go to jail more, or are harassed by police just because they are poor. Answer 'no' if the text implies that poor or homeless people are the victims of crimes, if people steal from the poor, or if the text is neutral. Here are some examples for you to learn from:

TEXT: Poor people are all criminals.
LABEL: Yes

TEXT: Jails are only for poor people.
LABEL: Yes

TEXT: Stealing from the poor is unacceptable.
LABEL: No

TEXT: Police arrest only poor people.
LABEL: Yes

TEXT: Police harass homeless people and destroy homeless encampments.
LABEL: Yes

TEXT: High crime rates and high homelessness are two main issues in our country.
LABEL: No

Based on these examples, does the following text relate poverty, criminality, and law enforcement? Answer yes or no only.
TEXT: <sentence>
ANSWER:

Flan-T5

Zero-shot prompt: Your task is to determine whether a given text associates poor people and criminality. Answer 'yes' if the text implies that poor or homeless people commit more crimes, are arrested more, go to jail more, or are harassed by police just because they are poor. Answer 'no' if the text implies that poor or homeless people are the victims of crimes, if people steal from the poor, or if the text is neutral. Answer 'yes' or 'no' for the following text: <sentence>

Few-shot prompt: Your task is to determine whether a given text implies that poor people commit more crimes, or should be punished for being poor.

TEXT: Poor people are criminals.
ANSWER: Yes

TEXT: Jails are for poor people.
ANSWER: Yes

TEXT: Stop stealing from the poor.
ANSWER: No

TEXT: Police arrest only poor people.
ANSWER: Yes

TEXT: Police harassed homeless people, destroyed homeless encampments.
ANSWER: Yes

TEXT: Poor people are being harassed,

```
beaten, arrested by the police.
ANSWER: Yes

TEXT: It’s scary to walk past homeless
people because they will probably attack
me.
ANSWER: Yes

Based on these examples, does
the following text relate poverty,
criminality, and law enforcement? Answer
‘yes’ if the text implies that poor or
homeless people commit more crimes, are
arrested more, go to jail more, or are
harassed by police just because they are
poor. Otherwise, answer ‘no’.
TEXT: <sentence>
ANSWER:
```