ACID: On the Perception of Online Classism

Arianna Muti[1], **Elisa Bassignana**[2,3], **Amanda Cercas Curry**[4]
Federica Durante[5], **Dirk Hovy**[1] **and Debora Nozza**[1]
[1]Bocconi University, Milan, Italy
[2]IT University of Copenhagen, Copenhagen, Denmark
[3]Pioneer Center for AI, Denmark
[4]CENTAI, Turin, Italy
[5]University of Milano-Bicocca, Milan, Italy
{arianna.muti, dirk.hovy, debora.nozza}@unibocconi.it
elba@itu.dk
amanda.cercas@centai.eu
federica.durante@unimib.it

Abstract

Socioeconomic status (SES) structures social inequality and underlies class-based discrimination that is often rationalised through stereotypes expressed in public discourse. However, despite extensive research on hate speech detection in Natural Language Processing, classism detection remains an underexplored phenomenon. We introduce ACID, a cross-cultural corpus with over 1.15 million instances, to investigate classism across YouTube and Twitter from 14 English-speaking countries. We examine (i) which stereotypes are invoked towards lower-SES, (ii) whether blame for lower-SES is attributed to individuals or structural factors, and (iii) whether these people are portrayed offensively. Across platforms, explanations are predominantly framed in terms of individual responsibility. Across countries, class stereotypes consistently revolve around moralized notions of dependency, laziness, and ignorance, revealing a shared global structure of class-based stigma. Our dataset and analysis are a foundation to advance research on class-based discrimination and its representation in online discourse.

Keywords: classism, stereotypes, social media analysis, hate speech, offensiveness, framing

1. Introduction

Socioeconomic inequality is one of the defining challenges of the 21st century, as it shapes access to resources and opportunities across societies worldwide. Classism is one of the mechanisms through which social inequality is reinforced, leaning on stereotypes about social strata that are used to justify inequality (Durante and Fiske, 2017a; Jost and Banaji, 1994). For example, one persistent stereotype portrays poor people as lazy or irresponsible rather than as individuals facing structural disadvantages. This narrative shifts attention away from structural factors and instead frames inequality as the result of individual failure.

Across cultures, stereotypes about poverty often serve a similar social function: stabilising unequal systems. However, the specific content of those stereotypes varies depending on personal and cultural values and social and historical context (Terol Cantero et al., 2023; Durante et al., 2017). In Europe, earlier social and religious traditions often framed poverty as “God’s will", an inevitable condition requiring charity (Katz, 2013), whereas modern industrial societies increasingly linked poverty to individual responsibility and work ethic, giving rise to the figure of the “undeserving” poor as a dominant cultural narrative. This stereotype continues to influence Western contemporary attitudes toward class, welfare, and social inequality (Tihelková, 2015).

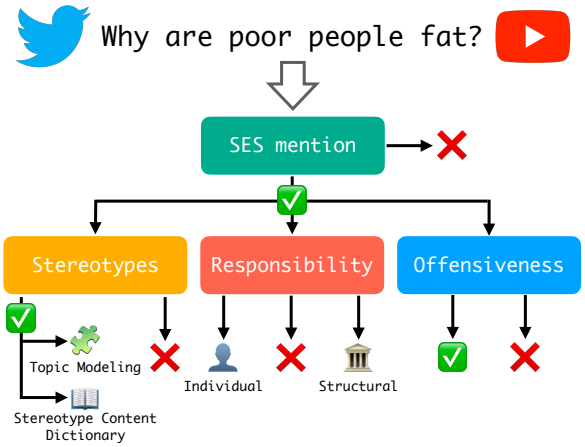


Figure 1: In ACID, every instance containing a mention to SES is annotated for offensiveness, stereotypes, and responsibility.

Stereotypes are often expressed and reproduced through everyday discourse, shaping how poverty and social class are discussed in public debates and media (Rose and Baumgartner, 2013). With the rise of digital communication platforms, these narratives are increasingly produced and circulated through large volumes of online text, creating new opportunities to study classist discourse at scale (e.g. Bose et al., 2024; Curto et al., 2024) but also new methodological needs.

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 10953–10969
11–16 May 2026.

Although classism falls within the broader scope of hate speech studied in NLP, it has received remarkably little attention to date (Cercas Curry et al., 2024b). Despite the extensive research devoted to other identity-based forms such as religion, racism, and sexism (Yu et al., 2025), classism is strikingly absent from mainstream hate speech taxonomies: Korre et al. (2024) found that only 16% of hate speech definitions explicitly include class and socioeconomic status as defining components, highlighting a significant gap in both conceptualization and detection.

To address this gap and enable research on class-based discrimination in NLP, computational social science, sociopsychology, and sociolinguistics, we introduce ACID (Automatic Classism Identification Dataset).[1]

We systematically analyze classist discourse across two different social media, YouTube and Twitter.[2] We compile a multicultural dataset of YouTube comments and Twitter posts in English across 14 countries and different cultural spheres. We manually annotate a subset of instances and combine unsupervised topic modeling with LLM-based classification to uncover stereotypes, responsibility attribution, offensiveness, and compare patterns across platforms and countries. Our analyses follow a grounded theory approach informed by sociopsychological principles based on the Stereotype Content Model (Fiske et al., 2002). Figure 1 illustrates our framework, which is designed to address the following **research questions**:

1. What stereotypes are mobilized towards lower-SES individuals? And how do they differ across countries?
2. To what extent do representations of lower-SES individuals emphasize individual responsibility versus structural causes?
3. Are stereotypes and responsibility attribution offensive?

Our findings reveal that stereotypes toward lower-SES groups consistently evoke moralized attributions of laziness, ignorance, dependency, and criminality, confirming morality as the primary dimension in the social perception. Across both platforms, explanations are predominantly framed in terms of individual responsibility. However, platform differences emerge: while YouTube discussions tend to emphasize individual blame across most countries, Twitter displays a more balanced distribution between individual and structural explanations. Country-level variation remains visible. For example, individual framing is particularly strong in the United Kingdom and the United States, two highly individualistic societies, while structural explanations appear more prominently in countries with lower individualism scores, such as India. Finally, offensiveness is high across media, especially when linked to individual blame.

To sum up, **our contributions** are as follows:[3]

We develop a theory-grounded framework for automatic classism identification across cultures and platforms, which includes the analysis of stereotypes, responsibility attribution, and offensiveness.

We compile the first multicultural Automatic Classism Identification Dataset (ACID).

2. Background

Understanding how responsibility is attributed in class-based discourse is crucial for revealing the ways classism operates in society. Media framing plays a central role in shaping these attributions. Iyengar's concept of **episodic vs. thematic framing** (Iyengar, 1991) explores how media shapes public understanding of social issues like poverty. Episodic frames focus on specific individuals or events and therefore encourage *individual attributions* of responsibility. In this type of framing, poverty or disadvantage is presented as the result of personal failure, poor choices, or individual circumstances. For example, when academic performance is explained primarily in terms of an individual's abilities or effort, the issue is framed episodically and responsibility is attributed to the individual. In contrast, thematic framing presents issues in a broader societal or historical context, and therefore encourage *structural attributions* of responsibility. Rather than focusing on isolated cases, thematic framing highlights systemic factors such as economic policy, institutional discrimination, or unequal access to resources. For instance, when differences in academic performance are explained through disparities in educational resources or socioeconomic background, the issue is framed thematically and responsibility is attributed to structural conditions.

Framing depends on the ambivalent stereotypes of deserving/undeserving lower-SES individuals. Homeless people are usually seen as undeserving who evoke contempt and neglect, while

[1] In this work, we follow the definition of classism as *the systemized discrimination of members of the lower classes due to the societal belief that having wealth is superior.* https://pct.libguides.com/anti-oppression/classism

[2] At the time of data collection, the platform was still called Twitter.

[3] We release our data and code at https://github.com/elisabassignana/acid-online-classism

disabled are seen as deserving, who merit pity and help (Fiske et al., 2002; Durante and Fiske, 2017b). Such stereotypes can influence whether audiences interpret disadvantage through an individual (episodic) or structural (thematic) lens. In terms of different media, Iyengar's research found that episodic frames are more common in television news, promoting individual blame. **In this work, we investigate whether users on social media frame class-based issues through individual responsibility or through structural explanations.**

Aporophobia vs Classism In contrast to aporophobia datasets (Curto et al., 2025; Kiritchenko et al., 2023), which focus on the stigmatization of poverty and negative attitudes toward the poor, ACID is grounded in the broader sociological understanding of classism. As Lamont (2000) shows, class boundaries are drawn not only through economic resources, but also through symbolic and cultural capital. Cultural capital exists in three distinct forms: embodied, objectified, and institutionalized. Embodied cultural capital refers to the skills, knowledge, and competencies that an individual acquires through socialization and education. Objectified cultural capital encompasses material objects and media, such as books, artworks, and instruments, that signify cultural competence. Institutionalized cultural capital is formal recognition, usually in the form of academic qualifications and credentials (Bourdieu, 1986). Drawing on Bourdieu (1986), we understand that exclusion often occurs symbolically: lower-SES groups are devalued not just for lacking economic resources, but for lacking cultural capital, for instance by possessing improper language, manners, cultural preferences, educational credentials or more generally, a limited access to resources. ACID operationalizes this notion of classism by including texts that target both the poor and the working class, directly or implicitly attacking their cultural capital or lack of resources. For instance, in the sentence *What kind of parents wouldn't provide their children with something as important as dental care?* disadvantage is framed as a moral failure of the parents rather than as the outcome of limited healthcare access. This illustrates how classism extends beyond aporophobia: rather than targeting poverty alone, it stigmatizes the behaviors, tastes, and perceived deficiencies of the lower class.

3. ACID: Automatic Classism Identification Dataset

3.1. Data Collection

We collect data from YouTube and Twitter. To capture discourse related to socioeconomic status (SES), we compile a keyword list covering different spectrum of lower-SES, including both economic disadvantage and working-class identity (e.g., *poor*, *homeless*, *on welfare*, *working class*, *blue collar*, *underprivileged*). [4]

Country	YouTube	Twitter
Australia	19,532	52,050
Canada	15,313	63,667
Ghana	2,062	5,316
India	90,786	69,921
Ireland	9,515	27,847
Kenya	4,258	13,452
New Zealand	5,445	12,028
Nigeria	7,037	62,533
Pakistan	10,095	16,140
Philippines	36,999	5,286
South Africa	43,919	35,561
Uganda	488	5,449
UK	33,461	87,491
US	226,674	194,246
Total	505,584	646,957

Table 1: Distribution of our data.

For Twitter, we use the DRAX dataset (Curto et al., 2025), which consists of English-language tweets collected between 25 August 2022 and 23 November 2022 using the query terms *the poor, poor people, poor ppl, poor folks, poor families, homeless, on welfare, welfare recipients, low-income, underprivileged, disadvantaged, lower class*. By using tweet location, they grouped tweets into the following six regions: North America, Europe, Africa, South Asia, Oceania, and Other.

For YouTube, we collected comments and metadata in September 2025. We use the YouTube Data API to retrieve videos whose titles contain both one of our selected keywords and the name of one of the 14 English-speaking countries represented in the DRAX dataset for comparison. The countries are US, Canada, UK, Ireland, Nigeria, South Africa, Kenya, Uganda, Ghana, India, Pakistan, Philippines, Australia, and New Zealand. We retrieve the top 50 search results per keyword. We retain all the comments associated with a video and filter out videos with fewer than 100 comments. We apply language identification with `fasttext-langdetect` to keep only English texts.

Table 1 shows the statistics of our dataset.

3.2. Annotation

To assess the consistency of the annotation process, we run a pilot annotation on a subset of 150 YouTube comments. The annotation was carried

[4] Full keyword list in Appendix A.

out by three annotators: two authors of this paper with expertise in and sensitivity to the topic of classism, and one Master's student who also works as a research assistant. All annotators were female, aged between 20 and 30. One annotator has a background in the humanities, while the other two have backgrounds in computer science. We compute pairwise inter-annotator agreement (IAA) between three annotators across the four levels of annotation: SES mentions (yes, no), Offensiveness (yes, no), Stereotypes (yes, no), and Responsibility (individual, structural, none). Guidelines can be found in Appendix B.

We measure agreement using Cohen's κ and show the results in Table 2. IAA is generally high for SES identification and offensiveness, with κ values in the substantial range and in some cases approaching almost perfect agreement. In contrast, stereotypes and responsibility yield more variable outcomes. For stereotypes, one comparison reaches substantial agreement, while others remain in the moderate range. Responsibility shows mostly moderate agreement, with only one case achieving almost perfect agreement.

After the pilot annotation, the three annotators met to review and discuss cases of disagreement. They compared their coding decisions, examined ambiguous comments, and resolved most discrepancies through discussion. In cases where consensus could not be reached, the final label was assigned by majority voting. One annotator is then selected to annotate 150 tweets as well.

4. Methodology

Analyzing public opinion at scale increasingly requires the ability to interpret large volumes of online discourse. Recent advances in natural language processing, particularly the rise of LLMs, offer promising tools for examining how people express views on complex societal topics (Ziems et al., 2024; Ranjit et al., 2024). First, we sample 100k instances from both datasets, balanced with respect to countries. We conduct analysis on this sample. We apply topic modeling to uncover salient themes related to socioeconomic status in each corpus, enabling a data-driven overview of how classism manifests across different platforms. Next, we use prompt-based LLM annotation to identify which stereotypes are mobilized towards lower-SES individuals (RQ1). Then, we classify whether lower-SES individuals are portrayed through individual or structural attributions, thereby quantifying the prevalence of each type of explanation (RQ2). We also investigate the intersection of stereotypes and framing with offensiveness (RQ3). To investigate cross-cultural variation, we compare stereotypes and framing patterns across countries.

To validate LLM annotations, we conduct preliminary experiments using the annotated subset of YouTube comments and tweets. Specifically, we employ one open-source and one closed-source model, OpenAI's GPT-4o-mini and Alibaba's Qwen 2.5, to perform four classification tasks: (1) detecting the presence of SES references, (2) evaluating whether such instances contain offensive language towards lower-SES, (3) identifying the presence of stereotypes and, when present, describing their type in free text, and (4) determining responsibility attribution, distinguishing between individual, structural, or no attribution. After observing that GPT-4o-mini performs best across all tasks, we exclude Qwen from the remainder of the experiments. [5] Table 3 reports GPT-4o-mini's performance on the annotated subset. The high recall for the positive class, except for Responsibility, indicates that the model captures relevant instances.

5. Class-based Discourse across Social Media Platforms

In order to observe differences in how class-based discourse is expressed in different social media, we employ BERTopic (Grootendorst, 2022), a topic-modeling framework that couples transformer-based sentence embeddings with density-based clustering to discover coherent themes. Each document is first embedded using `all-MiniLM-L6-v2`. We then feed the resulting high-dimensional vectors to HDBSCAN, which automatically determines the number of clusters and can label noise points as outliers. The corpora collected from the two social media platforms share a material core: homelessness, rent, welfare payments, and immigration, but they diverge in how those issues are framed. YouTube comments contain the richest intersectional vocabulary: race topics (Black, migrants), gender memes such as "Karen", caste terms, and religion labels (God/Jesus/Allah), alongside stigma markers that refer to appearance ("face-tattoos and jobs") and "slice-of-life" digressions about cooking or pets and crypto advice. Tweets skew toward punitive affect: clusters rely on crime lexicons ("steal, cops, jail"), sensory disgust ("smell, piss"), fear words ("terrified, scary") and moral blame ("lazy, fault").

6. Stereotypes

We extract stereotypes with a prompt-based LLM pipeline using GPT-4o-mini. We pass only texts flagged as SES-relevant by a separate SES-mention detector also based on GPT-4o-mini. For

[5] Qwen reaches an F1 of 0.79 on the positive class on the SES mention task on YouTube.

	SES identification	Offensiveness	Stereotypes	Responsibility
Ann1 vs Ann2	$\kappa = 0.798$	$\kappa = 0.833$	$\kappa = 0.458$	$\kappa = 0.555$
Ann1 vs Ann3	$\kappa = 0.761$	$\kappa = 0.794$	$\kappa = 0.797$	$\kappa = 0.866$
Ann2 vs Ann3	$\kappa = 0.765$	$\kappa = 0.740$	$\kappa = 0.518$	$\kappa = 0.563$

Table 2: Pairwise inter-annotator agreement (Cohen's κ) across annotation dimensions.

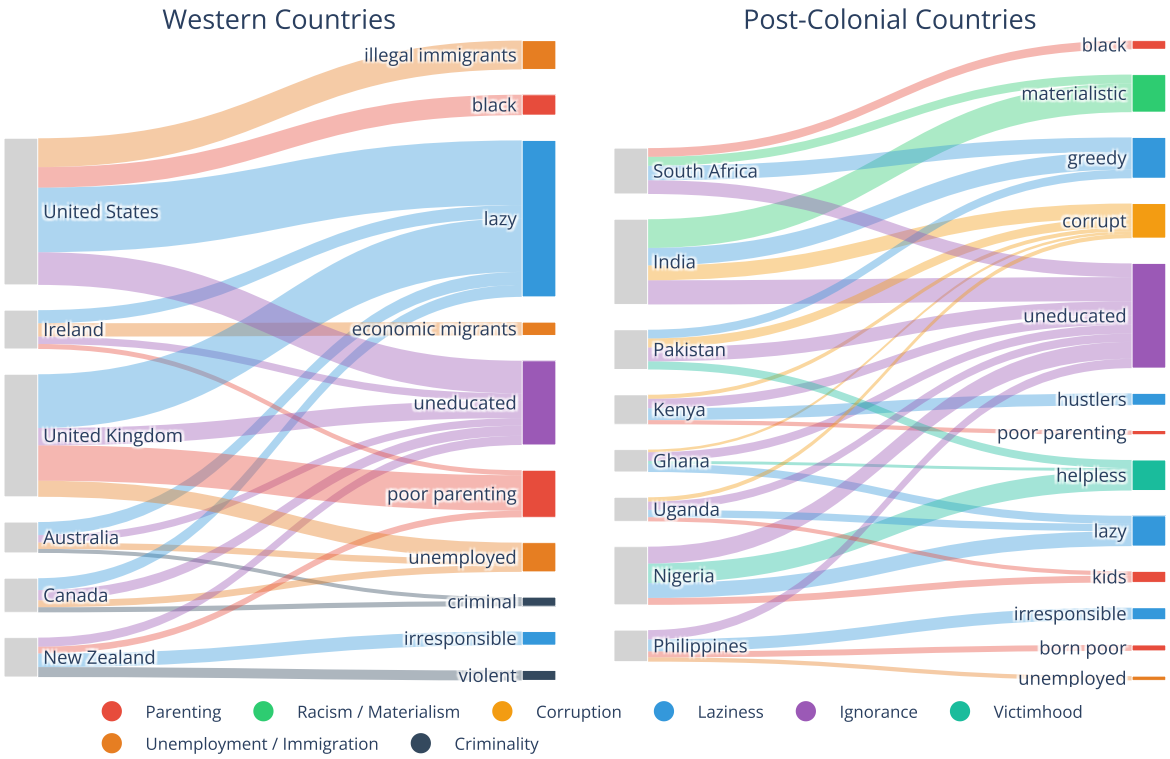


Figure 2: Top four clusters and the most frequent stereotype of that cluster by country.

Source	Task	Recall	F1
YouTube	SES mention	0.88	0.89
YouTube	Offensiveness	0.77	0.67
YouTube	Stereotypes	0.87	0.53
YouTube	Responsibility	0.63	0.54
Twitter	SES mention	0.96	0.95
Twitter	Offensiveness	0.96	0.49
Twitter	Stereotypes	0.94	0.43
Twitter	Responsibility	0.36	0.34

Table 3: Recall and F1 score on the positive class for SES mention, Offensiveness and Stereotypes; macro average for Responsibility.

stereotype retrieval, we use an instruction prompt that (i) asks whether stereotypes are present and (ii) if so, returns a free-text list of short stereotype phrases (one concept per bullet). We also include few-shot exemplars illustrating both stereotypical and non-stereotypical SES mentions. The prompt can be found in Appendix C.

Across the subset labeled as containing a SES mention by the LLM, the model identifies stereotypes in 63% of texts, confirming that a substantial portion of SES-related discourse embeds class-based generalizations. In total, 57,012 stereotypes have been identified, with 38,230 being unique. However, most of them are overly specific or semantically redundant. Therefore, we employ clustering techniques to aggregate them. Following Russo et al. (2026), we reduce the number of stereotypes in four steps: (i) we embed each stereotype using we embed each stereotype using either OpenAI's text-embedding-3-large or Word2Vec, (ii) we perform agglomerative clustering on these embeddings with cosine distance, (iii) we compute the Silhouette score on these clusters, (iv) we use gpt-4o-mini to assign each cluster a label that represents all the stereotypes in that cluster. Word2Vec yields the higher Silhouette score (0.140 with K=260), so we discard the OpenAI embeddings (0.040 with

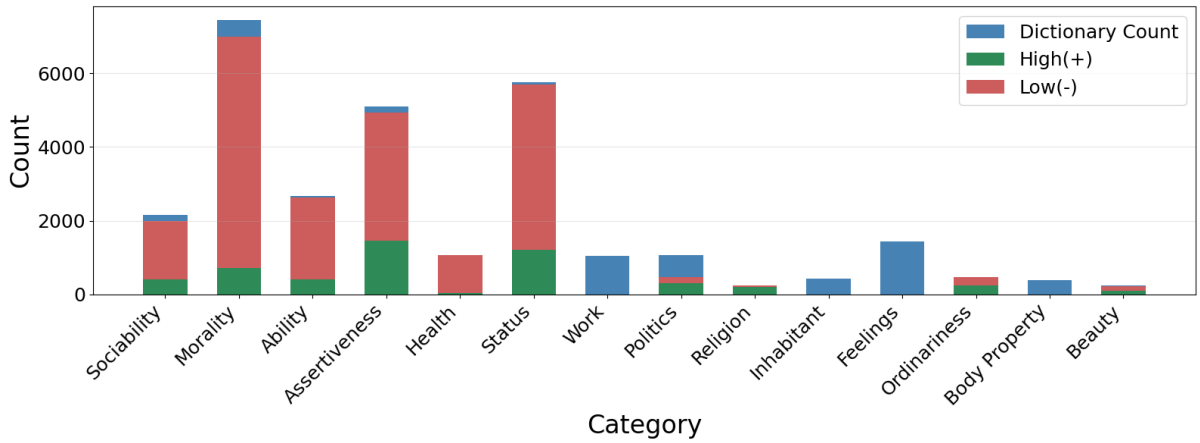


Figure 3: Stereotypes obtained with GPT-4o-mini are mapped onto the dimensions of the Stereotype Content Dictionary.

K=174). In Figure 2 we show the top 4 clusters for each country (for space reasons),[6] along with the most frequent stereotype for that cluster. In Appendix D we show the top ten clusters along with example stereotypes.

The Laziness cluster and the Ignorance cluster are nearly ubiquitous, highlighting a widespread tendency to attribute class disadvantage to individual moral failings or lack of education. These clusters represent the dominant moralizing discourse of classism, emphasizing personal deficiency rather than structural causes. The Criminality cluster emerge only in Western contexts, although it occurs also in post-colonial countries with a lower frequency. This suggest a global association between lower socioeconomic groups and moral deviance and criminality. The Oppression cluster appears primarily in India and the Philippines, reflecting local narratives around caste, agrarian identity, and colonial history. The Substance Abuse cluster appears primarily in Australia, Canada, and the United States, where poverty is frequently linked to substance use and personal irresponsibility. In contrast, the Parenting cluster is nearly universal across countries, though its framing varies from neglect and single parenthood in Western contexts to themes of generational poverty and family burden in post-colonial societies. For instance, in countries like Kenya and Uganda, references to "born poor" suggest that poverty is inherited, while in Ghana and Nigeria, stereotypes around "child marriage" and "family dependence" portray the poor as constrained by traditional obligations. However, we find some limitations in our clustering, most likely due to the embeddings encoding societal biases present in their training data, which can lead to misleading groupings. For example, embeddings conflate concepts such as "unemployed" and "immigrant," grouping them together despite their distinct meanings, amplifying pre-existing stereotypes.

Mapping Stereotypes to Stereotype Content Dictionary The Stereotype Content Dictionary (SCD) (Nicolas et al., 2021, 2022) is a lexical resource designed to map free-text descriptions of social groups into meaningful stereotype content dimensions. It was constructed via a semi-automated pipeline that began with theory-derived seed word lists, then expanded via WordNet and word-embeddings to yield a set of dictionaries covering 14,449 words, and achieved coverage of over 80% of open-ended stereotypes in a US sample. It maps words onto multiple dimensions, including Warmth (Morality, Sociability), Competence (Ability, Assertiveness), and broader domains such as Status, Beliefs, and Appearance. Each dimension is bidirectional, representing opposing poles (e.g., (+)honest (-)dishonest, (+)competent (-)incompetent). [7] We map all the stereotypes obtained onto the Stereotype Content Model (SCM) dimensions. Specifically, free-text stereotypes are matched to their corresponding SCM dimensions: *sociability, morality, ability, assertiveness, health, status, work, politics, religion, feelings, ordinariness, body property and beauty*, thereby situating the extracted stereotypes in a well-established sociopsychological framework. Out of the 38,230 unique stereotypes, 5.1% has exact phrase matches with the original word dictionary. We use GPT-5 to produce a simple one-word adjective for the remaining phrases, covering 69.68 % of our stereotype list. Figure 3 shows the dimen-

[6] The most frequent cluster, broadly about poverty, is excluded as it is overly generic and not distinctive across countries.

[7] Note that using the labels "positive" and "negative" can be misleading, as certain traits, such as violent, are coded as (+) agency but do not imply a positive valence.

sions occurring in our stereotype list with a higher frequency than 200.

Our mapping indicates that stereotypes frequently invoke negative moral judgment, low social positioning, low assertiveness and incompetence. Morality exhibits the highest overall frequency and the largest contrast between positive and negative poles, indicating that it is both the most salient and the most polarized evaluative domain. This pattern aligns with the moral primacy model (Brambilla et al., 2021), which argues that moral information serves as the most diagnostic cue for inferring others' intentions and predicting social behavior. In contrast, dimensions such as status, assertiveness, and ability play secondary roles, while sociability contributes comparatively less, reinforcing prior evidence that warmth-related impressions are primarily driven by moral traits. Moreover, the strong moral polarity observed resonates with prior findings in the context of poverty and group stereotypes, where moral attributions (e.g., laziness vs. misfortune) shape emotional and behavioral responses toward disadvantaged groups (Cozzarelli et al., 2001; Weiner et al., 2011). Overall, the quantitative pattern in our data corroborates previous literature in showing that morality dominates both in frequency and evaluative weight, serving as the core dimension through which social targets are perceived and judged.

7. Responsibility Attribution

We analyze how responsibility for classism is attributed in the comments. Specifically, we classify whether users attribute SES-related issues to individuals (e.g., personal choices, behavior, or moral failings), to structural factors (e.g., economic conditions, institutions, or inequality), or whether the attribution is unclear or mixed. Our analysis of responsibility attribution across countries reveals variation in the proportions of responses categorized as individual, structural, or none. Figure 4a and 4b show the statistics for YouTube and Twitter, respectively. On YouTube, individual responsibility is dominant across most countries, whereas structural responsibility generally appears less frequently. Structural responsibility is highest in India (54.04%), followed by Australia (43.63%), New Zealand (42.40%) and Canada (39.03%). In contrast, the lowest proportions of structural responsibility are observed in Uganda (9.09%), Kenya (11.29%), the Philippines (13.29%) and the United Kingdom (15.13%). One possible explanation relates to the digital divide: in countries where internet access remains uneven, social media users may disproportionately represent more affluent segments of the population. As a result, online discourse may reflect the perspectives of relatively privileged groups, which could contribute to a greater emphasis on individual responsibility.

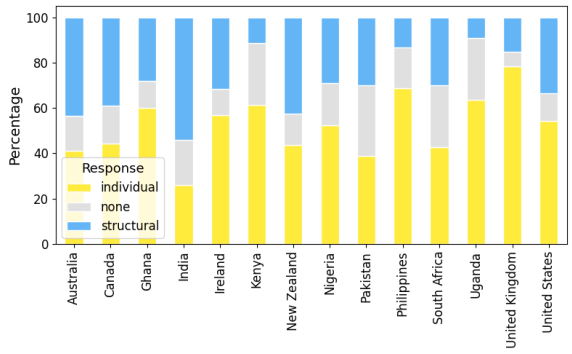
On Twitter, individual responsibility dominates across all countries, but the overall proportion of structural attributions is higher than on YouTube and displays a different cross-national pattern. Countries such as Pakistan (37.2%), New Zealand (35.8%), India (32.3%), Australia (32.5%), and Ireland (31.2%) show a comparatively strong emphasis on structural responsibility. In contrast, the United States (24.6%), the Philippines (18.1%), and Nigeria (19.1%) exhibit the lowest structural attribution rates, reflecting more individualistic framings.

These patterns can be interpreted in light of Hofstede's cultural dimensions, particularly Individualism vs. Collectivism (IDV) and Power Distance Index (PDI) (Hofstede, 1980; Hofstede et al., 2010).

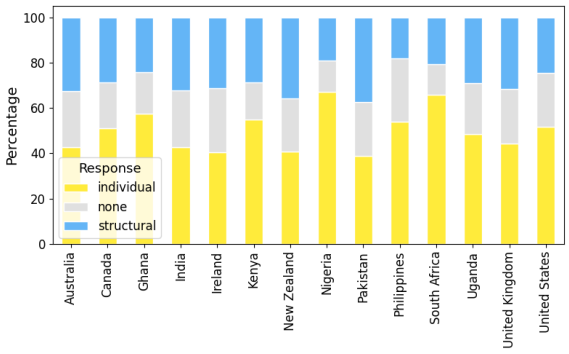
Individualism reflects the degree to which individuals are integrated into groups: in highly individualistic societies, people are expected to look after themselves and their immediate families, whereas in collectivist societies, individuals expect their ingroups to protect them in exchange for loyalty. High individualism is often associated with a tendency to attribute responsibility to personal agency rather than structural causes. The Power Distance Index further reinforces these tendencies: in high-power-distance societies, hierarchical authority is more accepted and people are more accustomed to systemic inequality. Figure 5 shows the scores for Australia, United Kingdom, India, Pakistan, South Africa and United States. The sample includes highly individualistic Western societies (e.g., United Kingdom, United States, Australia, New Zealand) as well as countries characterized by higher power distance and lower individualism (e.g., India and Pakistan), with South Africa representing an intermediate case.

To examine whether cross-national differences in responsibility attribution align with cultural values, we conducted country-level correlation analyses between responsibility attribution rates and Hofstede's cultural dimensions. Specifically, we computed Spearman's rank correlations between the proportion of structural and individual responsibility attributions on Twitter and Youtube and countries' scores on Individualism (IDV) and Power Distance (PDI).

For Twitter, Spearman's rank-order correlation indicates a moderate negative association between structural responsibility attribution and Individualism ($\rho = -0.43$), suggesting that higher levels of individualism tend to be associated with lower levels of structural attribution. Power Distance exhibits a weak positive correlation with structural attribution ($\rho = 0.20$). In contrast, individual responsibility attribution is moderately positively correlated with Individualism ($\rho = 0.60$), broadly consistent



(a) YouTube.



(b) Twitter.

Figure 4: Cross-country responsibility distribution of responsibility attribution across social media platforms.

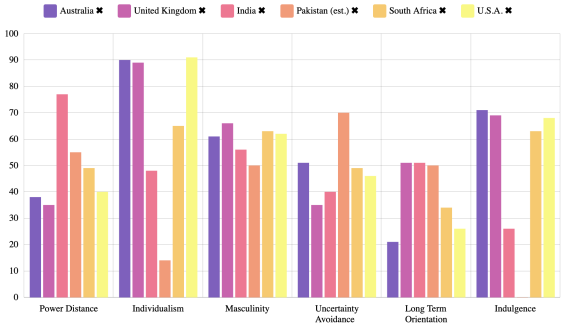


Figure 5: Hofstede scores for mentioned countries.

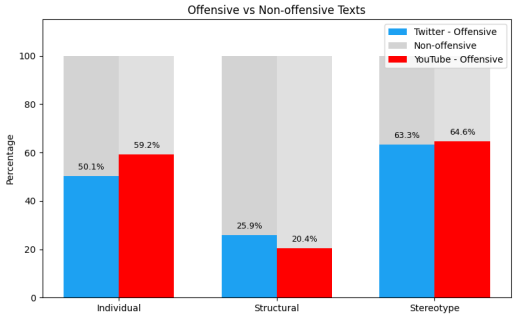


Figure 6: Offensive vs. non-offensive texts across responsibility categories and stereotypes in both platforms.

with Hofstede's theoretical expectations. Additionally, individual responsibility attribution shows a moderate negative correlation with Power Distance ($\rho = -0.43$).[8]

For YouTube, structural responsibility attribution shows no meaningful association with Individualism ($\rho = 0.14$) and a moderate positive association with Power Distance ($\rho = 0.43$). In contrast, individual

[8]Most correlations did not reach statistical significance, likely reflecting limited statistical power given the small country-level sample ($n = 6$). Results should therefore be interpreted in terms of effect size and directional tendencies rather than statistical significance.

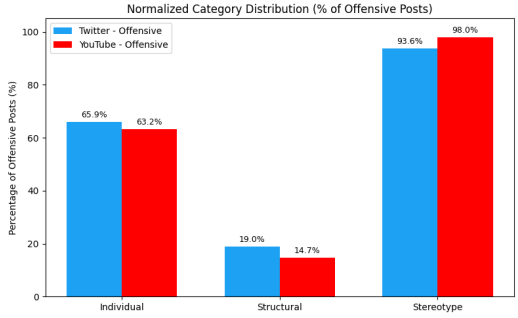


Figure 7: Contribution of each responsibility category and stereotypes to all offensive posts on both platforms.

responsibility attribution is positively correlated with Individualism ($\rho = 0.66$) and strongly negatively correlated with Power Distance ($\rho = -0.83$).[9]

To examine these cross-cultural differences in greater detail, we focus on the subset of responses classified as structural to uncover what kinds of social or systemic factors people tend to blame. We conduct this analysis by applying BertTopic. We obtain 74 topics which are listed in Appendix E. We identify four primary responsibility attribution frames across the corpus: a) Historical-structural explanations that locate inequality's origins in colonial exploitation and slavery; b) Government failure frames that attribute poverty to corrupt politicians and policy neglect; c) Market mechanism frames that critique capitalist dynamics (landlord rent-seeking, corporate subsidies, tax structures); d) Identity-based frames that organize inequality along caste, race, or immigration lines.

8. Offensiveness

YouTube exhibits a slightly higher proportion of offensive content compared to Twitter (41.6% vs

[9]This correlation reached statistical significance ($p = 0.042$).

38.4%). We compute the intersection between offensive posts, responsibility attribution and stereotypes. Figure 6 shows the proportion of offensive and non-offensive posts on Twitter (blue) and YouTube (red) across responsibility attribution and stereotypes. Overall, posts containing stereotypes tend to be offensive in 63–65% of cases, indicating that while many stereotypes are expressed in offensive ways, some are not. As for responsibility, most offensive posts assign blame to the individual, whereas structural-responsibility posts are the least offensive (20–26%), suggesting that discussions emphasizing systemic or institutional causes tend to employ more neutral and less hostile language. Fig. 7 shows the relative contribution of each responsibility category and stereotypes to all offensive posts. On both platforms, stereotypical content dominates the offensive discourse, representing around 94–98% of all offensive posts. Individual-responsibility posts occupy a middle ground (63–66%), reflecting a moderate level of hostility, while very few offensive posts place responsibility on structural factors.

Taken together, these results indicate that almost all offensive posts contain stereotypes, yet not all stereotypes are offensive, highlighting an important distinction between the presence of stereotypes and their offensive expression. Among non-offensive stereotypes we find the following top-terms: *financially unstable, dependent on welfare, homeless, uneducated*, describing conditions or social states rather than flaws. Non-offensive stereotypes contain also examples of benevolent classism (subjectively positive but condescending) (Jordan et al., 2021). An example of this is *FEED THE POOR!*, which positions the poor as dependent on others, instead of recognizing their autonomy. There are also examples of paternalistic classism (Sainz, 2024), like in the following example. *I dont care how poor you are... Plz take care of your hygiene... If you need help getting stuff let me know ill send ya some stuff.* The author expresses concern and willingness to help, yet assuming that poverty equates to neglect or incapacity, positioning the poor person as someone who needs guidance from a more capable benefactor.

9. Conclusion

This study introduces a framework for analyzing class-related discourse across platforms and countries, by combining stereotype extraction, clustering, and responsibility attribution. Our findings reveal that socioeconomic stereotypes are pervasive and predominantly negative, with morality emerging as the central evaluative dimension, followed by status and assertiveness. This pattern supports the moral primacy hypothesis, indicating that moral evaluations (e.g., lazy, criminal, irresponsible) serve as the dominant lens through which class is perceived. Across countries, responsibility for lower-SES is predominantly framed in terms of individual failings, while structural explanations appear less frequently. Individual responsibility framing is particularly strong in the UK and the US, two highly individualistic societies. Structural attributions are comparatively more visible on Twitter than on YouTube, suggesting that platform features influence how responsibility is assigned in online discussions. These findings highlight the importance of considering both cultural context and platform-specific dynamics in cross-national analyses of social media communication.

10. Related Work

Social Class in NLP Despite increasing attention to hate speech in NLP, classism remains a rarely studied phenomenon (Cercas Curry et al., 2024a). Liu (2024) introduces a continuous measure of stereotyping that captures graded biases against disadvantaged groups. Jeoung et al. (2023) propose STEREOMAP, a framework based on the Stereotype Content Model to measure how LLMs perceive social groups along dimensions like warmth and competence. Their analysis of groups such as 'rich', 'poor' and 'middle class' shows that the models reproduce class-based stereotypes: high-status groups are often rated as more competent but less warm, while marginalized groups cluster low on both dimensions. Perez Almendros et al. (2020) compile a dataset of patronizing and condescending language toward vulnerable communities (including poor families, refugees, homeless), showing that even when language expresses pity can reproduce power asymmetries tied to class. Ranjit et al. (2024) develop OATH-Frames to analyze online discourse around homeless people, pointing out how framing homelessness via crime, substance use, or "undeserving" narratives influences harmful stereotypes. Singh et al. (2024) show that most LLMs are unable to empathize with the socioeconomically underprivileged regardless of the situation. Arzaghi et al. (2024) investigate how LLMs encode socioeconomic biases, introducing a million-sentence dataset to quantify associations between demographic attributes and socioeconomic status. Their study highlights intersectional biases across models. Bassignana et al. (2025) analyze prompts used by people with a different socioeconomic background, showing that there are systematic differences in language technology usage (i.e., frequency, performed tasks), interaction styles, and topics.

Limitations

The reliance on keyword-based sampling may have excluded relevant content that does not explicitly contain the selected terms. Furthermore, all texts analyzed are English-language comments, which may not fully represent each country's sociolinguistic or cultural perspective. This linguistic constraint likely favors globally dominant discursive patterns while overlooking localized forms of class-based discourse expressed in other languages. While in some contexts English is widely used across socioeconomic strata, in others English proficiency is more strongly associated with higher socioeconomic status. In the latter case, we may disproportionately capture discourse produced by higher-SES individuals, while underrepresenting perspectives expressed in local languages that may be more prevalent among lower-SES groups. In addition, all analyses in this study rely on LLM-based predictions. Although we conducted manual validation on a small subset of the data to assess model performance, the models remain prone to classification errors across the tasks of offensiveness detection, stereotype identification, and responsibility attribution. Consequently, our findings should be interpreted as an underestimation of the real prevalence of online classism.

Ethical Considerations

This study uses publicly available social media data, which may raise privacy and consent concerns. We follow platform terms of service for the publication of our data.

Grouping cultures for analysis is a complex and nuanced issue. We compare Western and Post-colonial cultures in English-speaking contexts to capture historically distinct formations of class/status discourse shaped by colonial legacies and global economic hierarchies, but we do not intend to say these groups are homogeneous. We use these terms to describe historical relations of power, not inherent qualities of societies. We also acknowledge that most NLP resources are developed for Western varieties of English, which may bias model interpretation in post-colonial contexts.

We recognize that class hierarchies and associated stereotypes are shaped by colonial legacies, local economic structures, and cultural discourses. Western notions of class (e.g., tied to income or education) may not align with those in post-colonial societies, where class is often intertwined with ethnicity, language, or caste.

Acknowledgments

We thank the MilaNLP group at Bocconi University for feedback on an earlier version of this draft, as well as the reviewers for their helpful comments. Arianna Muti and Debora Nozza are supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 101116095, PERSONAE). Arianna Muti, Debora Nozza, and Dirk Hovy are members of the Data and Marketing Insights Unit of the Bocconi Institute for Data Science and Analysis. Elisa Bassignana is supported by a research grant (VIL59826) from VILLUM FONDEN.

References

Mina Arzaghi, Florian Carichon, and Golnoosh Farnadi. 2024. Understanding intrinsic socioeconomic biases in large language models.

Elisa Bassignana, Amanda Cercas Curry, and Dirk Hovy. 2025. The AI gap: How socioeconomic status affects language technology interactions. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 18647–18664, Vienna, Austria. Association for Computational Linguistics.

Paul Bose, Lorenzo Lupo, Mahyar Habibi, Dirk Hovy, and Carlo Schwarz. 2024. Beyond the stats: Realities, perception, and social media discourse on poverty. In *AEA Papers and Proceedings*, volume 114, pages 690–694. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Pierre Bourdieu. 1986. The forms of capital. In John Richardson, editor, *Handbook of Theory and Research for the Sociology of Education*, pages 241–258. Greenwood Press, New York.

Marco Brambilla, Simona Sacchi, Patrice Rusconi, and Geoffrey P. Goodwin. 2021. The primacy of morality in impression development: Theory, research, and future directions. In *Advances in Experimental Social Psychology*, volume 64, pages 1–66. Elsevier.

Amanda Cercas Curry, Giuseppe Attanasio, Zeerak Talat, and Dirk Hovy. 2024a. Classist tools: Social class correlates with performance in NLP. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12643–12655, Bangkok, Thailand. Association for Computational Linguistics.

Amanda Cercas Curry, Zeerak Talat, and Dirk Hovy. 2024b. Impoverished language technology: The lack of (social) class in NLP. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 8675–8682, Torino, Italia. ELRA and ICCL.

Catherine Cozzarelli, Anna V. Wilkinson, and Michael J. Tagler. 2001. Attitudes toward the poor and attributions for poverty. *Journal of Social Issues*, 57(2):207–227.

Georgina Curto, Svetlana Kiritchenko, Kathleen Fraser, and Isar Nejadgholi. 2024. The crime of being poor: Associations between crime and poverty on social media in eight countries. In *Proceedings of the Sixth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS 2024)*, pages 32–45, Mexico City, Mexico. Association for Computational Linguistics.

Georgina Curto, Svetlana Kiritchenko, Muhammad Hammad Fahim Siddiqui, Isar Nejadgholi, and Kathleen C. Fraser. 2025. Tackling social bias against the poor: a dataset and a taxonomy on aporophobia. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 6995–7016, Albuquerque, New Mexico. Association for Computational Linguistics.

Federica Durante and Susan T Fiske. 2017a. How social-class stereotypes maintain inequality. *Current opinion in psychology*, 18:43–48.

Federica Durante and Susan T. Fiske. 2017b. How social-class stereotypes maintain inequality. *Current Opinion in Psychology*, 18:43–48.

Federica Durante, Courtney Bearns Tablante, and Susan T Fiske. 2017. Poor but warm, rich but cold (and competent): Social classes in the stereotype content model. *Journal of Social Issues*, 73(1):138–157.

Susan T. Fiske, Amy J. C. Cuddy, Peter Glick, and Jun Xu. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6):878–902.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Geert Hofstede. 1980. *Culture's Consequences: International Differences in Work-Related Values*. Sage, Beverly Hills, CA.

Geert Hofstede, Gert Jan Hofstede, and Michael Minkov. 2010. *Cultures and Organizations: Software of the Mind*, 3rd edition. McGraw-Hill, New York.

Shanto Iyengar. 1991. *Is Anyone Responsible? How Television Frames Political Issues*. University of Chicago Press, Chicago.

Sullam Jeoung, Yubin Ge, and Jana Diesner. 2023. StereoMap: Quantifying the awareness of human-like stereotypes in large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12236–12256, Singapore. Association for Computational Linguistics.

Jessica A. Jordan, Joanna R. Lawler, and Jennifer K. Bosson. 2021. Ambivalent classism: The importance of assessing hostile and benevolent ideologies about poor people. *Basic and Applied Social Psychology*, 43(1):46–67.

John T Jost and Mahzarin R Banaji. 1994. The role of stereotyping in system-justification and the production of false consciousness. *British journal of social psychology*, 33(1):1–27.

Michael B Katz. 2013. *The undeserving poor: America's enduring confrontation with poverty: Fully updated and revised*. Oxford University Press.

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

Katerina Korre, Arianna Muti, Federico Ruggeri, and Alberto Barrón-Cedeño. 2024. Untangling hate speech definitions: A semantic componential analysis across cultures and domains.

Michèle Lamont. 2000. *The Dignity of Working Men: Morality and the Boundaries of Race, Class, and Immigration*. Harvard University Press.

Yang Liu. 2024. Quantifying stereotypes in language. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1223–1240, St. Julian's, Malta. Association for Computational Linguistics.

Gandalf Nicolas, Xuechunzi Bai, and Susan T. Fiske. 2021. The stereotype content dictionaries: A semi-automated tool to quantify the warmth and competence of words, groups, and texts. *Behavior Research Methods*, 53(6):2609–2627.

Gandalf Nicolas, Xuechunzi Bai, and Susan T. Fiske. 2022. A spontaneous stereotype content model: Taxonomy, properties, and prediction. *Personality and Social Psychology Bulletin*, 48(12):1678–1695.

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.

Jaspreet Ranjit, Brihi Joshi, Rebecca Dorn, Laura Petry, Olga Koumoundouros, Jayne Bottarini, Peichen Liu, Eric Rice, and Swabha Swayamdipta. 2024. OATH-frames: Characterizing online attitudes towards homelessness with LLM assistants. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 13033–13059, Miami, Florida, USA. Association for Computational Linguistics.

Max Rose and Frank R Baumgartner. 2013. Framing the poor: Media coverage and us poverty policy, 1960–2008. *Policy Studies Journal*, 41(1):22–53.

Giuseppe Russo, Debora Nozza, Paul Röttger, and Dirk Hovy. 2026. The pluralistic moral gap: Understanding judgment and value differences between humans and large language models. In *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics.

Mario Sainz. 2024. Identifying hostile versus paternalistic classism profiles: a person-based approach to the study of ambivalent classism. *Current Psychology*, 43(33):27316–27326.

Smriti Singh, Shuvam Keshari, Vinija Jain, and Aman Chadha. 2024. Born with a silver spoon? investigating socioeconomic bias in large language models.

M Carmen Terol Cantero, Maite Martin-Aragón Gelabert, Carolina Vázquez Rodríguez, Ana Lledó Boyer, and Jose Enrique García Soler. 2023. Perceptions of poverty in spain: differences in the attitudinal profiles between women and men. *Frontiers in Psychology*, 14:1229685.

Alice Tihelková. 2015. Framing the'scroungers': the re-emergence of the stereotype of the undeserving poor and its reflection in the british press. *Brno studies in English*, 41(2):121–139.

Bernard Weiner, Danny Osborne, and Udo Rudolph. 2011. An attributional analysis of reactions to poverty: The political ideology of the giver and the perceived morality of the receiver. *Personality and Social Psychology Review*, 15(2):199–213.

Zehui Yu, Indira Sen, Dennis Assenmacher, Mattia Samory, Leon Fröhling, Christina Dahn, Debora Nozza, and Claudia Wagner. 2025. The unseen targets of hate: A systematic review of hateful communication datasets. *Social Science Computer Review*, 43(5):1114–1144. © The Author(s) 2024.

Caleb Ziems, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. 2024. Can large language models transform computational social science? *Computational Linguistics*, 50(1):237–291.

A. Keyword List

We used the following keywords to collect texts referencing socioeconomic status (SES). The same list was applied across YouTube, X, and news sources:

benefits, blue collar, blue-collar, caste, classism, disadvantaged, economically disadvantaged, economically insecure, educational level, eviction, financial background, hobo, homeless, homelessness, impoverished, income, low status, low-income, lower class, marginalized groups, occupation, on welfare, welfare, poor, poor people, poor ppl, poor folks, poor families, precarious workers, prestige, renters, SES, low social class, low socioeconomic, low socioeconomic, struggling families, subsistence-level, the 99 percent, the have-nots, the poor, the unemployed, uneducated, underprivileged, unskilled labor, wage earners, welfare recipients, working class, working-class.

B. Classism Annotation Guidelines

These guidelines were designed to ensure consistency and conceptual clarity in the manual annotation of classist discourse across social media. They operationalize classism as a form of prejudice and discrimination based on socioeconomic status (SES), and they were used both to train human annotators and to validate large language model (LLM) predictions. Each text in the dataset is annotated for four dimensions: *SES Identification*, *Offensiveness*, *Stereotypes*, and *Responsibility Attribution*. These categories are grounded in sociological and social-psychological theory, reflecting how socioeconomic inequality is represented, evaluated, and moralized in online discourse.

B.1. A. SES Identification

Annotators first determine whether the comment explicitly or implicitly discusses socioeconomic status.
Question: Does this comment explicitly or implicitly discuss socioeconomic status (SES)? **Options:** `yes` / `no`
Examples:

- **No:** "In the Holy Bible scriptures it states to care for the poor, widows, elderly, foreigners and the sick."
- **Yes (implicit):** "Please share the location where you are giving out food—we want to send some there."

B.2. B. Offensiveness

If the comment contains an SES reference, annotators assess whether it is offensive toward individuals from low socioeconomic backgrounds.
Question: Do you find this text offensive toward individuals from low socioeconomic backgrounds? **Options:** `yes` / `no`
Examples:

- **No:** "This is just the poor areas."
- **Yes:** "Maxine could do loads of jobs well because she has inner strengths she didn't know she had, and Dean can do well too if he leaves behind the stereotypical loser with two Staffies appearance. He was like a no-hoper at the start, but emerged as a decent young lad willing to try to change for the better. I hope they're both doing well these days."

B.3. C. Stereotypes

If SES is mentioned, annotators evaluate whether the text expresses a stereotype about low-SES individuals. A *stereotype* is a generalized assumption about a group—in this case, lower-SES people. A sentence can still express a stereotype even when referring to an individual, if the attribution reflects a generalization about low SES rather than evidence about that specific person.
Question: Does this text contain stereotypes toward individuals from low socioeconomic backgrounds? **Options:** `yes` / `no`
Examples:

- **No:** "Ok, if you're homeless why the f would you move to Winnipeg—one of the coldest cities in the world?"
- **Yes (group):** "Why do those people always drop out of school and have so many kids they can't afford?"
- **Yes (individual):** "Of course you didn't finish high school."

B.4. D. Responsibility Attribution

For texts that mention SES, annotators determine whether responsibility for socioeconomic disadvantage is attributed to the individual (episodic framing), to social structures (thematic framing), or to neither.
Question: How is responsibility for SES issues attributed? **Options:** `individual` / `structural` / `none`
Definitions:

- **Episodic framing:** Focuses on individual behavior, personal choices, or moral failings (e.g., laziness, bad decisions).
- **Thematic framing:** Focuses on systemic or institutional causes (e.g., inequality, government policy, economic conditions).
- **None:** No clear attribution of responsibility.

Examples:

- **Individual:** "Bro, if you're still broke in 2025 it's because you're lazy and waste all your cash on vape and tattoos."
- **Structural:** "The welfare system keeps creating bums who sit around forever—thanks, government, for making a whole generation useless."
- **None:** "This video is depressing, can't believe people live like that."

C. Prompts

We report the prompts used to classify comments across the four tasks in our pipeline: (1) detection of socioeconomic status (SES) mentions, (2) offensiveness toward low-SES individuals, (3) stereotype detection, and (4) responsibility attribution.

C.1. SES Mention

> Does this comment explicitly or implicitly discuss socioeconomic status? Answer like in the examples. Do not provide explanations. Only output "yes" or "no".
>
> Examples:
> Text: In the Holy Bible scriptures it states to care for the poor, widows, elderly, foreigners and the sick.
> Answer: no
>
> Text: Please the location of where you are sharing the food I mean we want to send some there
> Answer: yes

C.2. Offensiveness

You are a classifier. Your task is to determine whether a given text contains offensive, derogatory, or negative stereotypes specifically directed toward people from low socioeconomic backgrounds.

If the text contains offensive or disparaging language, stereotypes, or implications targeting such individuals, answer "yes". If the text does not contain such content, answer "no".

Do not provide explanations. Only output "yes" or "no".

Examples:
Text: This is just the poor areas
Answer: no

Text: Maxine could do loads of jobs well because she has inner strengths she didn't know she had, and Dean can do well too if he leaves behind the stereotypical loser with two Staffies appearance. He was like a no-hoper at the start, but emerged as a decent young lad willing to try to change for the better. I hope they're both doing well these days.
Answer: yes

C.3. Stereotype Detection

Task: Read the following text. Determine whether it contains a stereotype about people from low socioeconomic status (SES). If yes, write a list of the stereotypes being expressed in free text. Each stereotype should be one word or short phrase.

Steps: 1. Decide if the text expresses stereotypes. (yes / no) 2. If it does, write a list of the stereotypes in free text (e.g., "lazy", "criminal", "uneducated", "dependent on welfare").

Output format:

Answer: yes / no
Stereotypes list:
- [stereotype 1]
- [stereotype 2]
- [stereotype 3]
Example:

Text: poor kid I hope they survive
Answer: no

Text: why those people always drop out of school and have so many kids they cant afford
Answer: yes
Stereotype list:
- uneducated
- have many kids

C.4. Responsibility Attribution

Read the following text. Decide whether responsibility for SES issues is attributed to the individual (episodic framing), to society/structural factors (thematic framing), or if it is unclear/mixed.

Definitions:

- Individual attribution (episodic framing): focuses on individuals, their behavior, personal choices, and moral failings (e.g., laziness, bad decisions).

- Structural attribution (thematic framing): focuses on social systems, institutions, or structural causes (e.g., inequality, government, economy).

- None: contains both or does not clearly assign responsibility.

Output format: Responsibility attribution: individual / structural / none

Examples:

Text: General concept of poverty. Lack of job or laziness.
Answer: individual

Text: Great informative video. It is disgraceful that we cannot provide homes for people. We need to use the Scandinavian model of houses first and then the support structures. The supports include people to teach others how to look after their homes, as well as drug therapy, mental health supports. How can you beat addiction if you live on the streets?
Answer: structural

Text: Can't even afford an rv park, nowadays in a cheaper state. Anyone that lives in CA is sick. No way i'd do that to myself. Its for the elite. Just drive to the South.
Answer: none

D. Top Stereotypes Clusters by Country

Table 4 shows the top 10 stereotype clusters by country, along with the most two frequent words for that cluster for that country.

E. Topic Modeling of Structural Responsibility

We employ BERTopic to extract topics, name of topics and representative documents automatically. Then, we pass the results to gpt-5 to label the topics and write a one-sentence explanation of the topic. Table **??** shows the results. Our topic model identified 75 distinct topics (including the outlier topic -1), revealing a highly structured discourse around responsibility attribution for class-based inequality. The distribution is heavily skewed, with the three largest topics accounting for 39.2% of all documents, while 42 topics contain fewer than 50 documents each.

We identify four primary responsibility attribution frames across the corpus:

Historical-structural explanations (topics 1, 2, 7) attribute contemporary inequality to colonial exploitation, slavery, and resource extraction. These topics invoke historical events (British rule in India, apartheid in South Africa, the transatlantic slave trade) as root causes that continue to structure present-day wealth distribution.

Government failure and corruption (topics 3, 4, 11, 12, 53, 59, 64) locate responsibility in contemporary state institutions. These topics critique politicians for "corruption," "self-enrichment," and "neglect," framing poverty as the result of policy choices. The discourse often adopts a moralistic tone, expressing "outrage" that governments "allow children to suffer" despite available resources.

Market mechanisms and elite (topics 9, 15, 16, 18, 21, 34, 67, 71) attribute inequality to capitalist dynamics, particularly rent-seeking behavior by landlords, real-estate speculators, and corporations. Topic 16 specifically targets "greedy landlords" for "rent profiteering," while topics 15 and 34 expose "tax cuts for the rich" and "corporate welfare" as mechanisms of wealth concentration.

Identity-based explanations emphasize how inequality is organized along lines of caste (topic 0), race (topics 2, 3), or immigration status (topics 8, 20, 26, 54). These topics highlight group-based discrimination and resource competition.

Country	Top Clusters (examples)
Australia	Poverty (poor, dependent on welfare), Laziness (lazy, irresponsible), Ignorance (uneducated, sick), Unemployment (unemployed, immigrants), Criminality (criminal, violent), Parenting (poor parenting, hardworking mums and dads), Victimhood (helpless, victim), Corruption (corrupt, exploited), Substance Abuse (alcohol abuse, drunk), Racism/Materialism (racist, victimhood)
Canada	Poverty (poor, homeless), Laziness (lazy, irresponsible), Ignorance (uneducated, sick), Unemployment (unemployed, unskilled), Criminality (criminal, violent), Substance Abuse (drug use, addicted), Parenting (poor parenting, old), Victimhood (helpless, victim), Corruption (corrupt, exploited), Elitism (elitist, plebs)
Ghana	Poverty (poor, homeless), Ignorance (uneducated, useless), Laziness (lazy, greedy), Victimhood (helpless, suffering), Corruption (corrupt, exploited), Parenting (child marriage, ordinary), Criminality (criminal, violent), Unemployment (unemployed, unskilled), Racism/Materialism (materialistic, materialism), Theft (stealing, looting)
India	Poverty (poor, dependent on welfare), Racism/Materialism (materialistic, jehadis), Ignorance (uneducated, illiterate), Laziness (greedy, lazy), Corruption (corrupt, exploitative), Oppression (oppressed, farmers), Parenting (kids, minorities), Unemployment (unemployed, jobless), Criminality (criminal, violent), Theft (looting, frauds)
Ireland	Poverty (poor, homeless), Unemployment (economic migrants, unemployed), Laziness (lazy, ungrateful), Ignorance (uneducated, physically handicapped), Parenting (poor parenting, child), Criminality (criminal, violent), Racism/Materialism (racist, racial bias), Victimhood (helpless, victim), Corruption (corrupt, exploitative), National Identity (irish, irish homeless)
Kenya	Poverty (poor, homeless), Laziness (hustlers, lazy), Ignorance (uneducated, unhappy), Parenting (poor parenting, born poor), Corruption (corrupt, exploited), Victimhood (helpless, suffering), Criminality (criminal, thieves), Racism/Materialism (materialistic, dependent on religion), Theft (stealing, steal), Unemployment (unemployed, jobless)
New Zealand	Poverty (poor, dependent on welfare), Laziness (irresponsible, lazy), Criminality (violent, abused), Ignorance (uneducated, brainwashed), Parenting (poor parenting, old), Victimhood (helpless, vulnerable), Unemployment (unemployed, locals), Elitism (elitist, fascist), Corruption (corrupt, exploitative), Oppression (peasants, persecuted)
Nigeria	Poverty (poor, dependent on charity), Victimhood (helpless, suffering), Ignorance (uneducated, illiterate), Laziness (lazy, greedy), Parenting (kids, poor parenting), Masses (masses, poor masses), Corruption (corrupt, easily manipulated), Criminality (criminal, criminals), Racism/Materialism (materialistic, classism), Illness (sickly, dirty)
Pakistan	Poverty (poor, poor people), Ignorance (uneducated, depressed), Corruption (corrupt, corruption), Laziness (greedy, dishonest), Victimhood (helpless, victim), Criminality (criminal, thieves), Unemployment (laborer, unemployed), Theft (looting, looters), Parenting (poor parenting, common people), Elitism (elite, elite class)
Philippines	Poverty (poor, tragedy for the poor), Laziness (irresponsible, ignorant), Ignorance (uneducated, untrustworthy), Parenting (born poor, single mom), Unemployment (unemployed, unskilled labor), Corruption (exploited, exploitative), Racism/Materialism (materialistic, materialism), Victimhood (helpless, victim), Criminality (criminal, violent), Oppression (oppressed, slaves)
South Africa	Poverty (poor, homeless), Laziness (greedy, irresponsible), Ignorance (uneducated, useless), Racism/Materialism (materialistic, racial bias), Parenting (black, poor parenting), Unemployment (unemployed, jobless), Criminality (criminal, violent), Victimhood (helpless, victim), Corruption (corrupt, exploitative), Theft (looting, steal)
Uganda	Poverty (poor, dependent), Ignorance (uneducated, untrustworthy), Laziness (lazy, irresponsible), Corruption (corrupt, corruption), Parenting (kids, poor parenting), Victimhood (helpless, vulnerable), Criminality (thieves, violent), Unemployment (unemployed, unemployment), Theft (stealing, steal), Mindset (victim mentality, mindset)
United Kingdom	Poverty (poor, working class), Laziness (lazy, irresponsible), Parenting (poor parenting, old), Ignorance (uneducated, unqualified), Unemployment (unemployed, illegal immigrants), Criminality (criminal, criminals), Racism/Materialism (racist, materialistic), Poverty-Slang (skint, dross), Victimhood (helpless, vulnerable), Corruption (corrupt, exploitative)
United States	Poverty (poor, homeless), Laziness (lazy, irresponsible), Ignorance (uneducated, mentally ill), Unemployment (illegal immigrants, illegals), Parenting (black, poor parenting), Criminality (criminal, crime), Substance Abuse (drug addicts, drug users), Racism/Materialism (racist, racism), Corruption (corrupt, exploitative), Wastefulness (irresponsible spending, wasteful)

Table 4: Top 10 stereotype clusters by country on both YouTube and Twitter. Cluster names shown with example stereotypes.

ID	N	Key Terms	Topic Label and Description
-1	2606	people, homeless, housing, rent	-
0	1792	caste, india, based	**Caste hierarchy:** Caste-based discrimination, reservation, and religious stratification.
1	1680	india, british, indian, poor	**British colonialism:** Economic exploitation that impoverished India.
2	428	white, black, south, africa	**Racial inequality in SA:** Post-apartheid divides and land redistribution debates.
3	209	australia, australian, govt	**Australian inequality:** Housing unaffordability and cost-of-living crises.
4	198	homeless, money, problem	**Homelessness funds:** Accusations of government corruption and inefficiency.
5	190	israel, billions, wars	**Wars and foreign aid:** U.S. spending abroad while citizens remain in poverty.
6	177	nz, zealand, new	**NZ housing crisis:** Foreign buyers and ineffective government regulation.
7	125	homeless, help, streets	**Compassion for unhoused:** Urges empathy, rejects moralizing poverty.
8	122	canada, trudeau, canadian	**Canadian housing:** Links shortages to immigration and political neglect.
9	109	landlords, rent, eviction	**Landlords and evictions:** Tenant rights and moral responsibility.
10	105	philippines, manila, filipino	**Philippine poverty:** Slum poverty and gendered exploitation.
11	104	children, help, sad, govt	**Children and neglect:** Moral outrage at children suffering in poverty.
12	103	nigeria, nigerian, africa	**Nigerian corruption:** Chronic poverty from government and foreign intervention.
13	99	housing, affordable, build	**Affordable housing:** Calls for public housing and de-commodification.
14	93	states, red, blue, state	**Red–blue inequality:** Poorer red states rely on wealthier blue states.
15	83	tax, billionaires, cuts	**Tax injustice:** Tax cuts for rich deepen inequality.
16	73	landlords, rent, tenants	**Rent profiteering:** Landlords as exploiters causing rent hikes.
17	70	ubi, income, basic, universal	**Universal Basic Income:** Guaranteed income as humane safety net.
18	69	rich, richer, poor	**Rich–poor divide:** Frustration at widening gaps between elites and workers.
19	68	wage, minimum, hour	**Minimum wage:** Living wages, inflation, and wage-based solutions.
20	65	uk, immigrants, housing	**Immigration and housing:** Perceived favoritism toward refugees over citizens.
21	60	estate, real, buying, market	**Real-estate speculation:** Corporate investors commodifying homes.
22	58	working, class, party	**Working-class politics:** Disillusionment with political elites.
23	57	welfare, mothers, education	**Welfare and family:** Dependency, education, and gendered structures.
24	56	capitalism, socialism	**Capitalism vs. socialism:** Critiques inequality, calls for reform.
25	56	philippines, pesos, salary	**Low Philippine wages:** Low salaries and cost-of-living struggles.
26	54	irish, ireland, private	**Irish housing tension:** Migrants receive aid while citizens face scarcity.
27	53	jobs, job, work, retail	**Post-pandemic work:** Job scarcity, low pay, and disillusionment.
28	48	healthcare, health, universal	**Universal healthcare:** Healthcare as a human right.
29	47	san, francisco, homeless	**SF homeless aid:** Debates stipends under "Care Not Cash."
30	46	tariffs, canada, canadian	**Canada's economy:** Post-COVID debt, tariffs, redistributive politics.
31	44	apartment, rent, bedroom	**Rising rent:** Unaffordable rent, wage stagnation, housing shortages.
32	39	richest, country, world	**Homelessness in wealth:** Contrasts national wealth with homelessness.
33	37	trump, maga, musk	**Right-wing populism:** Condemns populist politics, authoritarian threats.
34	37	welfare, corporate	**Corporate welfare:** Exposes subsidies and bailouts for corporations.
35	36	benefits, mps, disabled	**Disability and austerity:** Criticizes welfare cuts, politician hypocrisy.
36	34	right, work, unions	**Right-to-work laws:** Opposes anti-union legislation.
37	33	cpp, seniors, oas, pension	**Pensions and poverty:** Higher pensions, tax fairness for elderly.
38	33	cornwall, st, austell	**Urban decay in Cornwall:** Small-town decline, drugs, government neglect.
39	28	ireland, quangos, sector	**Irish public vs. private:** Inefficiency, privatization, inequality.
40	28	london, birmingham, live	**London gentrification:** Property speculation pushing out working class.
41	27	currency, banks, debt	**Central-bank debt:** Private banking enslaves nations through debt.
42	25	pakistan, climate, flood	**Pakistan climate disasters:** Floods, climate injustice, mismanagement.
43	24	finland, housing, high	**Finland's housing:** Praises "Housing First" policy.
44	24	businesses, covid, congress	**Pandemic fraud:** Misuse of PPP loans, political corruption.
45	23	negative, gearing, property	**Negative gearing:** Ban tax incentives inflating property prices.
46	23	richest, country, debt	**Debt in richest country:** Paradox of U.S. wealth and national debt.
47	22	paycheck, homeless	**Paycheck precarity:** Most Americans one paycheck from homelessness.
48	21	ai, ceo, executives	**AI and automation:** Automation, executive greed threaten jobs.
49	20	wv, coal, virginia, west	**Coal decline in WV:** Economic decay, mechanization, lost livelihoods.
50	20	manufacturing, china, jobs	**Offshoring:** Attributes unemployment to outsourcing.
51	18	ireland, irish, dublin	**Ireland's dysfunction:** Local bureaucracy, greed worsening housing shortage.
52	18	status, peterson, social	**Social engineering:** Social control, class hierarchy, gender oppression.
53	17	politicians, elites, money	**Corrupt politicians:** Self-enrichment, lobbyist influence, neglect of poor.
54	16	immigration, housing, crisis	**Immigration pressure:** Population growth exacerbates housing crises.
55	16	child, care, daycare, single	**Childcare issues:** Lack of affordable daycare, hardship for single mothers.
56	16	mexico, oklahoma, new	**Poverty in NM/OK:** Generational poverty, structural neglect, underfunded education.
57	16	family, planning, abortion	**Family planning:** Links overpopulation to lack of reproductive health education.
58	16	trump, elon, cut	**Funding cuts:** Predicts rising homelessness as welfare funds are reduced.
59	15	profit, govt, health	**Government greed:** Corruption perpetuates poverty and illness for profit.
60	14	florida, desantis, welfare	**Florida welfare:** Policies making welfare access difficult for vulnerable citizens.
61	14	daily, aur, ke, hai	**Daily-wage hardship:** Economic stagnation, inequality for day laborers in India.
62	14	ni, tax, er, week	**U.K. taxation:** Excessive income-tax and national-insurance deductions.
63	13	homeless, population, 2025	**Growing homelessness:** Predicts exponential homelessness due to inequality.
64	13	hain, hai, pakistan, hay	**Pakistani corruption:** Systemic corruption and elite control over poor.
65	13	canada, homes, built	**Canada housing solutions:** Advocates for publicly built homes, coordinated aid.
66	13	rent, half, income, room	**Rent consuming income:** Renters spending most earnings on housing.
67	12	elon, musk, guy	**Musk and subsidies:** Billionaires benefit from state subsidies, exploitation.
68	12	disabled, disability, security	**Disability struggles:** Disabled people denied aid and support.
69	12	poor, country, judge	**Poverty hypocrisy:** Wealth concentration in "poor" countries.
70	12	sydney, rental, afford	**Sydney rental crisis:** Unaffordable rents, shortage of social housing.
71	12	farmers, subsidies, welfare	**Farm subsidies:** Agricultural subsidies as hidden corporate welfare.
72	12	shops, shop, town	**Decline of local shops:** Empty town centers, online retail, stagnation.
73	10	rides, 600, san, month	**High SF costs:** Unaffordable living costs, declining incomes in Bay Area.
74	10	housing, given, free	**Free housing debate:** Debates fairness of public housing allocation.

Table 5: Topic modeling results for structural responsibility obtained with BERTopic and labeled with GPT-5.