Many Ways to Be Lonely: Fine-Grained Characterization of Loneliness and Its Potential Changes in COVID-19

Yueyi Jiang[1], Yunfan Jiang[2], Liu Leqi[3], Piotr Winkielman[1]

[1] University of California, San Diego
[2] Stanford University
[3] Carnegie Mellon University
{yujiang, pwinkielman}@ucsd.edu, yjiang05@stanford.edu, leqil@andrew.cmu.edu

Abstract

Loneliness has been associated with negative outcomes for physical and mental health. Understanding how people express and cope with various forms of loneliness is critical for early screening and targeted interventions to reduce loneliness, particularly among vulnerable groups such as young adults. To examine how different forms of loneliness and coping strategies manifest in loneliness self-disclosure, we built a dataset, FIG-Loneliness (**FI**ne-**G**rained **Loneliness**), by using Reddit posts in two young adult-focused forums and two loneliness related forums consisting of a diverse age group. We provided annotations by trained human annotators for binary and fine-grained loneliness classifications of the posts. Trained on FIG-Loneliness, two BERT-based models were used to understand loneliness forms and authors' coping strategies in these forums. Our binary loneliness classification achieved an accuracy above 97%, and fine-grained loneliness category classification reached an average accuracy of 77% across all labeled categories. With FIG-Loneliness and model predictions, we found that loneliness expressions in the young adult related forums were distinct from other forums. Those in young adult-focused forums were more likely to express concerns pertaining to peer relationship, and were potentially more sensitive to geographical isolation impacted by the COVID-19 pandemic lockdown. Also, we showed that different forms of loneliness have differential use in coping strategies.

1 Introduction

Feeling socially isolated or lonely is associated with social cognitive impairments, negative health consequences and even mortality (Holt-Lunstad et al. 2015; Hawkley and Cacioppo 2010). As a result of the COVID-19 pandemic, the "stay-at-home" or "shelter-in-place" orders mandated closure of schools and nonessential businesses, which has resulted in unprecedented change in the amount and type of social interactions, and concerns over increased risk for loneliness, especially among vulnerable groups such as young adults. A recent study conducted during social distancing policies found that loneliness affects young adults more than other age groups (Bu, Steptoe, and Fancourt 2020). Another study conducted with national samples in

the US found (surprisingly) no significant changes in loneliness during the pandemic lockdowns, though confirmed that young adults are especially vulnerable to loneliness (Luchetti et al. 2020). Although sizeable research has examined the prevalence and negative consequences of loneliness, it is little understood how loneliness manifests (Mushtaq et al. 2014; Cacioppo, Cacioppo, and Boomsma 2014).

Literature suggests that loneliness is a multidimensional experience, encompassing different manifestations and forms (De Jong-Gierveld and Raadschelders 1982; Cacioppo and Patrick 2008). For example, feelings stemming from the absence of intimate or romantic relationships need to be differentiated from feelings stemming from social isolation in family or friendship contexts. This literature also suggests differences due to interpretation of the situation as changeable or chronic. It is crucial to differentiate among forms of loneliness (e.g., romantic versus social loneliness, situational versus chronic loneliness) as research suggests that different sources of loneliness are associated with different forms of psychopathology (Lasgaard et al. 2011). Of different forms of loneliness, chronic loneliness should be given special attention because of its potential harmful consequences in mental health including increased risks for depression and suicidality (Perlman, Peplau, and Goldston 1984).

Notably, the persons' coping strategies for loneliness are important to examine. Active coping strategies, sometimes called problem-focused strategies, encourage the person to pursue solutions to alleviate loneliness, whereas passive coping strategies or emotion-focused strategies focus on eliminating the negative feelings associated with the source of stress. Previous studies demonstrate that active coping styles were associated with lower levels of loneliness compared to passive coping styles (Deckx et al. 2018). Note also that loneliness manifestations can fluctuate over time. The impact of the COVID-19 pandemic (a time when the forms and amount of social interactions are changed) on different loneliness manifestations remains to be answered.

Previous research has suggested that understanding online disclosure can facilitate early detection of psychological challenges (see review in Chancellor and De Choudhury (2020)). Reflecting the wide adoption of social network platforms (e.g., Twitter, Instagram and Reddit) that enable individuals to share their emotional and social ex-

periences publicly, online disclosure not only becomes an important therapeutic ingredient in improving psychological well-being (Ellis and Cromby 2012; Turner, Frankel, and Levin 1983), but also provides researchers opportunities to understand naturalistic patterns of mental states (Olteanu, Varol, and Kiciman 2017). Given that social media use has been rapidly growing and young adults constitute the majority of users (Center 2021), this raises an interesting question: how do people, especially young adults, use social media platforms to express or cope with their loneliness experiences?

Some answers were found in recent studies that examined the impacts of COVID-19 on online mental health discourse. Low et al. (2020) showed that the number of posts on Reddit mental health forums mentioning suicidality and loneliness-related topics was more than doubled. Another study examined the themes using tweets containing co-occurred words "lonely" and "COVID-19" (Koh and Liew 2020). They found that discussions on loneliness during the pandemic focused more on mental health effects of loneliness rather than community impact of loneliness from May to July 2020. Note though that these studies treated loneliness as a broad, unitary, and explicit theme. They used loneliness related keywords or forum memberships for posts' ground truth labels for loneliness and did not differentiate types, duration, and contexts of loneliness. Consequently, their selection method does not allow for other loneliness expressions with no explicit mentions of such keywords, and can result in underreporting due to the social stigma of admitting to feeling lonely (Rokach and Brock 1997).

In this study, we aim to provide *fine-grained* characterization of loneliness discourse including the forms of loneliness and coping strategies for loneliness as suggested in psychological theories and literature. To complement the selection method described above, we captured a wider range of loneliness expressions by using both human annotations and model predictions. We set out to examine three main research questions:

- Loneliness expressions: How do people express loneliness experiences in online communities? (Section 5.1)
- Coping strategies: Are different forms of loneliness associated with different coping styles? (Section 5.2)
- Impacts of the COVID-19 pandemic: How does the pandemic affect loneliness discourse? (Section 5.3)

In the analyses of loneliness expressions and the impacts of the pandemic, we will draw inferences about loneliness characterization from young adult-focused communities and communities of diverse age groups separately. We used Reddit to examine these questions. Reddit is a social media platform where people communicate on topic-specific forums, called subreddit. This platform can capture language used by people disclosing their pandemic experiences on public online communities in real-time. Because user accounts are anonymous, Reddit has also been widely used to investigate self-disclosure in mental illnesses (De Choudhury and De 2014). In addition, a Reddit post has a limit of $40,000$ characters, compared to a Tweet which has a 280-character limit and an Instagram status post which has a limit of $2,200$ char-

acters. Thus, drawing data from Reddit provides opportunities for more comprehensive content analyses.

Taking both qualitative and quantitative approaches, we characterize discussions about loneliness experience on Reddit forums and provide strong models for fine-grained loneliness classification. Specifically, we manually annotated thousands of Reddit posts that discussed loneliness experience across years (2018, 2019, and 2020). Using the labels from both human annotations and model predictions, we investigated how people used Reddit forums for different forms of loneliness expressions across different age groups, the relationship between loneliness forms and coping strategies, and explored how loneliness discussions changed since the COVID-19 pandemic. Our main contributions are three-folds:

1. We built a large human annotated dataset[1], FIG-Loneliness (**FI**ne-**G**rained **Loneliness**), that is suitable for content analysis on fine-grained loneliness (Section 3);
2. Using hierarchical distributional learning, we built strong classifiers to identify different forms of loneliness and coping strategies within online expressions (Section 4);
3. Using FIG-Loneliness and the obtained classifiers, we examined the proposed research questions with a focus on loneliness discourse from young adults, which is a group at greater risk for loneliness (Section 5).

Our study provides important insights into how loneliness experiences manifest in online discourses, which is not only helpful for improving loneliness screening tools but also for designing effective and targeted loneliness interventions for vulnerable groups, such as young adults.

2 Related Work

The current paper is motivated by the need to examine online discourses around mental health concerns through social media data, which we describe in Section 2.1. Our study is most related to two categories of work: classifying loneliness and characterizing loneliness expressions in social media. The former category can be divided into two kinds of classification problems, which are the predictions of an expression being lonely or non-lonely, and the specific forms of loneliness it expresses. The latter category is related to characterization of loneliness. We describe our loneliness classification scheme and summarize the supporting research and theories in Section 2.2 and Section 2.3.

2.1 Self-disclosure in Online Mental Health Communities

Due to the lack of access to mental health services, individuals with mental health concerns are turning to online mental health communities such as Reddit to share their emotional challenges and seek social support. The goal of such communities is to provide a "safe haven" for mental health disclosure and peer-to-peer support for stigmatized concerns (Saha et al. 2020). Previous research sug-

---

[1]The dataset can be found in https://huggingface.co/datasets/FIG-Loneliness/FIG-Loneliness

gests that self-disclosure through anonymous communications may promote better mental health outcomes. For example, Andalibi et al. studied self-disclosure of sexual abuse on Reddit and found that those who used a more private means of communication like "throwaway" accounts engaged more in seeking support than those who used identifiable accounts (Andalibi et al. 2016). Similarly, Yang et al. investigated how the use of private and public affected members' self-disclosure in an online health support group and found that negative self-disclosure in the private channels was associated more with receiving both informational and emotional support compared to the public channels (Yang et al. 2019).

As online mental health communities have become a promising venue for studying self-disclosure related to mental health concerns, researchers have used this venue to detect the presence of major depression, suicidality, schizophrenia and other mental health problems (Chancellor and De Choudhury 2020). Compared to the above-mentioned mental health concerns, loneliness, a pervasive and growing public health concern (Holt-Lunstad et al. 2015), has received relatively little attention. This paper fills in an important gap by focusing on understanding and predicting different forms of loneliness experiences in the context of online self-disclosure.

2.2 Loneliness Classification

Current research on loneliness classification is limited. Loneliness has often been reported as a negative emotion emerged from discussions surrounding mental health on Reddit (De Choudhury and De 2014; Low et al. 2020). Guntuku et al. (2019) built a Random Forests Classifier by using extracted linguistic features from tweets to predict mentions of loneliness that include the words "lonely" or "alone". It should be noted that this does not include expressions without the mentions of loneliness keywords. Instead of using keyword filters as ground truth metrics for loneliness, we provide a human annotated dataset that includes posts without mentions of loneliness related keywords. This avoids prediction bias for keywords when training a classifier and captures a wider range of loneliness expressions.

One great challenge is that the nature of loneliness is multifaceted. To our best knowledge, there has not been research on building predictive models to classify different forms of loneliness. Thus, a question emerges: How to leverage machine learning tools to understand different forms of loneliness in social media discourse? We aim to close this gap by systematically examining the forms of loneliness presented in online disclosure, and building classifiers on our dataset that contains fine-grained loneliness annotations. Our classification scheme enables us to explore the composition of different forms of loneliness in online loneliness expressions across different groups (e.g., by age demographics), and examine the relationship between specific loneliness forms and users' coping strategies. We provide more details on our fine-grained loneliness characterization below.

2.3 Characterization of Loneliness Expressions

The most relevant investigations to our study have focused on characterizing loneliness expressions using tweets. Ma-honey et al. (2019) captured loneliness discourse on Twitter containing a single term "lonely", and categorized the posts into themes. Another study conducted by Kivran-Swaine et al. (2014) qualitatively categorized loneliness expressed on Twitter in three dimensions: (i) the temporal bonding of loneliness; (ii) the context of loneliness; and (iii) interactivity (interaction with others or not) within the expression. In their study, tweets containing specific phrases such as "I'm so lonely" were selected for characterization. Both studies could potentially exclude other loneliness expressions outside the curated phrases.

Expanding upon Kivran-Swaine et al. (2014), we focus on four categories of loneliness disclosure: *duration of loneliness*, *contexts of loneliness*, *interpersonal relationships* involved in loneliness and *interaction styles* (including coping strategies) of the disclosure. We are particularly interested in these aspects because previous research noted that the temporality of loneliness and the strategies to cope with loneliness are associated with differential mental health outcomes (Perlman, Peplau, and Goldston 1984; Deckx et al. 2018). Moreover, different types of contexts and relationships meet different needs for social connection, and thus can contribute as different sources of loneliness (Cacioppo and Patrick 2008).

In our loneliness classification scheme, we annotated a given loneliness expression for different forms of loneliness. Specifically, the duration of loneliness refers to whether the loneliness experience is a transient "state" or a chronic "trait" (Perlman, Peplau, and Goldston 1984). The contexts of loneliness cover social (relationships with others), physical (the impact of environment change), somatic (physical or bodily states), and romantic (romantic relationship) domains, which are defined similarly to the coding scheme used in Kivran-Swaine et al. (2014). We further categorized the types of interpersonal relationships mentioned in the post into friendship, family, peers (classmates or colleagues), and romantic, which are not mutually exclusive from the contexts. This is similar to the Differential Loneliness Scale in Schmidt and Sermat (1983) which measures dissatisfaction with four types of relationships: family, group or community, friendships, and romantic/sexual relationships.

Finally, we used active and passive interaction seeking as a proxy for problem-focused and emotion-focused coping strategies respectively for loneliness. We classified the posts into five interaction strategies: (i) seeking advice such as checking if social rules are followed (e.g., "is it okay for me to tag along"); (ii) seeking affirmation and validation (e.g., "am I doing the right thing?"); (iii) reaching out (e.g., "want to talk to someone"); (iv) providing social support to combat loneliness[2] (e.g., "I joined a book club and feel less lonely now"); and (v) non-directed interaction such as venting and storytelling with no desire for interactions (e.g., "I'm ugly and will die alone"). We determined (i)-(iii) to represent the problem-focused (active) coping strategies and (v) to represent emotion-focused (passive) coping strategies.

---

[2]Some users shared their own loneliness experiences and coping strategies to help others, which could reflect community resilience.

3 FIG-Loneliness : A Dataset for Fine-grained Loneliness Characterization

Below we describe how we constructed FIG-Loneliness, a dataset for fine-grained loneliness characterization and subsequent model training. First, by using the Reddit's Pushshift API (Baumgartner et al. 2020), we collected all posts from two loneliness specific subreddits (`r/loneliness`, `r/lonely`) and two subreddits for young adults (`r/youngadults`, `r/college`) from 2018 to 2020. The idea behind this design is to get data on loneliness expressions not only from a wider user base but also from users who specifically belong to the young adult age group. Subreddits `r/lonely` and `r/loneliness` have over $244,000$ forum members and have a focus on discussions surrounding loneliness-related issues. For example, the homepage of `r/loneliness` sends a message that "if you are lonely enough to enter /r/loneliness in the address bar like I did, just to see if such a reddit exists, this might be the reddit for you. Say hello!". Thus, we assumed that both subreddits are communities for people who experience loneliness to post and connect, and contain users from diverse age groups. We also chose two non-loneliness specific subreddits (`r/youngadults`, `r/college`) which are created for discussing a variety of topics with a community of young adults or college-age students. After we removed posts that were deleted by the original posters or repeated entries and only retained English posts, our dataset before annotation includes $84,639$ posts from `r/lonely`, $3,382$ posts from `r/loneliness`, $3,689$ posts from `r/youngadults` and $101,751$ posts from `r/college`. To ensure meaningful content for annotation, we only consider posts with at least 25 words in the post title and body combined.

Next, we sampled data for annotations and model training. The posts used for fine-grained loneliness annotation were randomly sampled from all four subreddits. For our lonely samples, we consider all posts from the loneliness-related subreddits and only the posts that contain loneliness-related keywords (i.e., alone, lonely, lonesome, loner, loneli, loneness, isolated and left out) in `r/youngadults` and `r/college`. With this selection method, we included posts without the explicitly mentioned loneliness keywords. For our non-lonely samples, we randomly selected posts from `r/youngadults` and `r/college` that do not contain such keywords. These posts potentially do not contain loneliness expressions. The selected lonely and non-lonely samples constitute the basis of the dataset for annotation. To ensure that the sampled data is representative of the total data, we kept both ratios among the four subreddits and ratios between pre-pandemic (2018, 2019) and post-pandemic (2020) for the samples consistent with the total data.

To provide high-quality fine-grained ground truth labels for the selected posts, we had both trained undergraduate research assistants and Amazon's Mechanical Turk workers (MTurkers) with a Master certification provide annotation labels. Six research assistants labeled the sampled potential lonely posts. Each post was labeled by three of them. A posts was first labeled on whether it contains an expression
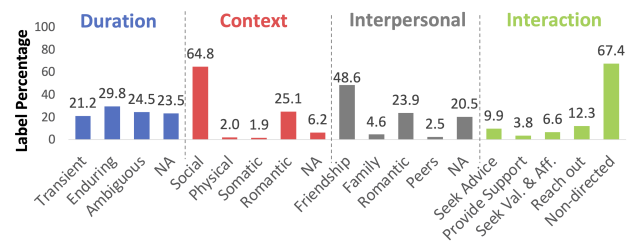

Figure 1: Distributions of labels for annotated lonely posts.

on self-disclosure of loneliness. If the majority of the annotators labeled a post as not containing such expression, the post would be discarded, otherwise it is further labeled according to a codebook that contains the following categories: (1) *duration*: the duration of the loneliness experience (transient, enduring, and ambiguous[3]), (2) *context*: the contexts of the experience (social, physical, somatic, and romantic), (3) *interpersonal*: the interpersonal relationships involved in the experience (romantic, family, friendship, and peers), and (4) *interaction*: user interaction styles (seeking advice, providing support, seeking validation/affirmation, reaching out and non-directed interaction). The codebook is intended for dissecting different forms of loneliness and users' coping strategies in the loneliness discourse. We also included a 'not applicable' (NA) label to accommodate situations that are not suitable for classification. For each category, the annotators gave one value which they think would best capture the source of loneliness in the post or the author's interaction intent. The hierarchical labeling structure of a post is illustrated in Figure 2. We explained the details of each label and the choice rationale in Section 2.3. The distributions of labels for the annotated lonely posts are shown in Figure 1. For the potential non-lonely samples, MTurkers were instructed to classify whether the posters express loneliness. Each post was annotated by three MTurkers, and only posts labeled as non-lonely by the majority would remain in the final annotated dataset.

All the labeled posts and annotations were included in FIG-Loneliness, which consists of roughly 3000 lonely and 3000 non-lonely posts. Finally, FIG-Loneliness also includes other information including the posters' demographics (e.g, gender, age and occupation) and mental health states if mentioned in the post. Consistent with our assumption, as shown in Table 1, the age groups from `r/lonely` and `r/loneliness` are more diverse than `r/youngadults` and `r/college` in the annotations for age. The per-category inter-rater similarity is shown in Figure 3.

4 Fine-Grained Loneliness Classification

We used FIG-Loneliness to build models for fine-grained loneliness classification. Our approach works with two unique features of our dataset: (1) hierarchical: as explained in Section 3, and (2) distributional: instead of being one-hot encoded, the label value for each loneliness categories (i.e.,

---

[3]"Ambiguous loneliness": annotators cannot determine if a post belongs to enduring or transient loneliness.

	% of lonely posts with age labels	before 18	18 - 25	25+
r/youngadults & r/college	28.6%	16.7%	83.3%	0
r/lonely & r/loneliness	24.5%	24.9%	47.8%	27.3%

Table 1: Age demographics of annotated lonely posts. This is an estimate of our data representation as not all annotated posts contain age information. The last three columns show the percentage of different age groups among lonely posts with age labels.
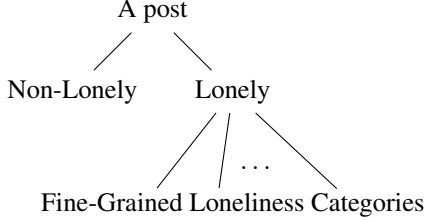
(a) Hierarchical annotations of a post.

Category	Labels
Duration	Transient, Enduring, Ambiguous, NA
Context	Social, Physical, Somatic, Romantic, NA
Interpersonal	Romantic, Friendship, Family, Peers, NA
Interaction	Seek Advice, Provide Support, Seek Val. & Aff., Reach Out, Non Directed

(b) Fine-grained loneliness categories and their labels. "Seek Val. & Aff." denotes "Seek Validation & Affirmation". NA indicates not applicable since posts can be irrelevant to such categories or labels.

Figure 2: The hierarchical labeling structure of a post and its fine-grained loneliness categories. (a) A post is first labeled as lonely or non-lonely. (b) The categories of loneliness will then be annotated with their associated labels.

Figure 3: Inter-rater agreements per category. Values range from 0 to 1, with higher values indicating greater similarity two raters' labels share.

duration, context, interpersonal and interaction) corresponds to a distribution of annotations (more details in Section 4.1).

4.1 Distributional Fine-Grained Loneliness Labels

For each loneliness category, a post might be assigned with different labels by different annotators. For example, in Figure 4, the given post was annotated by three annotators. The label value represents the fraction of the annotators who assigned the given label. In this example, the post was coded as "seeking advice" by two annotators, and "seeking validation and affirmation" by one annotator, which corresponds to $(2/3, 0, 1/3, 0, 0)$ for the label values. We do so mainly for two reasons. First, a distribution over the labels can be more informative than a single label alone. As shown in Lakoff (2008), when a unanimous decision is hard to reach, information from different perspectives is important. Second, samples with distributional labels can help train better models. Recent studies in natural-image classification, image-based diagnostic, and age estimation have shown that adopting distributional labels results in learning more robust classifiers with better generalization performance, even when using a small amount of labeled instances (Geng 2016; Peterson et al. 2019; Akbari et al. 2021).
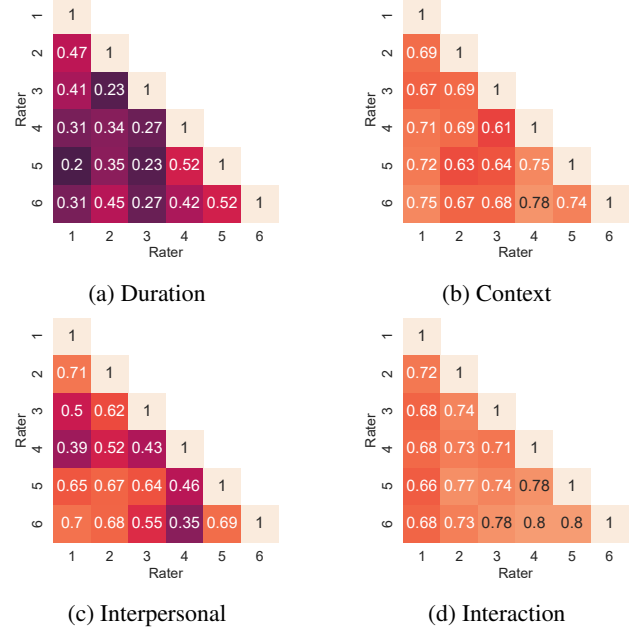
4.2 Hierarchical Distributional Learning

Our problem comprises two learning tasks: a loneliness binary classification task followed by four tasks of distributional fine-grained loneliness classification. Inspired by recent advances in distributional learning (Geng 2016) and hierarchical learning (Wehrmann, Cerri, and Barros 2018), we provide a framework for Hierarchical Distributional-label Learning (HDL). The HDL framework aims at learning the distributions of structured labels. To describe how HDL works under the setup of our problem, we introduce the following notations. Let $(\mathbf{x}^{(i)}, \mathcal{P}^{(i)})$ denote the i-th sample with $\mathbf{x}^{(i)}$ being the (tokenized) Reddit post and $\mathcal{P}^{(i)}$ being the set of labels for the post. The set of labels can be split into two parts $\mathcal{P}^{(i)} = \{\mathcal{P}^{(i)}_{lonely}, \mathcal{P}^{(i)}_{f.g.}\}$ where $\mathcal{P}^{(i)}_{lonely}$ gives the loneliness distributional label (i.e., lonely or non-lonely) and $\mathcal{P}^{(i)}_{f.g.}$ provides the fine-grained ones. $\mathcal{P}^{(i)}_{f.g.}$ contains four parts where each part $\mathcal{P}^{(i)}_c$ corresponds to the distributional label for category $c \in \mathcal{C} = \{$duration, context, interpersonal, interaction$\}$. For a non-lonely post, $\mathcal{P}^{(i)}_c$ is an all-zero vector; for a lonely post, $\mathcal{P}^{(i)}_c$ is non-negative, and its entries are summed up to 1. Finally,

"Have you ever feel bad when your friend talking about her crush? I think I am a introvert but I have being alone. I want someone beside whom I can go out or brag everything. But after I break up with my girlfriend, I feel like no one beside me anymore. And now there's that friends and they are sharing their feeling about their crush. And it make me feel something I don't know. I know I am not in love with them but hearing them talking about someone make me feel hurt. What should I do? Is there something wrong with me?"
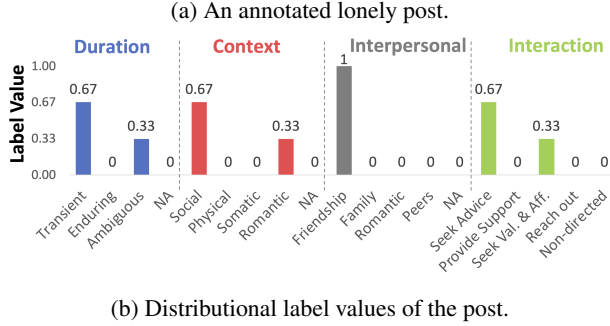
(a) An annotated lonely post.

(b) Distributional label values of the post.

Figure 4: An illustration of a post being annotated by three annotators. (a) The content of an exampled lonely post. (b) Distributional values of the post for all categories. For duration, two raters annotated "transient" and one annotated "ambiguous". For context, two raters annotated "social" and one annotated "romantic". For interpersonal relationship, three raters annotated "Friendship". For interaction styles, two raters annotated "seek advice" and one annotated "seek validation and affirmation". NA indicates not applicable.

for each sample $\mathbf{x}^{(i)}$, its predicted set of labels is denoted to be $\widehat{\mathcal{P}}^{(i)}$. Using these notations, the objective of our specific HDL task is to minimize $\mathcal{L}(\mathcal{P}, \widehat{\mathcal{P}})$ defined as

$$\frac{1}{N} \sum_{i=1}^{N} \left(\ell(\mathcal{P}_{lonely}^{(i)}, \widehat{\mathcal{P}}_{lonely}^{(i)}) + \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \ell(\mathcal{P}_c^{(i)}, \widehat{\mathcal{P}}_c^{(i)}) \right), \quad (1)$$

where N is the number of samples and ℓ is a loss function that measures the distance between two distributions (we used cross-entropy loss in our experiment).

4.3 Hierarchical Distributional Learning Methods

We consider two BERT-based methods for tackling HDL: a BERT + MLP method that adds an MLP classifier on top of a pre-trained BERT model (Devlin et al. 2019) for each distributional label; a Hierarchical Distributional Learning Network (HDLN) which is an adaption from HMCN-F network (Wehrmann, Cerri, and Barros 2018). For our baseline model, we use bidirectional LSTM (Hochreiter and Schmidhuber 1997) with an MLP classifier on top of it to learn each distributional label.

Both BERT + MLP and HDLN use a pre-trained BERT model fine-tuned on our dataset to obtain the embeddings of Reddit posts. The direct method individually minimizes cross-entropy loss for each classifier. On the other hand, HDLN incorporates label hierarchy into its architecture and concurrently outputs multiple distributional predictions. As illustrated in Figure 5, HDLN contains five local classifiers and one global classifier. Each local classifier individually predicts one of the five distributions (i.e., the loneliness binary distribution and four fine-grained distributions) and we denote the set of predicted distributional labels from the local classifiers to be $\widehat{\mathcal{P}}_L$. The global classifier predicts the concatenation of all five distributional labels, denoted as $\widehat{\mathcal{P}}_G$. Using both global and local information, HDLN is then trained by minimizing a joint loss function $\frac{1}{2}\mathcal{L}(\mathcal{P}, \widehat{\mathcal{P}}_L) + \frac{1}{2}\mathcal{L}(\mathcal{P}, \widehat{\mathcal{P}}_G)$ where $\mathcal{L}$ is defined in Eq. (1).

Following Wehrmann, Cerri, and Barros (2018), once HDLN is trained, we use a linear combination of the outputs from the global and local classifiers as the final predicted distributions $\widehat{\mathcal{P}}_F$, i.e.,

$$\widehat{\mathcal{P}}_F = \beta \widehat{\mathcal{P}}_L + (1 - \beta) \widehat{\mathcal{P}}_G, \quad (2)$$

where $\beta \in [0, 1]$ controls the amount of local and global information used in the final prediction.

Training To train the models, we split FIG-Loneliness into 70% for training, 20% for validation (e.g., hyperparameter tuning) and 10% for testing. We used AdamW (Loshchilov and Hutter 2019) as our optimizer with a warmup ratio of 0.1 and learning rate annealing from 2×10^{-5}. More details of the training procedure are in Appendix B.

Results Tables 2 and 3 show the results on the test set averaged over 3 random seeds. For binary loneliness classification, HDLN outperformed BERT + MLP and the LSTM baseline in all metrics. To rule out the possibility that the models were only learning the characteristics of different subreddits, we applied HDLN to the test dataset in the college subreddit only, and yielded an F1-score of 93.3% and an accuracy of 99.8%. For fine-grained classification, BERT + MLP performed better overall. However, it is worth noting that a separate BERT + MLP model is trained for each fine-grained category. Thus, a potential reason for BERT + MLP to outperform HDLN is that each BERT + MLP model has more optimized training parameters specifically for its corresponding fine-grained category. Despite the lower fine-grained classification accuracy, only a single HDLN model needs to be trained for all categories, thus significantly shortening both training and inference time (Appendix B). In addition, we found that learning the duration category was challenging for all models, consistent with the lower interrater similarity observed among human annotators (Figure 3a). In all subsequent sections, we will use HDLN for binary classification and BERT + MLP for fine-grained loneliness predictions.

5 Loneliness Characterizations

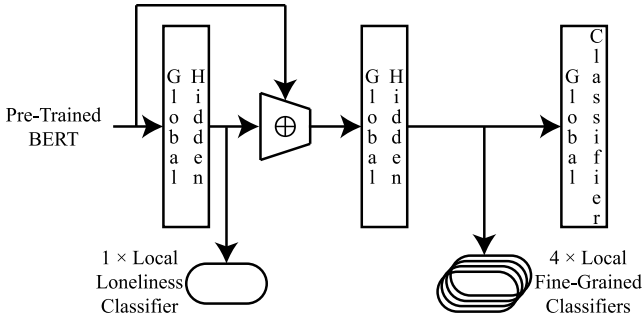Using FIG-Loneliness and the predictive models for fine-grained loneliness classification, we performed quantitative

Figure 5: Architecture of the proposed HDLN (more details in Appendix A). $\oplus$ denotes the vector concatenation operation. All local classifiers are two-layer MLPs. "Global Hidden" represents hidden layers in the global information flow.

	Acc. ↑	Precis. ↑	Rec. ↑	F1 ↑
LSTM Baseline	0.9024 ± 0.0014	0.8864 ± 0.0030	0.9014 ± 0.0018	0.8938 ± 0.0014
BERT + MLP	0.9722 ± 0.0046	0.9538 ± 0.0117	0.9870 ± 0.0036	0.9700 ± 0.0048
HDLN	**0.9763 ± 0.0041**	**0.9609 ± 0.0045**	**0.9883 ± 0.0063**	**0.9744 ± 0.0045**

Table 2: Results for loneliness binary classification. "Acc.": Accuracy. "Precis.": Precision. "Rec.": Recall. ↑ indicates higher values are better.

	Accuracy ↑	Clark ↓
Duration		
LSTM Baseline	0.4059 ± 0.0018	**1.4491 ± 0.0009**
BERT + MLP	**0.5992 ± 0.0114**	1.4682 ± 0.0060
HDLN	0.4539 ± 0.0073	1.4622 ± 0.0057
Context		
LSTM Baseline	0.7198 ± 0.0000	1.9656 ± 0.0005
BERT + MLP	**0.8573 ± 0.0102**	**1.9507 ± 0.0028**
HDLN	0.8560 ± 0.0220	1.9590 ± 0.0018
Interpersonal		
LSTM Baseline	0.5758 ± 0.0000	1.9369 ± 0.0010
BERT + MLP	**0.7976 ± 0.0031**	**1.9077 ± 0.0009**
HDLN	0.7795 ± 0.0191	1.9123 ± 0.0015
Interaction		
LSTM Baseline	0.6459 ± 0.0000	2.0005 ± 0.0005
BERT + MLP	**0.8352 ± 0.0120**	**1.9643 ± 0.0029**
HDLN	0.7237 ± 0.0138	1.9817 ± 0.0014

Table 3: Results for distributional fine-grained loneliness classification. ↑ (↓) indicates higher (lower) values are better. Accuracy and clark are defined formally in Appendix C.

analysis of data from all subreddits (r/youngadults, r/college, r/lonely and r/loneliness) to uncover (1) whether loneliness expressions differ between young adult-oriented communities and communities with diverse age groups; (2) the relationship between users' interaction strategies and forms of loneliness manifested in self-disclosure of loneliness; (3) the impact of COVID-19 on loneliness discourse across the forums.

5.1 Loneliness Expressions

How do young adults focused communities and communities with diverse age groups express loneliness? We explored the differences on loneliness discussions between Reddit communities that consists majority of young adults (r/youngadults, r/college; aged 18-25) and communities of a more diverse age group (r/lonely, r/loneliness). We leveraged the trained HDLN model to visualize the hidden representations of posts from these forums via t-SNE embedding (van der Maaten and Hinton 2008). We observed that among all the model predicted lonely posts, the young adult-focused group and the general group have well separated distribution centroids (Figure 6). This suggests that loneliness expressions are more similar within groups than between groups.

To further investigate the sources of such differences, we analyzed the composition of loneliness categories among posts from different subreddits. Table 4 presents the fraction of the posts annotated with each label in a loneliness category for all subreddits. These fractions were ob-
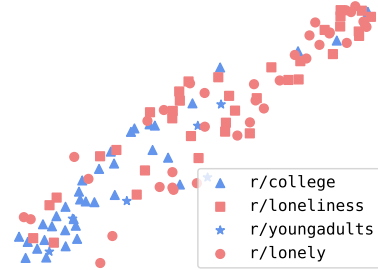

Figure 6: t-SNE embedding of the HDLN model predicted lonely posts across different subreddits.

tained by directly averaging the distributional labels of the labeled posts, and were normalized by removing the NA labels. Young adult-focused communities (r/college, r/youngadults) are different from communities of diverse age groups (r/lonely, r/lonely) in all loneliness categories. First, for duration of loneliness, posts from r/college and r/youngadults mentioned more transient loneliness and less enduring loneliness. One explanation is that young adults are more likely to be affected by situational factors, which could become the sources of loneliness (Bu, Steptoe, and Fancourt 2020). Another explanation is that loneliness related subreddits may attract users with more severe loneliness. Second, across four loneliness

contexts, young adult-oriented communities had more concerns related to somatic and physical loneliness. Interestingly, compared to other forums, college-oriented forum had the highest proportion of loneliness in the social domain and the least proportion of romantic loneliness. This emphasizes the importance of social relationships, especially for the college-age population. Third, regarding the specific relationships mentioned in the posts, we observed that individuals from `r/youngadults` and `r/college` experienced more family and peer related loneliness. Finally, we found that those in young adult-focused communities adopted more active coping strategies including seeking advice, validation and reaching out than passive coping strategies like non-directed interaction. Moreover, they also received greater amount of support from the communities, indicated by the greater proportion of posts labeled as providing support. This may suggest that although young adults are vulnerable to loneliness, they are building resilience in their own communities to combat loneliness.

We also examined loneliness-associated language markers using the Linguistic Inquiry and Word Count (LIWC) dictionaries (Pennebaker et al. 2015). We individually tested each dictionary output as a predictor for a logistic regression model controlling for year and subreddit, and corrected the standardized regression coefficients using Benjamini–Hochberg adjustment. Across all communities, posts with lonely labels are more present focused (example: *today*, *is*, *now*; β = .11; $P < .001$) and contain more words related to negative emotions (example: *hurt*, *ugly*, *nasty*; β = .21; $P < .001$), sadness (example: *crying*, *grief*, *sad*; β = .26; $P < .001$) and feeling (example: *feel*, *touch*; β = .15; $P < .001$). We also observed that lonely posts are associated with the use of more first-person pronouns (β = .18; $P < .001$) and references to cognitive processes (example: *cause*, *know*, *ought*; β = .13; $P < .001$). Analyses on a group level revealed that the young adult oriented group (`r/youngadults`, `r/college`) used more language related to peers such as buddy, coworker and neighbour (β = .18; $P < .001$) and expressed less sadness (β = .23; $P < .001$) compared to the other group (peer: β = .07; $P < .001$; sadness: β = .29; $P < .001$) in loneliness discourse. This is consistent with the higher proportion observed in *peers* and *transient* related loneliness in the young adult-focused forums.

5.2 Coping Strategies

How are different forms of loneliness associated with the types of coping strategies? To answer this question, we examined the conditional dependence between authors' interaction intents and the forms of loneliness in the domains of duration, context and interpersonal relationships in FIG-Loneliness. For example, to study the relationship between interaction intents and duration of loneliness, for posts with the same duration label, we computed the percentage of them annotated with different interaction labels that correspond to active or passive coping strategies. Active coping strategies are used to define and resolve the source of stress, and thus include "reaching out", "seeking advice" and "seeking validation" labels. Passive coping strategies

	r/youngadults	r/college	r/lonely	r/loneliness
Duration				
transient	31.7	29.8	29.4	28.8
enduring	29.5	34.8	34.5	35.2
ambiguous	38.8	35.4	36.2	36.0
Context				
social	69.8	77.9	71.6	73.1
physical	3.7	4.9	2.3	3.0
somatic	4.8	5.2	2.8	3.6
romantic	21.7	12.0	23.3	20.3
Interpersonal				
romantic	24.3	13.5	26.1	23.1
friendship	60.2	63.5	63.3	63.5
family	10.1	15.1	7.2	9.3
peers	5.4	8.0	3.4	4.2
Interaction				
seek advice	14.4	18.1	10.0	11.1
provide support	9.2	9.7	5.0	5.9
seek val. & aff.	9.4	12.2	6.6	7.5
reach out	23.8	18.2	16.9	15.5
non directed	43.2	41.8	61.5	60.0

Table 4: Fine-grained category composition. Percentages of all label values within a category are shown.

involve attempts to manage emotional stress without confronting the source of stress, which corresponds to our "non-directed interaction" label.

Duration of Loneliness Overall, more than 67% posts did not explicitly seek interactions with other users. Posts labeled as transient loneliness have a greater proportion (5.3% greater) than enduring loneliness in the use of active strategies. The most used active strategy is "reaching out" for transient loneliness (14.4%), and "seeking advice" for enduring loneliness (10.7%). A closer examination of the difference in coping strategies between the two types of loneliness revealed that those individuals with transient loneliness make relatively more efforts in reaching out or connecting with others (6.7% more), and fewer attempts at seeking validation or affirmation from people (1.6% less). Moreover, posts with enduring loneliness exhibited the greatest use of passive coping strategies.

Loneliness Context Across the four contexts, more than 66% of the posts used passive coping strategies (i.e., non-directed interaction). Seeking advice is the most popular strategy in physical (12.8%) and romantic loneliness (11.6%). In addition, the most used strategies for somatic and social loneliness are seeking validation (10.6%) and reaching out (14.4%).

Social Relationships for Loneliness Regarding interpersonal relationships involved in the loneliness discourse, consistent with previous findings, majority of the posts used passive coping strategies (over 65%). In the posts seeking for interactions, seeking advice is the most commonly used coping strategy for romantic (11.8%), family (11.0%) and peers (18.6%) related loneliness, whereas reaching out is the most used strategy to cope with loneliness related to friendships

(15.3%).

5.3 Impacts of the COVID-19 Pandemic

During the COVID-19 pandemic, understanding changes in mental health needs over time, especially related to loneliness, has become an important and urgent topic. The uncertainty in the pandemic, caused by the health threat from the virus and financial stress, has negatively affected many individuals' mental and physical health. Moreover, due to the social distancing policy, the prolonged state of physical isolation from peers, teachers and community networks may lead to a significant increase in the feelings of loneliness, especially among young adults (Bu, Steptoe, and Fancourt 2020). Thus, a better understanding of the trends of different loneliness categories before and during the pandemic can facilitate early intervention and design for reducing loneliness in public. In this section, we investigated the potential impacts of COVID-19 on people's loneliness experiences on Reddit.

To do this, we first examined the temporal trend of loneliness-related discussions on Reddit. We observed that both the total number of posts and loneliness-related discussions in FIG-Loneliness (i.e., the annotated posts) have been growing in `r/college`, `r/youngadults`, `r/lonely` and `r/loneliness` from January 2018 to December 2019 (Figure 7a). A similar pattern was observed in the HDLN model predicted labels on posts that were not annotated (Figure 7b). Moreover, the consistent high volume of posts in 2020 suggests that many individuals kept sharing and processing (loneliness) experiences during the pandemic on Reddit.

Next, we looked at the composition for the forms of loneliness across years in FIG-Loneliness. We found some noticeable changes in the types of loneliness people discussed in the pandemic (year 2020) compared to before (year 2018, 2019). Specifically, during the pandemic, there were increases in loneliness expressions on transient loneliness (by 8%), and concerns around romantic (context: by 5%; interpersonal: by 6%) and family relationship (by 3%). This indicates that romantic and family loneliness rose in 2020, perhaps as a result of spending quarantine away from romantic partners and family during the pandemic.

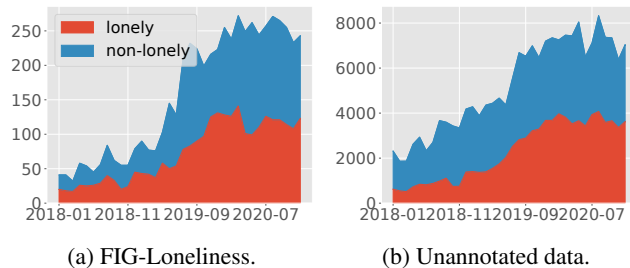(a) FIG-Loneliness.   (b) Unannotated data.

Figure 7: Number of lonely and non-lonely posts over time. (a) Amount of lonely vs. non-lonely posts in FIG-Loneliness. (b) Amount of lonely vs. non-lonely posts for unannotated data using model predictions from HDLN.

(a) Physical context.   (b) Non-directed interaction.
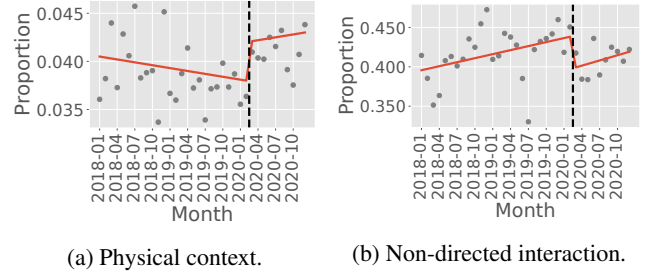
Figure 8: Trends of fine-grained loneliness category labels for (a) physical context and (b) non-directed interaction in young adult focused forums. Dashed lines represent the intervention (i.e., the national lockdown).

To investigate *the immediate effects* of the outbreak of the pandemic (the lockdown orders) on different forms of loneliness, using FIG-Loneliness and HDLN model predictions, we performed interrupted time-series analysis (Ferron and Rendina-Gobioff 2005) on the proportions of posts belonging to different category labels across months. We took March, 2020 [4] to be the time when the intervention has taken place. We used the following linear model to fit the data:

$$Y = b_0 + b_1 T + b_2 D + b_3 M + e,$$

where Y represents the proportions of different types of posts over time, T represents months since January 2018, D is a binary variable indicating whether the intervention is applied (i.e., before or after March, 2020), and M is the number of months passed since the intervention. Coefficients b_0, b_1, b_2, and b_3 correspond to the initial proportion, the pre-intervention trend, immediate effect of the intervention, and the difference between pre- and post-trends, respectively. e represents an error term.

Among the four loneliness categories across subreddits, context and interaction yielded statistically significant immediate effects, only in the young adult forums (`r/youngadults` and `r/college`). We speculate that the COVID-19 school lockdown may lead to increased loneliness in young adults such as college students, as a result of changes in physical environments. Specifically, the proportion of *physical* loneliness had a sudden increase from 3.8% to 4.2% in March, 2020 (Figure 8a; 95% CI $[-.0004, .0084]$, $P < .01$). This suggests that young adults were potentially immediately affected by the lockdown orders and changes in physical environments. In addition, we observed a sudden decrease in the proportion of loneliness expressions related to *non-directed interaction* or passive coping strategies, from 44.0% to 39.9% (Figure 8b; 95% CI $[-.0855, .0033]$, $P < .01$). This could result from an active coping approach adopted by the young adult population, given that they are overall more active in their interaction styles pertaining to loneliness discussions (see Table 4).

---

[4]National lockdown and stay-at-home orders were issued in the U.S. in March 2020.

6 Discussion and Conclusion

In this study, we provide a large human coded dataset FIG-Loneliness that includes annotations for fine-grained categories of loneliness. Using methods for hierarchical distributional learning, we built multi-label classifiers to examine how different forms of loneliness and users' interaction styles (including coping strategies) manifest in online discourse from two loneliness specific forums (r/lonely, r/loneliness) and two young adults focused forums (r/college, r/youngadults). With FIG-Loneliness and model predictions, we showed the differences in loneliness discourse across these Reddit forums, the relationships between loneliness forums and authors' coping strategies, and the impacts of the pandemic on different loneliness types.

We found that from 2018 to 2020, Reddit has become an increasingly popular platform for individuals to disclose sensitive information such as personal loneliness experiences. Inferred from loneliness discourse in r/college and r/youngadults, young adults, the primary users of social media, are more likely to use the Reddit platform to seek active interactions with others to cope with loneliness. They also receive some social support which could build resilience in the communities. Compared to other groups, young adults are more likely to experience loneliness in the physical and somatic domains, which highlights the potential harmful effects of loneliness on physical health due to geographical isolation during the pandemic. Further, family and peer related loneliness appear more prevalent in the young adult group, which can be explored in future research.

Consistent with prior literature, loneliness is associated with the use of first-person pronouns, references to words indicating thinking and reasoning (LIWC cognitive processes dictionary), and language related to negative emotions and sadness in the LIWC dictionaries (Guntuku et al. 2019). This suggests a preoccupation of self and a potential association between rumination and loneliness, which is also associated with depression. We found that those in young adult-focused forums are more likely to mention words related to *peers* in the loneliness discourse, suggesting that peer support is especially critical in promoting young adults' psychological well-being.

In addition, individuals with different forms of loneliness have different interaction styles. For example, in posts that request explicit social interactions, individuals with transient loneliness are more likely to seek social connections whereas those with enduring loneliness are more likely to seek advice or validation from the communities. Physical and romantic loneliness are also more associated with requests for advice, and somatic loneliness is more associated with validation seeking. These interaction styles are associated with active coping strategies because they are problem-focused. However, the majority of the posts are using an emotion-focused or passive coping approach (no explicit intent for social interactions). Future research efforts should work on methods to help individuals to adopt active coping strategies to alleviate the feeling of loneliness. Those with enduring loneliness can especially benefit from such work as they bear more negative consequences from loneliness, and

have a relatively greater use of passive coping strategies.

Moreover, COVID-19's social distancing and school lockdown orders may have a greater immediate effect on individuals from r/college and r/youngadults, as evidenced by a sudden increase in loneliness related to physical context and requests for active interactions within the communities. Linking our observation that those from young adult-focused forums experience relatively more physical loneliness, we speculate that young adults may be more sensitive to the influence of geographical isolation (e.g., school lockdown or community changes). Future studies should examine the risk factors involved in changes of physical environments for young adults' loneliness.

It is worth emphasizing that loneliness is a multidimensional concept and can be measured in different ways. The distinction between various forms of loneliness has been emphasized in psychological literature as it bears on understanding unique causes, manifestations, and consequences of each form. Our study confirms the possibility of automatic detection of different forms of loneliness and coping strategies, and provides a model framework for loneliness screening. The prediction of such information can help identify who may possibly need additional resources, and employ effective early intervention programs for loneliness. In future work, it is important to examine the causal relationship between the specific loneliness forms and their long-term adverse effects. For example, previous literature suggests that individuals who suffer from chronic loneliness exhibit long-term difficulties in interpersonal relationships, relative to transient loneliness which relates to an adaptation to situational changes (De Jong-Gierveld and Raadschelders 1982). This is because certain forms of loneliness, such as transient loneliness, are typically easier to cope with and recover from. It is possible that those who are with certain labels in their loneliness expressions such as "chronic" and "somatic" loneliness, are more likely to develop negative psychological outcomes in the long run compared to those with labels including "transient" and "physical" loneliness. Similarly, future work may follow-up on the effects of different coping strategies (e.g., seeking advice or validation and reaching out) on loneliness reduction because previous work suggests that coping strategies can make a difference in the psychological outcomes (Deckx et al. 2018).

Limitation and Data Disclaimer Several limitations to our work should be noted. We drew loneliness discourse from r/college and r/youngadults as a proxy for studying young adults' loneliness expressions. However, these data may not be representative of young adults' expressions. For example, 16.7% of r/college and r/youngadults posts were written by users below age 18. Another limitation of our data is that our non-lonely samples are only from r/college and r/youngadults. To capture a broader range of expressions not related to loneliness, one may sample non-lonely posts from a more diverse category of subreddits. We also acknowledge that our collected data is from a particular social media platform (Reddit) and written in English only. Thus, it may not represent loneliness discourse outside this scope, and may contain

cultural, platform-specific biases among many others. Since our models are trained on our choice of data, the pattern it learned may only be reflective to this particular source of data. The models' ability to generalize to other data sources needs to be further studied.

A Model Architectures

Architecture hyperparameters we used are given in Table 5. BERT + MLP and HDLN models take input from pre-trained BERT models implemented by Wolf et al. (2020). We used the `bert-base-cased` variant with default configurations. The output size of each classifiers equals to the number of supports of corresponding distribution.

Hyperparameter	Value
LSTM Baseline	
Input Size	768
Hidden Size	128
# LSTM Layers	1
# Classifier Layers	2
BERT + MLP	
Input Size	768
Hidden Size	50
# Layers	2
HDLN	
Input Size	768
Global Hidden Size	64
Local Hidden Size	64
# Layers	2

Table 5: Architecture hyperparameters for LSTM Baseline, BERT + MLP, and HDLN models.

B Training

During training we adopted a value of 16 for batch size and did not use any weight decay. For the BERT + MLP models, we found that they converged within 10 epochs. For the HDLN model, it took more epoches (20 epochs) to converge than the BERT + MLP models did. For all models, we terminated the training when the validation set accuracy no longer improved. However, since an individual BERT + MLP model was required for each loneliness category, the total training time of the BERT + MLP models were 2.5 times longer than that of the HDLN model (each individual BERT + MLP model took half the time of HDLN's, but we had to train five BERT + MLP models).

C Evaluation Metrics

Denoting a K-dimensional target real-value vector as $\boldsymbol{D} = [D_1, D_2, \ldots, D_K]$ and a K-dimensional predicted real-value vector as $\hat{\boldsymbol{D}} = [\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_K]$, both satisfy distribution constrains. Accuracy and clark are calculated as in Table 6.

Acknowledgments

We thank Terresa Eun and Weier Wan for their helpful discussions and comments. This work was supported by an un-

Metric	Formulation
Accuracy	$\mathbb{1}\left(arg\,max(\hat{\boldsymbol{D}}) \in arg\,max(\boldsymbol{D})\right)$
Clark	$\sqrt{\sum_k^K \frac{\left(D_k - \hat{D}_k\right)^2}{\left(D_k + \hat{D}_k\right)^2}}$

Table 6: Metrics calculation.

References

Akbari, A.; Awais, M.; Bashar, M.; and Kittler, J. 2021. How Does Loss Function Affect Generalization Performance of Deep Learning? Application to Human Age Estimation. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, 141–151. PMLR.

Andalibi, N.; Haimson, O. L.; De Choudhury, M.; and Forte, A. 2016. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, 3906–3918.

Baumgartner, J.; Zannettou, S.; Keegan, B.; Squire, M.; and Blackburn, J. 2020. The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media*, volume 14, 830–839.

Bu, F.; Steptoe, A.; and Fancourt, D. 2020. Who is lonely in lockdown? Cross-cohort analyses of predictors of loneliness before and during the COVID-19 pandemic. *Public Health*, 186: 31–34.

Cacioppo, J. T.; Cacioppo, S.; and Boomsma, D. I. 2014. Evolutionary mechanisms for loneliness. *Cognition & emotion*, 28(1): 3–21.

Cacioppo, J. T.; and Patrick, W. 2008. *Loneliness: Human nature and the need for social connection*. WW Norton & Company.

Center, P. R. 2021. Social Media Fact Sheet. https://www.pewresearch.org/internet/fact-sheet/social-media/. Accessed: 2022-04-10.

Chancellor, S.; and De Choudhury, M. 2020. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine*, 3(1): 1–11.

De Choudhury, M.; and De, S. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Eighth international AAAI conference on weblogs and social media*.

De Jong-Gierveld, J.; and Raadschelders, J. 1982. Types of loneliness. *Loneliness: A sourcebook of current theory, research and therapy*, 105–119.

Deckx, L.; van den Akker, M.; Buntinx, F.; and van Driel, M. 2018. A systematic literature review on the association between loneliness and coping strategies. *Psychology, health & medicine*, 23(8): 899–916.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. Minneapolis, Minnesota: Association for Computational Linguistics.

Ellis, D.; and Cromby, J. 2012. Emotional inhibition: A discourse analysis of disclosure. *Psychology & health*, 27(5): 515–532.

Ferron, J.; and Rendina-Gobioff, G. 2005. *Interrupted Time Series Design*. American Cancer Society. ISBN 9780470013199.

Geng, X. 2016. Label Distribution Learning. *IEEE Transactions on Knowledge and Data Engineering*, 28(7): 1734–1748.

Guntuku, S. C.; Schneider, R.; Pelullo, A.; Young, J.; Wong, V.; Ungar, L.; Polsky, D.; Volpp, K. G.; and Merchant, R. 2019. Studying expressions of loneliness in individuals using twitter: an observational study. *BMJ open*, 9(11): e030355.

Hawkley, L. C.; and Cacioppo, J. T. 2010. Loneliness matters: A theoretical and empirical review of consequences and mechanisms. *Annals of behavioral medicine*, 40(2): 218–227.

Hochreiter, S.; and Schmidhuber, J. 1997. Long Short-Term Memory. *Neural Computation*, 9(8): 1735–1780.

Holt-Lunstad, J.; Smith, T. B.; Baker, M.; Harris, T.; and Stephenson, D. 2015. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. *Perspectives on psychological science*, 10(2): 227–237.

Kivran-Swaine, F.; Ting, J.; Brubaker, J.; Teodoro, R.; and Naaman, M. 2014. Understanding loneliness in social awareness streams: Expressions and responses. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8.

Koh, J. X.; and Liew, T. M. 2020. How loneliness is talked about in social media during COVID-19 pandemic: text mining of 4,492 Twitter feeds. *Journal of psychiatric research*.

Lakoff, G. 2008. *Women, fire, and dangerous things*. University of Chicago press.

Lasgaard, M.; Goossens, L.; Bramsen, R. H.; Trillingsgaard, T.; and Elklit, A. 2011. Different sources of loneliness are associated with different forms of psychopathology in adolescence. *Journal of Research in Personality*, 45(2): 233–237.

Loshchilov, I.; and Hutter, F. 2019. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*.

Low, D. M.; Rumker, L.; Talkar, T.; Torous, J.; Cecchi, G.; and Ghosh, S. S. 2020. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study. *Journal of medical Internet research*, 22(10): e22635.

Luchetti, M.; Lee, J. H.; Aschwanden, D.; Sesker, A.; Strickhouser, J. E.; Terracciano, A.; and Sutin, A. R. 2020. The trajectory of loneliness in response to COVID-19. *American Psychologist*.

Mahoney, J.; Le Moignan, E.; Long, K.; Wilson, M.; Barnett, J.; Vines, J.; and Lawson, S. 2019. Feeling alone among 317 million others: Disclosures of loneliness on Twitter. *Computers in Human Behavior*, 98: 20–30.

Mushtaq, R.; Shoib, S.; Shah, T.; and Mushtaq, S. 2014. Relationship between loneliness, psychiatric disorders and physical health? A review on the psychological aspects of loneliness. *Journal of clinical and diagnostic research: JCDR*, 8(9): WE01.

Olteanu, A.; Varol, O.; and Kiciman, E. 2017. Distilling the outcomes of personal experiences: A propensity-scored analysis of social media. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 370–386.

Pennebaker, J. W.; Boyd, R. L.; Jordan, K.; and Blackburn, K. 2015. The development and psychometric properties of LIWC2015. Technical report.

Perlman, D.; Peplau, L. A.; and Goldston, S. 1984. Loneliness research: A survey of empirical findings. *Preventing the harmful consequences of severe and persistent loneliness*, 13–46.

Peterson, J. C.; Battleday, R.; Griffiths, T.; and Russakovsky, O. 2019. Human Uncertainty Makes Classification More Robust. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 9616–9625.

Rokach, A.; and Brock, H. 1997. Loneliness and the effects of life changes. *The Journal of psychology*, 131(3): 284–298.

Saha, K.; Ernala, S. K.; Dutta, S.; Sharma, E.; and De Choudhury, M. 2020. Understanding moderation in online mental health communities. In *International Conference on Human-Computer Interaction*, 87–107. Springer.

Schmidt, N.; and Sermat, V. 1983. Measuring loneliness in different relationships. *Journal of personality and social psychology*, 44(5): 1038.

Turner, R. J.; Frankel, B. G.; and Levin, D. M. 1983. Social support: Conceptualization, measurement, and implications for mental health. *Research in community & mental health*.

van der Maaten, L.; and Hinton, G. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(86): 2579–2605.

Wehrmann, J.; Cerri, R.; and Barros, R. 2018. Hierarchical Multi-Label Classification Networks. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, 5075–5084. PMLR.

Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; Davison, J.; Shleifer, S.; von Platen, P.; Ma, C.; Jernite, Y.; Plu, J.; Xu, C.; Scao, T. L.; Gugger, S.; Drame, M.; Lhoest, Q.; and Rush, A. M. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. Online: Association for Computational Linguistics.

Yang, D.; Yao, Z.; Seering, J.; and Kraut, R. 2019. The channel matters: Self-disclosure, reciprocity and social support in online cancer support groups. In *Proceedings of the 2019 chi conference on human factors in computing systems*, 1–15.