Quantifying Social Norms and Anxiety in Social Media Text

Anonymous ACL submission

Abstract

Social norms can induce anxiety within members of a society when they feel pressured to conform. While researchers have previously examined the psychological impact of specific norms or norms collectively, little is known about how different types of norms vary in association to anxiety. In this work, we propose a framework to extract and categorize social norms and their sources (norm drivers) from social media using large language model (LLM). We conduct a human evaluation to assess the reliability of LLM annotations on the obtained categories and systematically examine the relationship between different norm types, norm drivers, and the users' levels of anxiety. Our findings reveal that romantic partners and norms concerning physical appearance are most strongly linked to anxiety. We share the norm types, norm drivers, their rankings in association to anxiety, and the social norms extraction tool to help advance the study of social norms found through language.

1 Introduction

Social norms are standards of acceptable behavior shared by social groups (Chung and Rimal, 2016). While they can contribute to the overall stability of society (Bicchieri et al., 2018) as a framework for increasing the predictability of people in different situations (Kiesler, 1973), they also make members of society feel anxious from the perceived pressure to meet the expectations of norms (Elster, 1994).

In an ideal setting, anxiety induced by norms can be natural and helpful to navigate complex social landscapes and lead to social harmony and collective well-being (Petrie, 2002). However, social norms, especially those that function as subjective culturally-specific rules can be distorted to promote harmful behaviors (Amiot et al., 2013) which are at odds with one's wishes or desires (i.e., cognitive dissonance) (Balestrino and Ciardi, 2008) or even stigmatize people (Norman et al., 2008). Such downsides of social norms can lead to excessive anxiety, creating a persistent state of distress that negatively impacts one's mental health (Wong et al., 2017; Frost et al., 1990) and daily functioning (Ferguson and Rodway, 1994). While the literature in social psychology is extensive, it mostly studies specific norms and little is known on differences in norms in terms of their effects on anxiety. LM-based encoding techniques along with development of more robust language-based assessments of anxiety (Kjell et al., 2023) can provide a valuable window into the connection between anxiety and social norms.

In this work, we propose an approach that extensively explores social norms expressed in social media and their relations to anxiety. We specifically pay attention to the social expectations that an individual gets from the people that exercise influence over that person (Kemper, 1966) comprising behaviors and manners expected for oneself and within a relationship with another, i.e., interpersonal norms. We extract such expectations and the entities that impose the norms, or **norm drivers** (Legros and Cislaghi, 2020), from Reddit posts, categorize them, and annotate each instance accordingly by prompting large language models (LLM). We then predict the Reddit users' levels of anxiety using a language-based prediction model (Son et al., 2023; Mangalik et al., 2024) and examine how different types of norms and their sources vary in association with anxiety.

Our **contributions** include: (1) proposal of approaches for extraction and categorizations of social norms and norm drivers from social media; (2) human assessment of LLM annotations to validate the labeling reliability, and (3) a ranking of the connection from different norm types and norm drivers with anxiety. We release the norm types, drivers, and their rankings in relation to anxiety along with the social norms extraction tool to help facilitate future work in the area.

2 Related Work

Social norms exist on a spectrum, from widely accepted common sense such as *"Cover your mouth when you sneeze"* or *"Be quiet when watching a movie in a theater"* to subjective and culturally influenced rules such as *"Prioritize family over work"* or *"Study hard and go to a prestigious university"* Latter encompass interpersonal expectations (e.g., providing support to romantic partner) (Ohbuchi et al., 2004), and self-oriented obligations often shaped by others (e.g., pressure to study or pursue career goals). It is primarily these latter that contribute to anxiety, as they involve pressures shaped by close relationships and cultural contexts (Hur et al., 2009).

Researchers in the area of social psychology have studied the specific norms that belong to this latter range and their impact on anxiety, including stigma on unemployment (Staiger et al., 2018), workaholic culture (Andreassen et al., 2016), academic pressure (Kumaraswamy, 2013), gender roles (Mahalik et al., 2003), marriage expectations (Gui, 2023), and beauty standards (Dakanalis et al., 2014).

Studies in the fields of ML and NLP have also explored social norms in various directions, such as detection of social roles (Beller et al., 2014; Kim et al., 2016) or stigma (Straton et al., 2020) from social media, and identifying (Park et al., 2021) or analyzing (Moon et al., 2023) norm violations within online communities, and integrating norms into (Forbes et al., 2020) or measuring norms of language models (Yuan et al., 2024). Rai et al. (2024) studied the cultural differences in the expression of shame and pride between the United States and India. Nonetheless, a gap still remains in that studies tend to focus on individual types of norms or treat them as a whole. Our work addresses this by comprehensively exploring social norms expressed in social media, summarizing them into distinct categories, and analyzing their connection to anxiety.

3 Dataset

We collected Reddit posts from subreddits that represent language usage from a variety of ethnic and cultural backgrounds, including `r/AsianParentStories`, `r/asianamerican`, `r/KoreanAmerican`, `r/ABCDesis`, `r/Hispanic`, `r/NativeAmerican`, `r/italianamerican`, `r/Blackpeople`, and `r/blackladies`. We also examine subreddits for demographics that we deem are likely to deal with social norms or expectations, regardless of ethnicity, such as `r/family`, `r/teenager` and `r/firststgenstudents`.

By using the extraction method described in the following section, the posts are filtered to those containing norm phrases, resulting in 17,448 posts authored by 11,958 Reddit users. We utilized this set of posts to define the categories of the norms and investigate the variance of the prevalence of each norm by culture. We also collected the posts that the same set of users wrote outside of the selected ethnic subreddits to estimate their baseline level of anxiety.

4 Method

Extraction of Norm Phrases and Drivers We first applied coreference resolution using a modified version of AllenNLP [1] model [2] to the collected posts to replace personal pronouns with their corresponding entities, excluding first- and second-person pronouns. We then filter posts containing specific linguistic patterns indicative of perceived social norms, such as `[expect|want|tell|force|allow] me to VB` and `let me VB`, (i.e., **norm patterns**), using regular expressions. Each post is split into sentences, from which we extract the verb phrases as **norm phrases** and their preceding subjects as **norm drivers** using constituency parsing from Stanza library (Qi et al., 2020). For example, given the sentence *"My friends want me to hang out with them"*, the norm driver is *"My friends,"* and the norm phrase becomes *"hang out with my friends"* after resolving the pronoun *"them."*

Categorizing Social Norms and Norm Drivers To identify types of social norms from our dataset, we use LLooM (Lam et al., 2024), an LLM-based text analysis tool that generates semantically coherent, human-interpretable concepts from large text corpora. Unlike traditional topic modeling or clustering methods, which often rely on surface-level lexical features and produce groupings that require extensive manual interpretation, LLooM produces higher-level conceptual summaries that align more closely with human perceptions. While not all generated concepts are immediately usable, making decisions on keeping or combining useful topics and discarding irrelevant ones still enabled

[1] https://github.com/allenai/allennlp-models

[2] Proposed by Neurosys: https://neurosys.com/blog/effective-coreference-resolution-model#article-2

more efficient and principled topic derivation from text clusters.

One pitfall of LLooM is its lack of scalability, it performs best when generating concepts from a few thousand texts at most, whereas our dataset of extracted norm phrases exceeds this scale. To address this, we first drop the samples where either the norm driver or norm phrase is parsed to be empty or the norm driver is "please" or "thanks" and convert all norm phrases to lowercase. Then we prepend "not" to norm phrases that entail negated norm patterns (e.g., *"doesn't want me to", "refuses to let me"*). This resulted in 17,448 unique phrases. We compute the frequency of each phrase and apply weighted random sampling to select 2,000 representative phrases.

To structure the input for LLooM, we embed the sampled phrases using Twitter-RoBERTa-base (Barbieri et al., 2020) and apply KMeans clustering to partition them into 10 roughly similar groups, providing thematically narrowed subsets to facilitate concept generation. Then LLooM is applied to each cluster to generate norm type candidates. Conceptually overlapping topics were manually merged, and those deemed less relevant to social norms were discarded. The final set consists of 12 norm types along with their classification criteria, as detailed in Table 2. We would like to note that while VERBAL OR PHYSICAL ABUSE may not represent a social norm in the conventional sense and is rather heterogeneous compared to other categories, due to the nature of our data collection we observed a high frequency of expressions such as *"(told me to) kill myself"* or *"(told me to) fuck off"*. Given their prevalence and relevance to interpersonal expectations and harm, we chose to include this norm type in our schema.

We adoped a simpler approach for categorizing norm drivers given their lower diversity. We first asked ChatGPT to group norm drivers mentioned at least 10 times into broad entity types. We then re-framed these categories to emphasize the relationship between each entity and the first-person author of the post, i.e., MY PARENTS, MY ROMANTIC PARTNERS, MY FRIENDS, AND PEERS. Mentions of entities not directly related to the author (e.g., *'his parents'*, *'their friends'*) were also classified as the GENERAL PEOPLE OR OTHERS type. A first person's family members other than parents, such as siblings, grandparents, aunts, and uncles, were merged into MY NON-PARENT FAMILY MEMBERS due to their comparatively trivial role in imposing norms. NON-HUMAN OR ABSTRACT is a category introduced to capture subjects of the sentences like *"[my job] allows me to have work-life balance"* or *"[a family emergency] that kind of forced me to stay at home"*. While such entities may not be norm drivers in the strictest sense, we included them in our analysis rather than arbitrarily removing language patterns that express external pressures. ETC. comprises informal words or fragmented tokens such as interjections (e.g., *'ah'*), abbreviations (e.g., *'idk'*), or numbers, which arise from the challenges of parsing noisy social media text.

Annotation We prompted GPT-4.1-mini for annotation of norm types and norm driver types.

For norm types, annotation was conducted using a pair consisting of a norm phrase and its surrounding norm sentence. While the norm phrase alone ideally provides enough information to determine the norm type, it can sometimes be ambiguous or underspecified. In such cases, we instructed the LLM instructed to refer to the norm sentence for additional context. For example, in the pair (*"eventually get married"*, *"my parents want me to eventually get married"*), the phrase reflects expectations related to ROMANTIC RELATIONSHIPS. In contrast, in the pair (*"come over"*, *"my friend told me to come over"*), the norm phrase alone is ambiguous, but the norm sentence clarifies that the example falls under SOCIAL RELATIONSHIPS.

Anxiety Prediction For each user's most recent post containing an expression of norms, we collected posts written by the same user outside the aforementioned subreddits, selecting those that appeared immediately or after the norm post based on temporal proximity. We continued collecting until there were at least three posts with a total word count of 500 or more, and the number of users was reduced to 7,733 as a result. This decision reflects our treatment of the user's anxiety associated with social norms as a state rather than a trait.

We also collected posts from users whose writing in the selected subreddits did not contain any norm statements, following a similar procedure by retrieving their most recent posts instead.

We then predicted the level of anxiety for each user by applying a pre-trained anxiety weighted-lexicon (Son et al., 2023; Mangalik et al., 2024) on the frequencies of the words comprising the collected posts. The lexicon was originally trained on a source domain of Facebook language along-

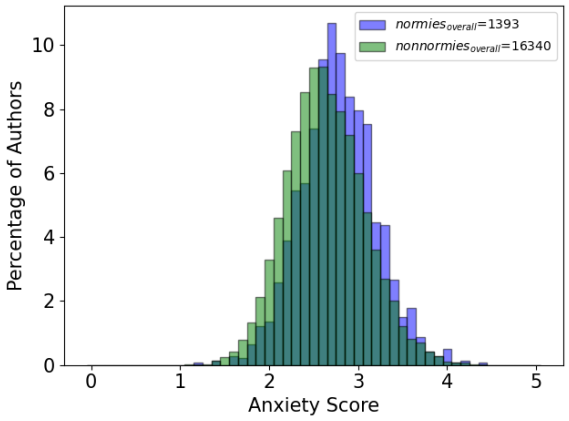

Figure 1: Histogram of anxiety scores for $normies_{\text{overall}}$ and $nonnormies_{\text{overall}}$.

side assessments of anxiety and adapted to the target domain of 2019–2020 Twitter to control for domain-specific language effects.

5 Evaluation

Association Between Social Norms and Anxiety We grouped users based on how frequently they mentioned social expectations. Specifically, we define $normies_{\text{overall}}$ as the 1,393 users whose posts contained references to social norms or expectations at least three times. In contrast, $normies_{\text{overall}}$ includes 5,936 users who mentioned norms once or twice, as well as 10,000 additional users randomly sampled from those who posted at least once in the selected subreddits but never wrote norm statements.

We further divided $normies_{\text{overall}}$ into two subsets, $normies_{\text{specific}}$ and $normies_{\text{other}}$. A user belongs to the former if they mentioned a specific norm type, norm driver type, or participated in a specific ethnic subreddit; all others were labeled with the latter.

Then we compute Cohen's d using the following equation,

$$d = mean\left(\zeta \text{anx}_{group_1}\right) - mean\left(\zeta \text{anx}_{group_0}\right)$$

where ζ denotes the z-score (mean-centered, standardized) of a user's predicted level of anxiety, and $(group_1, group_0) = (normies_{\text{overall}}, nonnormies_{\text{overall}})$ or $(group_1, group_0) = (normies_{\text{specific}}, normies_{\text{other}})$. The results are shown in Table 1.

We obtained an effect size of 0.317 when comparing the anxiety scores of $normies_{\text{overall}}$ to $nonnormies_{\text{overall}}$, This indicates a modest relationship, suggesting that the presence of social norm expressions in one's writing can be meaningfully linked with elevated anxiety levels. In other words, social norms may not be the sole or most dominant driver of anxiety, but their influence is non-negligible.

From comparing the types of norm drivers with respect to their association with anxiety, we observed that MY ROMANTIC PARTNERS and MY FRIENDS AND PEERS are ranked the highest, followed by MY NON-PARENT FAMILY MEMBERS and MY PARENTS. While it is surprising to see that the most frequently mentioned entities are not the most correlated with anxiety, such result aligns with prior findings that individuals experience higher anxiety in romantic relationships, followed by friendships, and the least in family relationships (Kamenov and Jelić, 2005). This may be pertinent to the differences in perceived relational stakes. That is, while parents regularly communicate norms and expectations to their children, the typically stable nature of parent-child relationship may make children less worried about going against them. In contrast, as romantic relationships and friendships are formed and maintained by choice (Khullar et al., 2021; Newcomb and Bagwell, 1995), they are more dependent on ongoing approval and more prone to breaking apart in the face of conflict, which may lead individuals to feel more anxious about failing to meet their expectations.

Among the norm types, APPEARANCE AND PRESENTATION shows the highest association with anxiety. While it is difficult to clearly understand the reason behind this outcome, we can conjecture that norms around physical appearance, such as being told to stay skinny, dress a certain way, or conform to beauty standards, are often pervasive and have a strong impact on self-confidence (Irving, 1990). This result is also reasonable given that MY ROMANTIC PARTNERS are most relevant to anxiety among norm drivers, as such norms are also closely tied to dating and romantic relationships where appearance tends to carry more value (Swami et al., 2021; Rollero, 2022).

INNER DEVELOPMENT AND MENTAL HEALTH shows the second strongest association with anxiety, which is natural given that many statements in this category (e.g., *"Everyone tells me to move on"*, *"My dad told me to will myself to be less depressed"*, *"My parents let me get therapy suddenly"*) imply

Overall	d	$normies_{\text{overall}}$			
	.317	1,393			
Norm Type	d	$normies_{\text{specific}}$	**Norm Driver Type**	d	$normies_{\text{specific}}$
APPEARANCE & PRESENTATION	.169	109 (8%)	MY ROMANTIC PARTNERS	.207	59 (4%)
INNER DEV. & MENTAL HEALTH	.151	231 (17%)	MY FRIENDS & PEERS	.176	50 (4%)
FAMILY DYNAMICS	.073	1,006 (72%)	MY OTHER FAMILY MBRS	.081	273 (20%)
ROMANTIC RELATIONSHIPS	.067	160 (11%)	MY PARENTS	.025	1,246 (90%)
ACADEMIC PURSUIT	.054	296 (21%)	AUTHORITY FIGRS / PROFNLS	-.029	53 (4%)
PHYSICAL HEALTH	.009	189 (14%)	GENERAL PEOPLE / OTHERS	-.055	559 (18%)
NOT A NORM	.002	470 (34%)	NON-HUMAN / ABSTRACT	-.120	288 (21%)
VERBAL / PHYSICAL ABUSE	-.005	428 (31%)	ETC.	-.171	39 (3%)
FINANCIAL PLANNING	-.020	111 (8%)			
CULTURAL INFLUENCE	-.032	186 (13%)			
SOCIAL RELATIONSHIPS	-.055	167 (12%)			
CAREER DECISIONS	-.101	274 (20%)			
INDEPENDENCE & AUTONOMY	-.168	562 (40%)			
Subreddit	d	$normies_{\text{specific}}$			
`r/family`	.202	337 (24%)			
`r/AsianParentStories`	-.074	989 (71%)			
`r/ABCDesis`	-.152	79 (6%)			
`r/asianamerican`	-.286	13 (1%)			
`r/teenagers`	-.717	27 (2%)			
`r/Hispanic`	×	0 (0%)			
`r/KoreanAmerican`	×	0 (0%)			
`r/Blackpeople`	×	0 (0%)			
`r/blackladies`	×	0 (0%)			
`r/NativeAmerican`	×	0 (0%)			

Table 1: Cohen's d of predicted anxiety scores between $normies_{\text{specific}}$, or Reddit users who mentioned specific norm types, norm driver types, or participated in specific subreddits, and $normies_{\text{other}}$.

that the user is navigating emotionally difficult or stressful circumstances.

FAMILY DYNAMICS show a higher correlation with anxiety than ROMANTIC RELATIONSHIPS, which may seem contradictory to our finding that ROMANTIC PARTNERS exhibit a stronger connection to anxiety than family members. This can be explained by the fact that the 97 users whose norm statements about ROMANTIC RELATIONSHIPS imposed by family members have lower predicted level of anxiety than the 26 users imposed by MY ROMANTIC PARTNERS, diluting the overall association of this type with anxiety.

Unfortunately, comparing the connection between social norms and anxiety across different ethnicities proved challenging, as half of the selected subreddits were relatively small in size and did not yield users that qualified for analysis. $normies$ from `r/family` have shown to be the most anxious, followed by those from `r/AsianParentStories`, `r/ABCDesis`, and `r/asianamerican`, the subreddits that represent Asian demographics. We also observed the level of anxiety for the $normies$ from `r/teenagers` to be significantly lower than the rest, likely because this subreddit mostly features meme posts rather than venting. Despite its large size (3.2 million members), it may be a less suitable place for identifying anxious users expressing external pressures.

Human Evaluation of LLM Annotation We randomly selected 25 samples that were labeled each norm type or norm driver type and asked two human judges to evaluate whether they agreed that such samples fall into the categories. The average percentages of their agreements to the annotations and inter-annotator agreements computed via Cohen's κ are recorded in Table 2 for norm types and Table 3 for the norm drivers. We would like to clarify that κ of 0 does not indicate a complete lack of agreement but rather comes from one evaluator responding 'yes' to all samples, resulting in zero variation in their responses and making the measure

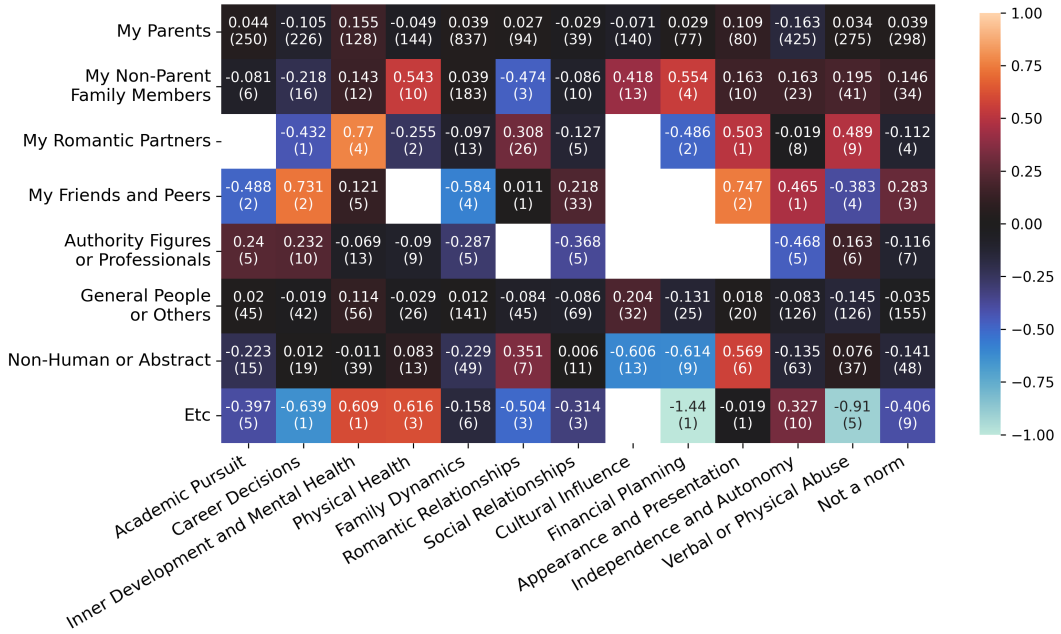

Figure 2: Heatmap of Cohen's d focused on anxiety level of $normies_{\text{specific}}$ whose norm statements indicate each norm type imposed by each norm driver type compared to $normies_{\text{other}}$.

uninformative.

While evaluators generally showed high agreement with the LLM's classification of norm drivers, they were less consistent for norm types. This is attributable to several factors, such as the inherent difficulty in defining clear boundaries between the norm types, LLM occasionally deviating from the annotation criteria, and variability in human judges' interpretations of the samples in relation to norms. We discuss these issues in detail in the Limitations section.

6 Conclusion

Social norms can cause anxiety when individuals perceive pressure to conform. While prior research has explored the psychological impact of specific norms or norms collectively, a gap remains in how different types of norms vary in their relation to anxiety. We developed a framework to extract and categorize social norms and norm drivers from social media using LLM. We conducted a human evaluation to assess the reliability of LLM annotations and analyzed the association between different norm types, norm drivers, and the users' levels of anxiety. APPEARANCE AND PRESENTATION among norm types and MY ROMANTIC PARTNERS among norm drivers are revealed to be most strongly linked to anxiety. We share the norm types, norm drivers, their rankings in association to anxiety, and the social norms extraction tool to help support future research on the complex relationship between social norms found through language and mental health.

Ethics Statement

We anticipate our study of social norms to offer valuable insights into the expectations shaping individuals' behaviors that can be observed in online communities. By identifying the types and frequency of norm expressions, this work can contribute to an enhanced understanding of psychological burden imposed by certain norms. Such comprehension could support mental health care by helping clinicians identify harmful internalized social norms that contribute to conditions.

At a broader level, our findings may help institutions in making effort to promote mental health and foster supportive environments by identifying norms and expectations that are inducing anxiety and providing interventions.

We also acknowledge potential risks. For instance, the same tools and findings could be misused to target individuals or communities with manipulative advertising or political messaging. We

therefore emphasize the need for responsible use of these methods and maintain caution in how insights are applied.

Limitations

Our study has several key limitations. First, since our data is limited to English, social norms expressed in other languages (Popitz, 2017) or in cultures and societies other than English-speaking ones (Heinrichs et al., 2006) may not be fully represented in our findings.

We abstracted diverse social norms and expectations into 12 types, which, while allowing for a structured analysis, may obscure meaningful distinctions between subtypes that differ in their relationship to anxiety. For example, ROMANTIC RELATIONSHIPS includes both expectations directed toward romantic partners (e.g., showing affection, spending quality time), and parental expectations for the individual to eventually marry. These subtypes may carry different emotional implications and levels of psychological pressure, yet are grouped under the same level.

In addition, the boundaries between norm types are not always clear-cut. The statement "My dad wants me to major in computer science" could be categorized under ACADEMIC PURSUIT if focusing solely on studying the field, but it may also fall under CAREER DECISIONS if interpreted as pressure to choose a major with better job prospects.

Low inter-annotator agreement is partly attributable to the judges having different interpretations toward social norms. For example, in the statement *"My sister really misses me and wants me to come home"*, one interpreted this as the sister's expectation for physical proximity as a family (Simola et al., 2023), which is an interpersonal norm that can be commonly found across a society. The other saw this as a personal interaction rather than a norm. These discrepancies arose partly due to trying to keep the prompt concise to avoid confusion for LLM, while human evaluators referred to the same guidelines for judgement. Having a unified annotation guide for the validity of evaluation thus inevitably involved a trade-off between prompt specificity and inter-annotator agreement.

Norm statements are often difficult to classify when the surrounding context is limited. For instance, *"(want me to) lose weight"* could either reflect concerns about obesity which would fall under PHYSICAL HEALTH, or imply pressure surrounding diet culture relevant to APPEARANCE AND PRESENTATION, depending on the surrounding context.

Another limitation lies in the LLM annotation process. Despite providing explicit instructions to prioritize the content of the norm phrase over the sentence, the model often focused did otherwise. For instance, given a pair of phrase and sentence (*"cut my hair"*, *"My mom won't let me cut my hair"*), the intended label was APPEARANCE AND PRESENTATION, whereas the model assigned INDEPENDENCE AND AUTONOMY by paying attention to *"let me"* rather than the core action. This highlights a recurring challenge with LLMs deviating from annotation criteria (Tan et al., 2024).

Furthermore, LLMs occasionally interpret statements that diverge from human understanding of norms, likely due to their limited understanding of social and cultural contexts (Ziems et al., 2024; Choi et al., 2023; Havaldar et al., 2023; V Ganesan et al., 2023). While human evaluators considered the pair (*"do things"*, *"My dad forces me to do things"*) an instance of INDEPENDENCE AUTONOMY, LLM classified it as VERBAL OR PHYSICAL ABUSE, likely due to emotionally charged verbs like *force.*

Despite these challenges, our research design was the most practical and effective strategy given the scope of this study. Our dataset comprises a wide variety of social norms and includes tens of thousands of instances, making manual annotation less feasible under time and resource constraints. The use of LLMs enabled large-scale annotation that reasonably approximated human perceptions of social norms, facilitating a systematic analysis of norm expressions across a diverse set of themes and contexts with respect to anxiety. They have proven to be valuable in social science research (Dey et al., 2024; Bail, 2024), and their growing influence in this domain highlights the importance of integrating their capabilities with care while acknowledging their limitations.

References

Catherine E Amiot, Sophie Sansfaçon, and Winnifred R Louis. 2013. Investigating the motivations underlying harmful social behaviors and the motivational nature of social norms. *Journal of Applied Social Psychology*, 43(10):2146–2157.

Cecilie Schou Andreassen, Mark D Griffiths, Rajita Sinha, Jørn Hetland, and Ståle Pallesen. 2016. The

relationships between workaholism and symptoms of psychiatric disorders: A large-scale cross-sectional study. *PloS one*, 11(5):e0152978.

Christopher A Bail. 2024. Can generative ai improve social science? *Proceedings of the National Academy of Sciences*, 121(21):e2314021121.

Alessandro Balestrino and Cinzia Ciardi. 2008. Social norms, cognitive dissonance and the timing of marriage. *The Journal of Socio-Economics*, 37(6):2399–2410.

Francesco Barbieri, Jose Camacho-Collados, Luis Espinosa Anke, and Leonardo Neves. 2020. TweetEval: Unified benchmark and comparative evaluation for tweet classification. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1644–1650, Online. Association for Computational Linguistics.

Charley Beller, Rebecca Knowles, Craig Harman, Shane Bergsma, Margaret Mitchell, and Benjamin Van Durme. 2014. I'm a belieber: Social roles via self-identification and conceptual attributes. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 181–186, Baltimore, Maryland. Association for Computational Linguistics.

Cristina Bicchieri, Ryan Muldoon, and Alessandro Sontuoso. 2018. Social norms. *The Stanford encyclopedia of philosophy*.

Minje Choi, Jiaxin Pei, Sagar Kumar, Chang Shu, and David Jurgens. 2023. Do LLMs understand social knowledge? evaluating the sociability of large language models with SocKET benchmark. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 11370–11403, Singapore. Association for Computational Linguistics.

Adrienne Chung Adrienne Chung and Rajiv N Rimal Rajiv N Rimal. 2016. Social norms: A review. *Review of Communication Research*, 4:01–28.

Antonios Dakanalis, Massimo Clerici, Manuela Caslini, L Favagrossa, Antonio Prunas, Chiara Volpato, Giuseppe Riva, and MARIA ASSUNTA Zanetti. 2014. Internalization of sociocultural standards of beauty and disordered eating behaviours: the role of body surveillance, shame and social anxiety. *Journal of Psychopathology*, 20:33–37.

Gourab Dey, Adithya V Ganesan, Yash Kumar Lal, Manal Shah, Shreyashee Sinha, Matthew Matero, Salvatore Giorgi, Vivek Kulkarni, and H. Andrew Schwartz. 2024. SOCIALITE-LLAMA: An instruction-tuned model for social scientific tasks. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 454–468, St. Julian's, Malta. Association for Computational Linguistics.

Jon Elster. 1994. Rationality, emotions, and social norms. *Synthese*, pages 21–49.

Kirsten L Ferguson and Margaret R Rodway. 1994. Cognitive behavioral treatment of perfectionism: Initial evaluation studies. *Research on Social Work Practice*, 4(3):283–308.

Maxwell Forbes, Jena D. Hwang, Vered Shwartz, Maarten Sap, and Yejin Choi. 2020. Social chemistry 101: Learning to reason about social and moral norms. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 653–670, Online. Association for Computational Linguistics.

Randy O Frost, Patricia Marten, Cathleen Lahart, and Robin Rosenblate. 1990. The dimensions of perfectionism. *Cognitive therapy and research*, 14:449–468.

Tianhan Gui. 2023. Coping with parental pressure to get married: Perspectives from chinese "leftover women". *Journal of Family Issues*, 44(8):2118–2137.

Shreya Havaldar, Bhumika Singhal, Sunny Rai, Langchen Liu, Sharath Chandra Guntuku, and Lyle Ungar. 2023. Multilingual language models are not multicultural: A case study in emotion. In *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 202–214, Toronto, Canada. Association for Computational Linguistics.

Nina Heinrichs, Ronald M Rapee, Lynn A Alden, Susan Bögels, Stefan G Hofmann, Kyung Ja Oh, and Yuji Sakano. 2006. Cultural differences in perceived social norms and social anxiety. *Behaviour research and therapy*, 44(8):1187–1197.

Taekyun Hur, Neal J Roese, and Jae-Eun Namkoong. 2009. Regrets in the east and west: Role of intrapersonal versus interpersonal norms. *Asian Journal of Social Psychology*, 12(2):151–156.

Lori M Irving. 1990. Mirror images: Effects of the standard of beauty on the self-and body-esteem of women exhibiting varying levels of bulimic symptoms. *Journal of social and clinical psychology*, 9(2):230–242.

Željka Kamenov and Margareta Jelić. 2005. Stability of attachment styles across students' romantic relationships, friendships and family relations. *Review of psychology*, 12(2):115–123.

Theodore D Kemper. 1966. Self-conceptions and the expectations of significant others. *The Sociological Quarterly*, 7(3):323–343.

Thomas H Khullar, Miriam H Kirmayer, and Melanie A Dirks. 2021. Relationship dissolution in the friendships of emerging adults: How, when, and why? *Journal of Social and Personal Relationships*, 38(11):3243–3264.

Sara B Kiesler. 1973. Preference for predictability or unpredictability as a mediator of reactions to norm violations. *Journal of Personality and Social Psychology*, 27(3):354.

Sunghwan Mac Kim, Stephen Wan, and Cécile Paris. 2016. Detecting social roles in Twitter. In *Proceedings of the Fourth International Workshop on Natural Language Processing for Social Media*, pages 34–40, Austin, TX, USA. Association for Computational Linguistics.

Oscar NE Kjell, Katarina Kjell, and H Andrew Schwartz. 2023. Ai-based large language models are ready to transform psychological health assessment. *Preprint at https://doi. org/10.31234/osf. io/yfd8g*.

Narasappa Kumaraswamy. 2013. Academic stress, anxiety and depression among college students: A brief review. *International review of social sciences and humanities*, 5(1):135–143.

Michelle S Lam, Janice Teoh, James A Landay, Jeffrey Heer, and Michael S Bernstein. 2024. Concept induction: Analyzing unstructured text with high-level concepts using lloom. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–28.

Sophie Legros and Beniamino Cislaghi. 2020. Mapping the social-norms literature: An overview of reviews. *Perspectives on Psychological Science*, 15(1):62–80.

James R Mahalik, Benjamin D Locke, Larry H Ludlow, Matthew A Diemer, Ryan PJ Scott, Michael Gottfried, and Gary Freitas. 2003. Development of the conformity to masculine norms inventory. *Psychology of men & masculinity*, 4(1):3.

Siddharth Mangalik, Johannes C Eichstaedt, Salvatore Giorgi, Jihu Mun, Farhan Ahmed, Gilvir Gill, Adithya V. Ganesan, Shashanka Subrahmanya, Nikita Soni, Sean AP Clouston, et al. 2024. Robust language-based mental health assessments in time and space through social media. *NPJ Digital Medicine*, 7(1):109.

Jihyung Moon, Dong-Ho Lee, Hyundong Cho, Woojeong Jin, Chan Park, Minwoo Kim, Jonathan May, Jay Pujara, and Sungjoon Park. 2023. Analyzing norm violations in live-stream chat. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 852–868, Singapore. Association for Computational Linguistics.

Andrew F Newcomb and Catherine L Bagwell. 1995. Children's friendship relations: A meta-analytic review. *Psychological bulletin*, 117(2):306.

Ross MG Norman, Richard M Sorrentino, Deborah Windell, and Rahul Manchanda. 2008. The role of perceived norms in the stigmatization of mental illness. *Social psychiatry and psychiatric epidemiology*, 43:851–859.

Ken-Ichi Ohbuchi, Toru Tamura, Brian M Quigley, James T Tedeschi, Nawaf Madi, Michael H Bond, and Amelie Mummendey. 2004. Anger, blame, and dimensions of perceived norm violations: Culture, gender, and relationships. *Journal of Applied Social Psychology*, 34(8):1587–1603.

Chan Young Park, Julia Mendelsohn, Karthik Radhakrishnan, Kinjal Jain, Tushar Kanakagiri, David Jurgens, and Yulia Tsvetkov. 2021. Detecting community sensitive norm violations in online conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3386–3397, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Murray Petrie. 2002. Institutions, social norms and well-being. Technical report, New Zealand Treasury Working Paper.

Heinrich Popitz. 2017. Social norms. *Genocide Studies and Prevention: An International Journal*, 11(2):4.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108, Online. Association for Computational Linguistics.

Sunny Rai, Khushang Jilesh Zaveri, Shreya Havaldar, Soumna Nema, Lyle Ungar, and Sharath Chandra Guntuku. 2024. Social norms in cinema: A cross-cultural analysis of shame, pride and prejudice. *arXiv preprint arXiv:2402.11333*.

Chiara Rollero. 2022. Mass media beauty standards, body surveillance, and relationship satisfaction within romantic couples. *International journal of environmental research and public health*, 19(7):3833.

Anna Simola, Vanessa May, Antero Olakivi, and Sirpa Wrede. 2023. On not 'being there': Making sense of the potent urge for physical proximity in transnational families at the outbreak of the covid-19 pandemic. *Global Networks*, 23(1):45–58.

Youngseo Son, Sean AP Clouston, Roman Kotov, Johannes C Eichstaedt, Evelyn J Bromet, Benjamin J Luft, and H Andrew Schwartz. 2023. World trade center responders in their own words: predicting ptsd symptom trajectories with ai-based language analyses of interviews. *Psychological medicine*, 53(3):918–926.

Tobias Staiger, Tamara Waldmann, Nathalie Oexle, Moritz Wigand, and Nicolas Rüsch. 2018. Intersections of discrimination due to unemployment and mental health problems: the role of double stigma for job-and help-seeking behaviors. *Social psychiatry and psychiatric epidemiology*, 53:1091–1098.

Nadiya Straton, Hyeju Jang, and Raymond Ng. 2020. Stigma annotation scheme and stigmatized language detection in health-care discussions on social media.

In *Proceedings of The 12th Language Resources and Evaluation Conference (LREC 2020)*, pages 1178–1190. European Language Resources Association.

Viren Swami, Charlotte Robinson, and Adrian Furnham. 2021. Associations between body image, social physique anxiety, and dating anxiety in heterosexual emerging adults. *Body Image*, 39:305–312.

Zhen Tan, Dawei Li, Song Wang, Alimohammad Beigi, Bohan Jiang, Amrita Bhattacharjee, Mansooreh Karami, Jundong Li, Lu Cheng, and Huan Liu. 2024. Large language models for data annotation and synthesis: A survey. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*.

Adithya V Ganesan, Yash Kumar Lal, August Nilsson, and H. Andrew Schwartz. 2023. Systematic evaluation of GPT-3 for zero-shot personality estimation. In *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 390–400, Toronto, Canada. Association for Computational Linguistics.

Y Joel Wong, Moon-Ho Ringo Ho, Shu-Yi Wang, and IS Miller. 2017. Meta-analyses of the relationship between conformity to masculine norms and mental health-related outcomes. *Journal of counseling psychology*, 64(1):80.

Ye Yuan, Kexin Tang, Jianhao Shen, Ming Zhang, and Chenguang Wang. 2024. Measuring social norms of large language models. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 650–699, Mexico City, Mexico. Association for Computational Linguistics.

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.

Appendix

Norm Type	Inclusion	Exclusion	Average Annotation Agreement (%)	Inter-annotator Agreement (κ)
ACADEMIC PURSUIT	education in general, studying, academic responsibilities, or school-related activities	pursuing a specific career path, such as entering medical school or law school	84%	.0
CAREER DECISIONS	choosing or aspiring to a specific career or profession	improving life in general, becoming a better person, or general educational goals such as studying or focusing on schoolwork	90%	.779
INNER DEVELOPMENT AND MENTAL HEALTH	emotional well-being, psychological challenges, healing from trauma, or improving life in general	choosing or aspiring to a specific career, studying or educational goals, taking care of physical needs, or family dynamics or pressure	88%	.0
PHYSICAL HEALTH	maintaining or improving physical condition, dealing with physical illness or injury, exercise, diet, sleep, or taking care of "my" body	pursuit of a medical career, such as becoming a doctor or going to medical school, or taking care of someone else's health	88%	.254
FAMILY DYNAMICS	relationships, interactions, or expectations between "my" family members, including doing house chores or providing financial support within the family	social interactions with people outside "my" family	46%	.359
ROMANTIC RELATIONSHIPS	relationships, interactions, or expectations between romantic partners	marrying someone of a specific ethnicity, or interactions with family members, friends, co-workers, or any other non-romantic social connections	76%	.565
SOCIAL RELATIONSHIPS	relationships, interactions, or expectations between friends, co-workers, or other non-romantic social connections	interactions with family members or a romantic partner	84%	.194
CULTURAL INFLUENCE	situation where cultural beliefs or values influence a decision or behavior, such as marrying or dating someone of a specific ethnicity, learning a specific language, or prioritizing one's own cultural traditions or ethnic roots	-	84%	.118
FINANCIAL PLANNING	saving money, budgeting, or planning expenditures for "my" future	financially supporting someone else	86%	.194
APPEARANCE AND PRESENTATION	taking care of appearance or maintaining a socially expected presentation, such as dressing appropriately, wearing makeup, or conforming to beauty standards	personal hygiene for medical reasons, or working out for health	86%	.516
INDEPENDENCE AND AUTONOMY	themes of independence, self-reliance, or autonomy, including making responsible decisions or prioritizing personal needs and boundaries	-	84%	.405
VERBAL OR PHYSICAL ABUSE	being insulted, threatened, harmed, or subjected to controlling, demeaning, or violent behavior by others	internal or self-imposed pressure	56%	.677
NOT A NORM	does not belong to any of the types	-	66%	.262

Table 2: Inclusion and exclusion criteria for annotation on the types of social norms.

Norm Driver Type	Criteria	Examples	Average Annotation Agreement (%)	Inter-annotator Agreement (k)
MY PARENTS	one's own parents (biological, adoptive, step, or culturally specific)	(my) parent(s), (my) mom and dad, (my) mother and father, (my) asian parent(s) / ap(s), (my) mom, our mom, (my) asian mom / am, (my) dad, (my) father, my folks, (my) asian dad / ad, (my) asian father / af	96%	.0
MY NON-PARENT FAMILY MEMBERS	siblings, grandparents, aunts, uncles, cousins, etc., when the speaker is referring to their own family	(my) family, (my) sister(s), (my) brother(s), (my) sibling(s), (my) aunt(s), (my) uncle(s), (my) grandmother/grandma, (my) grandmother, (my) grandparent(s), (my) grandfather/grandpa, (my) sister in law / sil, (my) brother in law / bil, (my) father in law / fil, (my) mother in law / mil	94%	.648
MY ROMANTIC PARTNERS	romantic partners in a personal context	(my) husband, (my) wife, (my) partner, (my) boyfriend/bf, (my) girlfriend/gf, (my) ex	88%	.627
MY FRIENDS AND PEERS	one's own friends or peers	(my) friend(s), (my) best friend, a friend	88%	.627
AUTHORITY FIGURES OR PROFESSIONALS	people in roles of authority or professional support	my teacher, the teacher, my manager, my therapist	100%	×
GENERAL PEOPLE OR OTHERS	others' family members, friends, or partners, generic people, or general groups	you, you guys, we, they, them, everyone, someone, anyone, others, all, no one, people, your / his / her / their parents, your / his / her / their mom, your / his / her / their dad, your / his / her / their family, this woman, this girl	78%	.651
NON-HUMAN OR ABSTRACT	objects, concepts, or vague references not tied to people	it, this, that, the one, the type, something, things, anything, a job, my brain, yesterday	92%	.0
ETC.	-	ah, idk, 8, wich	70%	.719

Table 3: Definition and examples for annotation on the types of norm drivers.