Empowering Predictions of the Social Determinants of Mental Health through Large Language Model Augmentation in Students' Lived Experiential Essays

Mohammad Arif Ul Alam[1,2,3], Madhavi Pagare[1], Susan Davis[4], Geeta Verma[4], Ashis Biswas[4], Justin Barbern[5]
[1] *University of Massachusetts Lowell*
[2] *University of Massachusetts Chan Medical School*
[3] *National Institute on Aging, National Institute of Health*
[4] *University of Colorado Denver*
[5] *Pearson*

ABSTRACT
Recognizing the Social Determinants of Mental Health (SDMHs) among students is essential, as lower backgrounds in these determinants elevate the risk of poor academic achievement, behavioral issues, and physical health problems, thereby affecting both physical and emotional well-being. Leveraging students' self-reported lived experiential essays yields substantial insights into the SDMHs. However, constructing an automated prediction tool for SDMHs necessitates thorough planning, efficient design of a web-based tool for data collection, articulation of SDMHs within the context of lived experiences, expert data annotation, and the implementation of an efficient multi-label classifier for prediction. This paper investigates the capabilities of Large Language Models (LLMs) in the development of a multi-label SDMHs prediction system in students' lived experiential essays covering the above aspects. In this regard, we propose a novel Human-LLM Interaction for Annotation (HLIA) method to label texts pertaining to predetermined SDMHs and develop a Multitask Cascaded Neural Network (MTCNN) classification algorithm to predict these determinants in students' experiential essays. Additionally, we developed a web-tool based lived experience essay data collection system, developed a dataset of 1̃500 lived experiential essays collected from 800+ students annotated by 4 educational experts (IRB approved) and evaluated the proposed framework.

Keywords
educational data mining, lived experience, multitask learning, multilabel classification, social determinants of mental health, large language model, human large language model interaction

1. INTRODUCTION
With the increasing focus within the educational sector on extracting varied student experiences for purposes beyond micro-credentials, our team of educational professionals identified a critical aspect of mental health known as Social Determinants of Mental Health (SDMHs). This area represents a substantial part of the narratives shared by students on our platform [10, 12]. Social determinants of mental health (SDMHs), encompassing factors like socioeconomic status, access to nutritious food, education, housing, and physical environment, exhibit a significant correlation with students' mental well-being (stress, depression, suicide) and academic performance (dropout, failure). For instance, social disruptions (e.g., relationship breakdowns, financial instability, legal issues, or exposure to childhood adversity) are widely recognized as triggers for students' mental distress, potentially leading to depression and withdrawal from academic pursuits [28]. To develop effective policies addressing students' mental health, it is crucial to move beyond predictor identification and ascertain the strength of the relationship between SDMHs and mental well-being [32]. However, a major challenge has been the limited availability of comprehensive and reliable SDMHs data in large-scale population databases, with researchers traditionally relying on structured data such as survey instruments. Structured data often lack comprehensive information on SDMHs, particularly when primarily designed for scoring purposes.

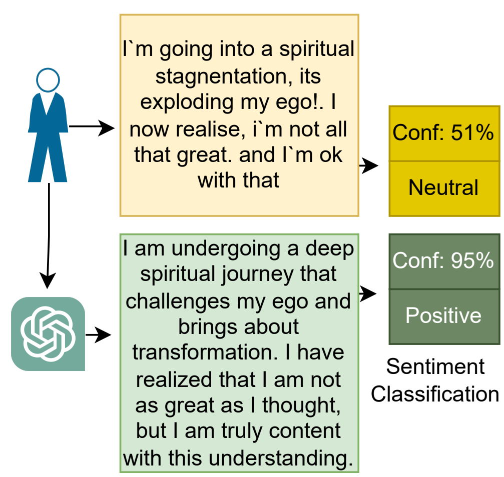

Figure 1: Example BERT-based sentiment analysis on a lived experience essay and its equivalent ChatGPT texts while attempting to correct its errors

Natural Language Processing (NLP) approaches have been employed to predict users' mental health status based on self-explanatory

M. A. U. Alam, M. Pagare, S. Davis, G. Verma, A. Biswas, and J. Barbern. Empowering predictions of the social determinants of mental health through large language model augmentation in students' lived experiential essays. In B. Paaßen and C. D. Epp, editors, *Proceedings of the 17th International Conference on Educational Data Mining*, pages 617–627, Atlanta, Georgia, USA, July 2024. International Educational Data Mining Society.

https://doi.org/10.5281/zenodo.12729900

text, typically derived from social media posts or self-reported experiential essays [32]. However, these methods heavily rely on a large volume of data that needs to be annotated by mental health diagnosis experts who possess the expertise to identify social determinants from textual content. This annotation process is costly, time-consuming, cumbersome, and prone to potential errors. To address these limitations, this paper introduces a novel approach called LLM-augmented expert annotation, which focuses on annotating mental health social determinants in students' self-reported essays. This methodology represents the first of its kind in the field.

LLMs, the new pocket calculator of text generation for modern era, while offering numerous practical applications in NLP research, also pose certain risks [17, 30]. These risks include the potential amplification of biases present in training data [23], leading to biased or discriminatory language generation. There is a risk of malicious use, as LLMs can be exploited to generate misinformation or harmful content at scale [37]. Overreliance on LLMs without human review may diminish critical thinking, judgment and potentially demolitions of one's own textual authenticity. Mitigating these risks requires responsible development, fact-checking mechanisms, transparency, and adherence to ethical guidelines in LLM usage. For example, Fig. 1 shows BERT-based sentiment classification [31] framework's application on an original students' experiential essays and a ChatGPT [29] generated equivalent texts in attempting to mitigate errors, shows significant difference in sentiment classification (Neutral for original and Positive for ChatGPT one).

We propose an efficient Human-LLM Interaction for Annotation (HLIA) framework to generate appropriate strategy for LLM-Augmented expert annotation of SDMHs from students' lived experience texts. Additionally, we propose a Multitask Cascaded Neural Network on BERT embedding (MTCNN) to classify and evaluate the prediction of SDMHs (Fig. 2 presents the overview).

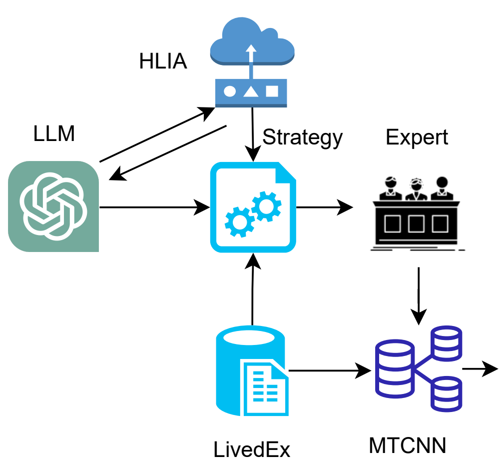

Figure 2: Overview of the proposed framework, HLIA: Human-LLM Interaction for Annotation, LivedEx: lived experience texts from students, MTCNN: Multitask Cascaded Neural Network

2. RELATED WORKS

Most of the prior ML-augmented text annotation works proposed efficient active learning based approaches to select best subset to annotate [34, 45, 33, 39] while others focused on crowdsourcing toolkit design either via collaborative [6, 46] or intervention [49]. Further researches have been done to design efficient interface for annotators [13, 22] via important information extraction [25, 4, 24] to help social or expert annotators [9].

LLM has demonstrated its potential in several practical applications within the field of NLP research. Researchers have proven that LLM can be used to generate synthetic training data or naturalistic text generation tasks [14], such as generating product reviews [11] or dialogue systems [16], and for abstractive summarization [43], aiding in the creation of concise summaries of longer texts to help improve the performance and robustness of NLP models. Beyond that LLM can also be employed in tasks such as semantic search [35] and information retrieval [38, 47], sentiment analysis [18], language translation [48], and error correction [48], offering a broad range of practical uses in the field of NLP. With its ability to understand and generate human-like text, LLM serves as a versatile tool for researchers, developers, workers and more importantly students to aid their everyday work involving texts and documentation [2].

Text annotation utilizing LLM by taking advantage of its explainability has been rarely explored before [15]. One of the major contributions in this new domain of research, Xingwei et. al. [15], proposed an 'explain-then-annotate' method to generate expert augmented annotation for text labeling. Here, experts request a comprehensive explanation by providing textual commands to language models (LLMs). Nevertheless, this approach lacks a mechanism for determining which queries are more effective in elucidating the output. This raises concerns about the reliability and sustainability of the annotation process. In this paper, we introduce a prompt design strategy that utilizes a trial-and-error technique in conjunction with statistical measures of prompt quality estimation to identify the most effective strategies. Furthermore, we formulated an expert-augmented definition, including keywords and clear definitions of target labels for Student Mental Health Datasets (SDMHs) to enhance the interpretibility of LLMs annotation for experts' confirmation. By employing our proposed method, we identified the optimal prompt strategy and enhanced the multi-class multi-label classification of SDMHs for students.

3. METHODOLOGIES

3.1 Lived Experience Background

Students' potential extends beyond academic achievements and professional recognition. Their lived experiences are rich sources of academic skills, offering credentials equivalent to formal education. Our goal is to foster a just and ethical world—one individual, one story at a time—by redefining the narrative around academic and professional journeys. Skills such as problem-solving, teamwork, and leadership are equally vital as formal qualifications (e.g., degrees, languages, certifications) for personal and professional development. Consequently, there is a growing emphasis on integrating and evaluating these transferable skills within education and recruitment. However, existing assessment methods are subjective, inefficient, and inconsistent, disproportionately affecting marginalized groups—such as minorities and women—who often possess valuable skills like resilience and conflict resolution but lack the means to formally demonstrate them.

To address this challenge, we introduced a platform, LivedX dedicated to capturing students' lived and learned experiences, transforming them into stackable, transferable skill credentials that can be converted into college credits through partnerships with public universities and our organization [5]. This initiative aims to empower students to pursue further education or career opportunities. On our web-based platform, students articulate their experiences via a straightforward questionnaire. Utilizing a patented machine learning system, which includes an innovative Natural Language Processing (NLP) algorithm and responsible AI design, the platform analyzes these narratives. It then employs evidence-based social-emotional assessment frameworks to issue micro-credentials that affirm the transferable skills inherent in each experience. Over 150 micro-credentials are organized into a three-tier hierarchy, with the accumulated data presented on a user- and institution-accessible dashboard. Our platform employs a prototype guiding students in narrative entry and accessing their micro-credential dashboards. This innovative approach acknowledges the value of individuals' real-life experiences—work, volunteering, projects, personal achievements—and translates them into recognized credentials. By capturing and evaluating diverse experiences, our technology enables the acknowledgment of essential life skills often overlooked by traditional educational frameworks [40].

3.2 Data Collection

To collect students' lived experience effectively, we have utilized our core lived experience data collection platform where students were guided and encouraged to systematically document their personal lived experiences using this tool. We recruited four experienced academic mental health researchers and trained them with our target Social Determinants of Mental Health (SDMHs) components to annotate our collected data. We engaged 860 students over 4 years to participate in the lived experience web-based tool to share their lived experiences voluntarily that constitutes 1580 texts in total. We eliminated the texts that contain < 50 characters given the existence of potential classification ambiguity due to the lack of enough texts, that results a total 1498 texts.

3.3 Scripting Social Determinants of Mental Health (SDMHs)

The role of Social Determinants of Mental Health (SDMHs) in the academic and personal lives of students is significant [3]. Research indicates a correlation between social determinants, such as food insecurity, and adverse effects on mental health, leading to lower academic performance among students [27]. Among undergraduate students, key social determinants influencing mental health encompass academics, family dynamics, employment, economic conditions, romantic relationships, and religious affiliations [3].

In a European study, approximately one-third of first-year university students were identified as having mental health-related issues [7]. Academic performance has been consistently linked to mental health, with poor academic outcomes contributing to worsened mental health and vice versa [20]. Additional factors contributing to mental health deterioration include economic challenges, high parental expectations, strained relationships, and an unhealthy lifestyle [44, 1, 36]. Common mental disorders are prevalent among individuals aged 14 to 24 years, a demographic that includes many university students [19]. These students often face new experiences, such as relocating from home and assuming adult financial responsibilities. Furthermore, shifts in health behaviors are frequently observed in this age group [42]. Adapting to these changes can be challenging, especially as students are simultaneously expected to fulfill coursework requirements and participate in exams. Failure to effectively manage these challenges may lead to mental health difficulties [42].

Our team of experts have been investigating lived experience texts for 3 years. We have found that lived experiences offer valuable insights into the development and manifestation of mental health challenges. Traumatic events, such as abuse or witnessing violence, increase the risk of conditions like PTSD or depression. Marginalized communities also face unique stressors and discrimination that contribute to mental health difficulties. These challenges further compound the existing stressors faced by minoritized students due to a history of racism, classism, and oppression, impacting their academic achievement and career readiness. The relationship between mental health challenges and academic achievement among minoritized students is multifaceted and influenced by various factors, including barriers to access and support, school climate and cultural competence, intersectionality, resilience, and community support. Identifying mental health challenges in this context is hindered by stigma, lack of awareness, cultural and contextual factors, the development and normalization of symptoms (especially in youth), and limited communication and expressive skills.

Following a thorough review of existing literature, and study on lived experience data, our team of experts conceptualized and formulated 13 distinct categories of social determinants of mental health, each with potential impacts on students' academic, physical, and mental well-being. Every determinant is meticulously defined, encompassing various criteria linked to keywords frequently used by students in their daily written or spoken communication. Furthermore, each determinant is complemented by a comprehensive, plain-description paragraph that illuminates its individual significance facilitating the identification process for annotators (refer to Appendix Table 3) for comprehensive information. These determinants can potentially be identified through the analysis of students' lived experience essays.

4. HUMAN-LLM INTERACTION FOR ANNOTATION

LLM is being widely used for various purposes, such as text classification, conversational intelligent interpretation, and code generation. However, the application of LLM for annotating texts as a substitute for expert annotation is relatively unexplored area that holds potential benefits and risks in the future of NLP [15]. In this paper, we propose a novel LLM-augmented text annotation approach (HLIA), leveraging the interpretability capabilities of LLM technology in the field of NLP. The main objective of this approach is to alleviate the extensive efforts required by expert annotators through the utilization of LLM's interpretability. We introduce a technique called Human-LLM Interaction for Annotation (HLIA) to facilitate this process. The data annotation process was conducted in two phases: in the first phase, trained experts were asked to annotate a small amount of data without Language Model (LLM) interaction. In the second phase, annotators received detailed instructions on how to utilize LLMs for designing LLM prompt strategies, combined with statistical quality assessment. Finally, with the best prompt strategy (highest partial correlation coefficient), annotators continued the annotation process until all the texts were annotated.

4.1 LLM Augmented Annotation via HLIA

In our HLIA (Algorithm 1), with the expert literature studies, we input our expert developed set of SDMHs (Appendix Table 3 (Appendix)), a small amount of data, D_l, and, large unlabeled data, D_u. We trained four mental health experts with a set of LLM query strategies. These series of strategies include, asking LLM interface to classify the text with multiple SDMHs, existence of any of the SDMHs and so on. The purpose of the strategies selection was to ask LLM interface not only the label of an input text, but, also obtain why LLM decided to classify this text with certain SDMS, and, why LLM did not label it with some certain SDMs. Once we obtain the labels annotated by the LLM interface, we apply that on D_l and calculate partial correlation coefficient, $\mathcal{R}$ on the existence of any form of SDMHs in the text considering the other 13 SDMHs as control variables [32]. Then, we consider if the $\mathcal{R} > Th$, where $Th = .8$ (standard practice [32]), we stop the query selection, otherwise continue the exploration to step 1. After obtaining desired $\mathcal{R}$, we run the strategy on entire unlabeled dataset, D_u, to generate a labeled D_u. This labeled D_u has been further investigated by our SDMHs experts to mitigate any unwanted biases in annotation. After running 15 iteration, we obtained an optimal strategy which is stated in details in Appendix Figure 6 that resulted first $\mathcal{R} > .8$ partial correlation coefficient on existence of any form of SDMHs considering other 13 SDMNs as control variables.

Algorithm 1: Human-LLM Interaction for Annotation (HLIA)

```
Input  : Social Determinants, Small Labeled Data (D_l), Large
         Unlabeled Data (D_u)
Output : labeled D_u
continue_loop ← True;
while continue_loop do
    1. Strategy ← Selected strategy;
    2. L ← Extracted annotation and interpretation;
    3. Calculate partial correlation coefficient, R, of SDMHs;
    4. if R > Th then
        continue_loop ← False;
    end
end
Label D_u via Strategy-augmentation;
return labeled D_u
```

4.2 Iterative HLIA to Annotate SDMHs

The solve purpose of designing Human-LLM Interaction for Annotation (HLIA) system is to reduce text efforts with the aid of LLM augmented text explanation. In this regard, at first, an annotator who is expert in extracting SDMHs from texts, annotated a set of 200 texts manually without any help from LLM. Then, 4 mental health expert annotators trained by our computer scientists and educators team in using the most popular LLM technology, ChatGPT [29], with appropriate instruction about the HLIA framework. The annotators followed the pre-defined list of strategies to obtain the best performing strategies (partial correlation coefficient $\mathcal{R} > .8$ as stated in Section 4). Note that, while assessing the performance of any strategy, experts also considered the explainability of LLM produced annotations. The best performing strategy has been shown in Appendix Fig. 6 stating the ChatGPT provided single text annotation as well as explanation. After finding the best strategy, the annotators are engaged in generating text annotations on each of the collected students' lived experience texts, checking the validity of the explanation provided LLM annotation, modifying the annotation and finally saving the LLM provided and LLM-augmented expert provided annotations.

5. MULTITASK CASCADED NEURAL NETWORK ON BERT EMBEDDING

5.1 Preliminaries and Notations

We denote a text written by a student about their recent experience as X. The set of social determinants of mental health (SDMHs) labels is denoted as Y, where $Y = 0, 1, \ldots, 12$ represents the 13 different labels.

5.2 Model Architecture

We propose a multitask cascaded learning framework (MTCNN) for SDMH detection. The core architecture of the model consists of an NLP embedding layer (e.g., Base BERT, RoBERTa) that maps the input text X to a continuous representation.

5.2.1 *Task 1: Binary Classification*

The first task is binary classification, aiming to detect the existence of one or more SDMHs in the text X. The output of the embedding layer is connected to a dense layer with 32 hidden units and a tangent activation function. Subsequently, it is connected to a final dense layer with a single output using softmax activation, representing the probability of the existence of SDMHs in the text.

5.2.2 *Task 2: Multilabel Classification*

After detecting the existence of SDMHs in the text (Task 1 output is true), the model develops a multilabel multitask learning model for the 13 different labels of SDMHs. Task 1 output is connected to 13 different dense layers, each representing one of the binary classes of the 13 SDMHs. Each dense layer starts with a dense layer with tangent activation and 32 hidden units, followed by a single output softmax layer representing the probability of the corresponding SDMH label.

5.3 Loss Function and Training Procedures

The final loss function is a weighted sum of Task 1 categorical cross-entropy loss and Task 2 categorical cross-entropy losses. Let α be the weight coefficient for Task 1 and $\lambda_1, \lambda_2, \ldots, \lambda_{13}$ be the weight coefficients for Task 2 (corresponding to the 13 SDMH labels). The multitask cascaded loss function can be defined as $\mathcal{L}_{\Updownarrow}$:

$$\mathcal{L}_{\Updownarrow} = -\alpha \sum_{i=1}^{N} y_i^0 \log(p_i) + \sum_{j=1}^{13} -\lambda_i \sum_{i=1}^{N} y_i^j \log(p_i) \qquad (1)$$

where y^0 refers to the output of task 1 and y^{1-13} refer to task 2 multilabel outputs. The model is trained using backpropagation and stochastic gradient descent (SGD) optimization, with the objective of minimizing the defined loss function. The training procedures involve iteratively updating the model parameters to maximize the overall performance in both Task 1 and Task 2 classification.

Note: The specific value of the weight coefficient α has been determined based on the relative importance of each task (.8) and λ_i has been determined based on hyperparameter tuning ranging from 0 to .5.

6. EXPERIMENTAL EVALUATION

6.1 Datasets

As part of this study, we have generated three datasets. **D1**: 200 small amount of lived experience texts along with human annotated labels of SDMHs. **D2**: 1498 lived experience texts with HLIA augmented expert annotation and solely LLM annotation of SDMHs. **D3**: 1498 LLM (ChatGPT) generated texts that has already HLIA augmented expert annotation as well as solely LLM annotation of SDMHs. of SDMHs. Fig. 3 shows HLIA based annotated class distributions in D2 dataset.

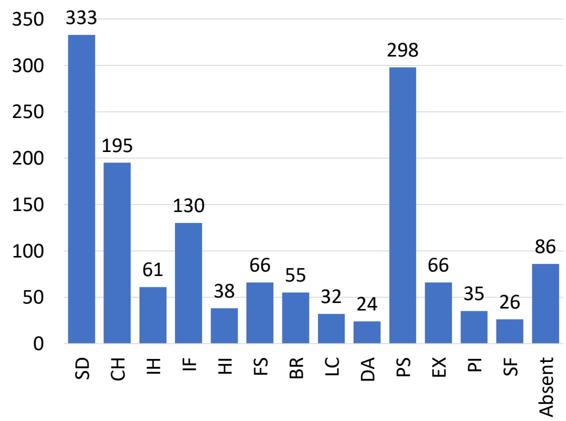

Figure 3: HLIA Annotated Class Distribution for 13 different SDMHs and Absence/Presence of any of the 13 SDMHs. Here, "SD", "CH", "IH", "IF", "HI", "FS", "BR", "LC", "DA", "PS", EX", "PI" and "SF" refer to socially disconnected or psychological symptoms, care handoff, impediments to healthcare, instability in finance, housing insecurity, food scarcity, brutality, legal challenges, drug abuse, excruciation, patient impairment and suicide fatality respectively

6.2 MTCNN Performance in Predicting SDMHs

We have implemented MTCNN algorithm using python-based tensorflow platform. Our MTCNN model training-testing policies are as follows:

6.2.1 *Training-Testing Policies*

- **(P1)**: Trained on D2 dataset (original lived experience texts) taking HLIA annotations as labels and tested on D2 dataset with HLIA annotations as labels (70%/30% split)
- **(P2)**: Trained on D2 dataset taking HLIA annotations as labels (100%) and tested on D1 dataset (human annotated labels).
- **(P3)**: Trained on D2 dataset taking HLIA annotations as labels (100%) and tested on D3 dataset (ChatGPT generated lived experience equivalent texts).

6.2.2 *Baseline Algorithms*

We implement the following benchmark algorithms for multi-label problem using huggingface and pytorch libraries (problem_type = 'multi_ label_ classification').

- **BERT**: The Bidirectional Encoder Representations from Transformers (BERT) [8], a language model pre-trained on unlabeled English texts, has garnered significant attention in the field of natural language processing. Transformers [41] served as the basis for its pre-training. BERT's primary pre-training objective revolves around acquiring contextualized representations of words, proving to be valuable for various downstream applications. Its outstanding performance across numerous natural language understanding tasks has been widely acknowledged [8].
- **ALBERT**: A variant of BERT, known as A Lite BERT (ALBERT), was introduced with the aim of reducing parameter size and minimizing memory usage. ALBERT, featuring 12 million parameters, stands in contrast to the 110 million parameters found in the original BERT model. This design is particularly advantageous in scenarios with constrained computing memory, such as low-resource settings [21].
- **RoBERTa**: The Robustly optimized BERT approach (RoBERTa), which is a modified version of BERT with adjusted pretraining objectives for enhanced robustness, has demonstrated superior performance compared to BERT across various Natural Language Processing (NLP) benchmark tasks [26].

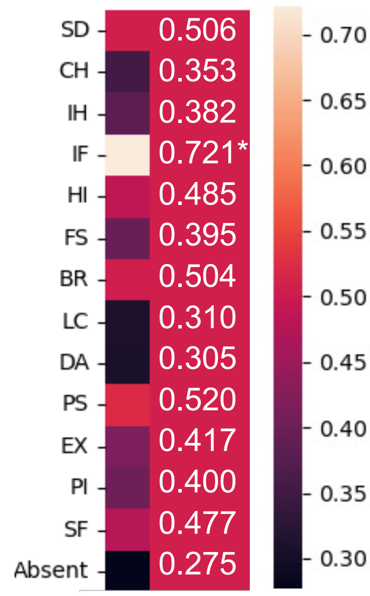

Figure 4: Pearson correlation coefficient with the significance pointer s(*) at the .05 level, p<.05

6.2.3 *Correlation among annotations*

Fig. 4 presents the Pearson correlation coefficient (r) of HLIA augmented annotation and LLM-only annotation in terms of 13 different SDMHs and existence/absence of any SDMHs (significant at p<0.05 has been pointed as *). We can observe that in case of LLM generated annotation, only one class (IF: Instability in Finance) has higher correlation with HLIA augmented annotation, others are not significantly correlated, that signifies that LLM generated annotation is still highly erroneous.

6.2.4 *Classification Performance*

Table 1 presents the accuracy, F-1, precision and recall performance comparisons among plain BERT, ALBERT and RoBERTa based multi-label classification as well as comparisons among different versions of proposed MTCNN framework taking three different pretrained embeddings (BERT, ALBERT and RoBERTa) as backbone as stated in Section 6.2 model architecture. We can see that MTCNN models significantly outperforms their corresponding baseline multilabel algorithms while RoBERTa backbone embedding provides the higehst accuracy among all.

Fig. 5 presents the MTCNN performance accuracies on each of 13 SDMHs in terms of three different policies. We can observe

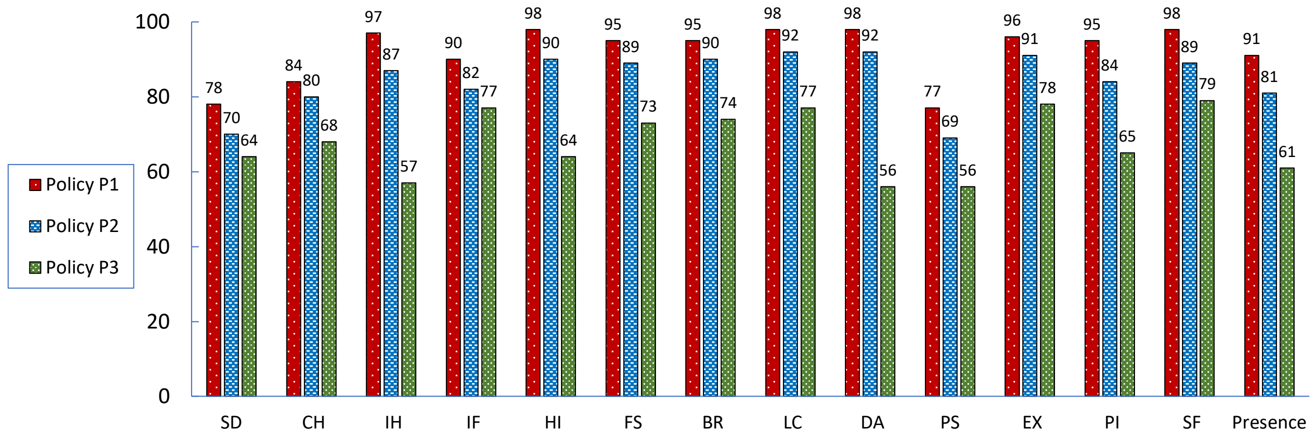

Figure 5: Accuracies of Social determinants of mental health (13 SDMHs) across three different policies (P1, P2 and P3)

that for all of them, policy P1 (trained and tested on D2 dataset) provides higher accuracy, that signifies that HLIA model provide appropriate pattern. Also, policy P2 (trained on D2 and tested on D1) provides slightly lower, but, still significantly close to P1, that signifies that our proposed HLIA produced annotation can predict human annotated texts with higher accuracy. However, for policy P3 (trained on D2 and tested on D3), accuracy decreases significantly, which signifies that HLIA proposed annotation based MTCNN model cannot predict ChatGPT produced texts accurately. Table 2 show details accuracy, precision, recall and f-1 measure of our proposed training policies P1, P2 and P3 respectively.

Class	Acc	F-1	Prec	Recall
BERT	.69	.78	.65	.68
ALBERT	.72	.71	.71	.72
RoBERTa	.70	.79	.70	.70
BERT+MTCNN	.90	.91	.91	.94
ALBERT+MTCNN	.89	.90	.90	.92
RoBERTa+MTCNN	.92	.92	.92	.97

Table 1: (P1) Details performance of multi-label version of baseline algorithms vs proposed MTCNN with different pre-trained embedding as backbone. Here, the model was trained on D2 Dataset and HLIA annotated labels; tested on D2 dataset and HLIA annotated labels with 70-30 splits

7. HUMAN-LLM INTERACTION (HLI): ANNOTATORS' EXPERIENCE STUDY

While Human-LLM Interaction (HLI) is a brand new area of research, our 4 annotators' invaluable experience in solving a real-world annotation problem can be beneficial for future LLM researchers. Annotators participated in this study expressed that, ChatGPT is extremely powerful in annotating texts as it can conversationally learn some of the examples, definitions of each class, potential keywords; and as per request, it can provide multitask annotation with explanation simultaneously. This made it more efficient for handling complex annotation tasks that involve multiple aspects or classes. ChatGPT based annotation offered annotators with consistent annotation based on its trained knowledge and language understanding. It eliminated potential human biases or variations that can arise when multiple human annotators are involved. This annotation through ChatGPT helped in processing and analyzing the text quickly, making it suitable for annotation tasks that require fast turnaround times. The annotators felt that, ChatGPT augmented text annotation of SDMHs helped them label 10 times faster than without ChatGPT help. Moreover, though there had been a lot of mis-annotation with ChatGPT, the tool's efficient explanation helped spot easily the mis-annotation, further enhances the annotation accuracy 2 times more accurately than their normal annotation skills. They also supported the notion that this paper proposed HLIA method is capable of teaching non-experts enough about any topic through its classification and explanation, which can be the key to the next generation citizen science research. The annotators also felt that over time, with the explanations from ChatGPT, they became more efficient in annotating texts.

8. LIMITATIONS

ChatGPT displayed variability in its annotations even when presented with identical input. While AI effectively reduced human bias and efforts in the annotation process, inconsistencies were evident in the use of ChatGPT. Despite these limitations, ChatGPT presents promising prospects for capturing social determinants of mental health, facilitating prompt intervention when necessary. It is important to acknowledge, however, that human intervention has not yet reached a stage where it can be entirely replaced by AI. Through collaborative efforts, this approach may contribute to a more expedited response in conjunction with human involvement. For instance, we defined the social determinants for mental health from Class 1 to Class 13 (Appendix Table 3) and sought the assistance of ChatGPT to obtain annotated labels for the given text:

Text: *"I worked with my sister to resolve a family conflict related to my mother's health. She did not fulfill her responsibilities, leading to my mother's deteriorating health. I realized that I cannot trust and rely on my sister. I would have sought other individuals to take care of my mother's health and manage the situation. Moving forward, I will identify 2-3 reliable people and maintain regular contact with them to ensure that such incidents do not recur."*

Upon analyzing this text with ChatGPT, we observed that it was classified as Class 2: Transition of care, while the other labels were deemed irrelevant. Subsequently, when generating a response for

	P1				P2				P3			
Class	Acc	F-1	Prec	Recall	Acc	F-1	Prec	Recall	Acc	F-1	Prec	Recall
SD	.78	.80	.84	.87	.70	.69	.74	.75	.64	.68	.74	.70
CH	.84	.90	.91	.86	.80	.88	.85	.80	.68	.67	.69	.73
IH	.97	.98	.97	.96	.87	.89	.90	.88	.57	.76	.67	.66
IF	.90	.91	.90	.91	.82	.81	.80	.82	.77	.74	.70	.72
HI	.98	.98	.98	.98	.90	.91	.91	.91	.64	.67	.66	.68
FS	.95	.96	.99	.97	.89	.88	.82	.89	.73	.66	.74	.87
BR	.95	.94	.95	.96	.90	.91	.88	.90	.74	.73	.85	.66
LC	.98	.96	.96	.98	.92	.93	.94	.91	.77	.77	.78	.65
DA	.98	.96	.97	.96	.92	.93	.91	.91	.56	.76	.56	.43
PS	.77	.76	.74	.79	.69	.70	.72	.70	.56	.58	.77	.35
EX	.96	.96	.98	.99	.91	.91	.93	.91	.78	.88	.71	.73
PI	.95	.96	.96	.96	.84	.83	.84	.84	.65	.80	.78	.63
SF	.98	.98	.98	.98	.89	.88	.87	.89	.79	.60	.59	.70
Presence	.91	.91	.94	.95	.81	.82	.82	.84	.61	.74	.60	.89

Table 2: Consolidated performance details of MTCNN trained on D2 Dataset and annotated labels; P1 tested on D2 dataset with HLIA annotated labels, P2 tested on D1 dataset with HILA annotated labels, and P3 tested on D3 dataset manually labeled by human only, all with 70-30 splits.

the same text, ChatGPT indicated that it did not correspond to any of the established classes. In another attempt, ChatGPT categorized the text under a different annotation as Class 3: Barriers to care. Based on our experiences, we found that the ChatGPT annotated labels are not reliable. Despite providing clear instructions, including the name, definition, and keywords for each class label, ChatGPT occasionally misclassified the labels or failed to assign any label at all. Therefore, we conclude that ChatGPT lacks true human understanding and context. It relies solely on patterns and information from its training data, which limits its ability to fully grasp the nuanced or subjective aspects of text. As a result, occasional misinterpretations or inaccurate annotations, as demonstrated in the aforementioned case, can occur.

9. CONCLUSION

This study explores the utilization of Large Language Models (LLMs) in the development of a SDMHs prediction system using students' experiential essays. The proposed Human-LLM Interaction for Annotation (HLIA) framework and the Multitask Cascaded Neural Network (MTCNN) classification algorithm are introduced to enhance the annotation and prediction processes. However, the findings indicate that LLMs, such as ChatGPT, may lack true human understanding and context, resulting in occasional misinterpretations and inaccurate annotations. These limitations highlight the need for human expertise and thorough evaluation when incorporating LLMs into complex tasks like mental health determinants prediction. Future research should focus on refining LLMs to address the challenges of authenticity, sentiment distortion, and hidden biases, ensuring their reliable and ethical use in sensitive domains such as mental health.

10. ACKNOWLEDGEMENTS

Due to limitations imposed by the Institutional Review Board (IRB) and our company's policies, we are unable to share the specific lived experience data that was collected.

The research adheres to IRB guidelines, ensuring informed consent and anonymization of personal data to protect participants' privacy and confidentiality. Data protection measures and bias mitigation strategies were employed, involving mental health experts to review and verify LLM-generated labels for fairness and accuracy. Algorithmic transparency and explainability were prioritized, explaining the reasoning behind LLM predictions. The study also emphasizes the responsible use of LLMs, acknowledging their limitations and validating results with human expertise, while maintaining ethical standards in data collection, storage, and analysis.

11. REFERENCES

[1] A. Atri, M. Sharma, and R. Cottrell. Role of social support, hardiness, and acculturation as predictors of mental health among international students of asian indian origin. *Int. Q. Community Health Educ.*, 27(1):59–73, 2006.

[2] A. Azaria, R. Azoulay, and S. Reches. Chatgpt is a remarkable tool – for experts, 2023.

[3] A. Bhattacharjee, S. M. T. Haque, A. Hady, S. M. R. Alam, M. Rabbi, M. A. Kabir, and S. I. Ahmed. Understanding the social determinants of mental health of the undergraduate students in bangladesh: Interview study. *CoRR*, abs/2109.02838, 2021.

[4] T. Bikaun, M. Stewart, and W. Liu. QuickGraph: A rapid annotation tool for knowledge graph extraction from technical text. pages 270–278, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[5] A. K. Biswas, G. Verma, and J. O. Barber. Improving ethical outcomes with machine-in-the-loop: Broadening human understanding of data annotations. *CoRR*, abs/2112.09738, 2021.

[6] D. Brook Weiss, P. Roit, O. Ernst, and I. Dagan. Extending multi-text sentence fusion resources via pyramid annotations. In *North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1854–1860, Seattle, United States, July 2022. Association for Computational Linguistics.

[7] R. Bruffaerts, P. Mortier, G. Kiekens, R. P. Auerbach, P. Cuijpers, K. Demyttenaere, J. G. Green, M. K. Nock, and R. C. Kessler. Mental health problems in college freshmen: Prevalence and academic functioning. *J. Affect. Disord.*, 225:97–103, Jan. 2018.

[8] J. Devlin, M. Chang, K. Lee, and K. Toutanova. 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, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1*, pages 4171–4186. Association for Computational Linguistics, 2019.

[9] H. Dong, W. Wang, K. Huang, and F. Coenen. Joint multi-label attention networks for social text annotation. 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)*, pages 1348–1354, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[10] D. Dorr, C. A. Bejan, C. Pizzimenti, S. Singh, M. Storer, and A. Quinones. Identifying patients with significant problems related to social determinants of health with natural language processing. *Stud. Health Technol. Inform.*, 264:1456–1457, Aug. 2019.

[11] E. Ferrara. Should chatgpt be biased? challenges and risks of bias in large language models, 2023.

[12] B. Funk and S. Sadeh-Sharvit. A framework for applying natural language processing in digital health interventions. *Journal of Medical Internet Research*, 22(8), 2020.

[13] T. Goyal, J. J. Li, and G. Durrett. FALTE: A toolkit for fine-grained annotation for long text evaluation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 351–358, Abu Dhabi, UAE, Dec. 2022. Association for Computational Linguistics.

[14] H. Hassani and E. S. Silva. The role of ChatGPT in data science: How AI-assisted conversational interfaces are revolutionizing the field. *Big Data and Cognitive Computing*, 7(2):62, Mar. 2023.

[15] X. He, Z. Lin, Y. Gong, A. Jin, H. Zhang, C. Lin, J. Jiao, S. M. Yiu, N. Duan, and W. Chen. Annollm: Making large language models to be better crowdsourced annotators. *CoRR*, abs/2303.16854, 2023.

[16] M. Heck, N. Lubis, B. Ruppik, R. Vukovic, S. Feng, C. Geishauser, H.-C. Lin, C. van Niekerk, and M. Gašić. Chatgpt for zero-shot dialogue state tracking: A solution or an opportunity?, 2023.

[17] Y. Huang, Q. Zhang, P. S. Y, and L. Sun. Trustgpt: A benchmark for trustworthy and responsible large language models, 2023.

[18] M. Karanouh. Mapping chatgpt in mainstream media: Early quantitative insights through sentiment analysis and word frequency analysis, 2023.

[19] S. D. Kauer, S. C. Reid, A. H. D. Crooke, A. Khor, S. J. C. Hearps, A. F. Jorm, L. Sanci, and G. Patton. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. *J. Med. Internet Res.*, 14(3):e67, June 2012.

[20] G. Kiekens, L. Claes, K. Demyttenaere, R. P. Auerbach, J. G. Green, R. C. Kessler, P. Mortier, M. K. Nock, and R. Bruffaerts. Lifetime and 12-month nonsuicidal self-injury and academic performance in college freshmen. *Suicide Life Threat. Behav.*, 46(5):563–576, Oct. 2016.

[21] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut. ALBERT: A lite BERT for self-supervised learning of language representations. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.

[22] Y. Li, B. Yu, L. Quangang, and T. Liu. FITAnnotator: A flexible and intelligent text annotation system. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations*, pages 35–41, Online, June 2021. Association for Computational Linguistics.

[23] Y. Li and Y. Zhang. Fairness of chatgpt, 2023.

[24] Y. Lin, T. Ruan, M. Liang, T. Cai, W. Du, and Y. Wang. DoTAT: A domain-oriented text annotation tool. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 1–8, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[25] S. Liu, Y. Sun, B. Li, W. Wang, F. T. Bourgeois, and A. G. Dunn. Sent2Span: Span detection for PICO extraction in the biomedical text without span annotations. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1705–1715, Punta Cana, Dominican Republic, Nov. 2021. Association for Computational Linguistics.

[26] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692, 2019.

[27] S. M. Martinez, E. A. Frongillo, C. Leung, and L. Ritchie. No food for thought: Food insecurity is related to poor mental health and lower academic performance among students in california's public university system. *J. Health Psychol.*, 25(12):1930–1939, Oct. 2020.

[28] M. Mofatteh. Risk factors associated with stress, anxiety, and depression among university undergraduate students. *AIMS Public Health*, 8(1):36–65, 2021.

[29] OpenAI. Chatgpt (mar 14 version) [large language model]. https://chat.openai.com/chat, 2023.

[30] F. Panagopoulou, C. Parpoula, and K. Karpouzis. Legal and ethical considerations regarding the use of chatgpt in education, 2023.

[31] J. M. Pérez, J. C. Giudici, and F. Luque. pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks. 2021.

[32] A. Richter, M. Sjunnestrand, M. Romare Strandh, and H. Hasson. Implementing school-based mental health services: A scoping review of the literature summarizing the factors that affect implementation. *Int. J. Environ. Res. Public Health*, 19(6):3489, Mar. 2022.

[33] C. Schröder, K. Bürgl, Y. Annanias, A. Niekler, L. Müller, D. Wiegreffe, C. Bender, C. Mengs, G. Scheuermann, and G. Heyer. Supporting land reuse of former open pit mining sites using text classification and active learning. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 4141–4152. Association for Computational Linguistics, 2021.

[34] T. Searle, Z. Kraljevic, R. Bendayan, D. Bean, and R. J. B. Dobson. Medcattrainer: A biomedical free text annotation interface with active learning and research use case specific customisation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019 - System Demonstrations*, pages 139–144. Association for Computational Linguistics, 2019.

[35] G. Shi, D. Gao, S. Ma, M. Yang, Y. Xiao, and X. Xie. Mathematical characterization of signal semantics and rethinking of the mathematical theory of information, 2023.

[36] R. G. Silva and M. Figueiredo-Braga. Evaluation of the relationships among happiness, stress, anxiety, and depression in pharmacy students. *Curr. Pharm. Teach. Learn.*, 10(7):903–910, July 2018.

[37] Y. Sun, J. He, S. Lei, L. Cui, and C.-T. Lu. Med-mmhl: A multi-modal dataset for detecting human- and llm-generated misinformation in the medical domain, 2023.

[38] S. Tian, Q. Jin, L. Yeganova, P.-T. Lai, Q. Zhu, X. Chen, Y. Yang, Q. Chen, W. Kim, D. C. Comeau, R. Islamaj, A. Kapoor, X. Gao, and Z. Lu. Opportunities and challenges for chatgpt and large language models in biomedicine and health, 2023.

[39] A. Tsvigun, L. Sanochkin, D. Larionov, G. Kuzmin, A. Vazhentsev, I. Lazichny, N. Khromov, D. Kireev, A. Rubashevskii, and O. Shahmatova. Altoolbox: A set of tools for active learning annotation of natural language texts. In *Proceedings of the The 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022 - System Demonstrations, Abu Dhabi, UAE, December 7-11, 2022*, pages 406–434. Association for Computational Linguistics, 2022.

[40] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017.

[41] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008, 2017.

[42] R. M. Viner, E. M. Ozer, S. Denny, M. Marmot, M. Resnick, A. Fatusi, and C. Currie. Adolescence and the social determinants of health. *Lancet*, 379(9826):1641–1652, Apr. 2012.

[43] G. Wang, W. Li, E. M.-K. Lai, and Q. Bai. Aakos: Aspect-adaptive knowledge-based opinion summarization, 2023.

[44] M. C. Whatnall, A. J. Patterson, S. Brookman, P. Convery, C. Swan, S. Pease, and M. J. Hutchesson. Lifestyle behaviors and related health risk factors in a sample of australian university students. *J Am. Coll. Health*, 68(7):734–741, Oct. 2020.

[45] Y. Yan, S. Huang, S. Chen, M. Liao, and J. Xu. Active learning with query generation for cost-effective text classification. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 6583–6590. AAAI Press, 2020.

[46] J. Yang, Y. Zhang, L. Li, and X. Li. YEDDA: A lightweight collaborative text span annotation tool. In *Proceedings of ACL 2018, System Demonstrations*, pages 31–36, Melbourne, Australia, July 2018. Association for Computational Linguistics.

[47] Z. Zheng, O. Zhang, C. Borgs, J. T. Chayes, and O. M. Yaghi. Chatgpt chemistry assistant for text mining and prediction of mof synthesis, 2023.

[48] J. W. Zimmerman, D. Hudon, K. Cramer, J. S. Onge, M. Fudolig, M. Z. Trujillo, C. M. Danforth, and P. S. Dodds. A blind spot for large language models: Supradiegetic linguistic information, 2023.

[49] M. Zlabinger, M. Sabou, S. Hofstätter, and A. Hanbury. Effective crowd-annotation of participants, interventions, and outcomes in the text of clinical trial reports. In *Findings of the Association for Computational Linguistics: EMNLP 2020, Online Event, 16-20 November 2020*, volume EMNLP 2020 of *Findings of ACL*, pages 3064–3074. Association for Computational Linguistics, 2020.

APPENDIX

Extracted Variables	Description	Sample Tokens/Keywords
(SD) Socially Disconnected	Identifying individuals' loneliness, absence of social connections or family assistance, their marital or relationship situation, Despair, diminished social engagement, abnormal eating patterns	Solitary, isolated, divorce, separated, widowed
(CH) Care Handoff	alteration in medication and/or healthcare professional;Modification in admission status (release, relocation, etc.)	Release,Enrollment,Medication alteration,Relocation
(IH) Impediments to healthcare	challenges in communication,individuals with intellectual disabilities,Difficulties with transportation, absence of trust or connection	Unintelligible speech,Transportation challenges,communication difficulties
(IF) Instability in Finance	Financial challenges, employment difficulties, impoverished conditions.	Jobless, impoverished, joblessness, reintegration, reemployment.
(HI) Housing insecurity	Housing problems, housing concerns	Unsheltered,Displacement,Houseless
(FS) Food scarcity	Poor dietary intake and/or nutrition, inadequate access to nutritious meals, reliance on food assistance programs or resources such as food charities, food vouchers, or food stamps.	starving, food storage,deprivation,meal coupon
(BR) Brutality	Availability and/or acquisition of deadly resources; harassment; intimate partner violence; any mistreatment/abuse/trauma; racial discrimination; thoughts of committing homicide; feeling fearful/vulnerable.	Guns, aggression, battery, armament, mistreatment, murderous, racial discrimination
(LC) Legal challenges	Incarceration; legal proceedings; confinements; punitive measures; protection orders; encounters with the justice system; criminal allegations; infractions of the law.	Incarceration, conditional release, indictable offense, under arrest, jail, probe etc
(DA) Drug abuse	Substance use disorder; alcohol abuse; alcohol use disorder; dependency; drug overdose	Liquor, smoking products, Narcotic, stimulant, Tobacco use, excessive dose etc.
(PS) Psychological symptoms	Hopelessness; Insomnia; Problem solving difficulty; decreased psychosocial, functioning; psychiatric hospitalization; eating disorder; mention of any psychiatric disease	Mental health hospitalization, Reference to a mental health condition, clinical depression, chronic sleep problems,nervousness,hallucinations and delusions,Psychotic disorder,traumatic stress syndrome
(EX) Excruciation	Pain physically	Agony, Suffer, Pangs, Discomfort, Anguish, Ache etc.
(PI) Patient Impairment	Dependency on compensation for disabilities, Service-connected disability ratings,Dependence on assistive devices	Visually impaired, Mobility aid, Handicapped, auditory impairment,
(SF) Suicide fatality	Suicidal behavior, Contemplation of self-harm	Life not worth living,profound lack of will to live,shoot myself

Table 3: (Appendix) Expert Identified Social Determinants of Mental Health and Tokens for NLP

Classify a text with the following binary classes with explanation for each of them, why ChatGPT classified them such?

class 0: No social determinants of mental health exists
class 1: one or more social determinants of mental health exist

Now, if ChatGPT classifies as class 1: one or more social determinants of mental health exist, please classify further one or more of the following classes of social determinants of mental health with appropriate explanation, why ChatGPT classified it such.

class 1: Socially Disconnected; description: Identifying individuals' loneliness, absence of social connections or family assistance, and their marital or relationship situation; Tokens/keywords: Solitary, isolated, separated, widowed.

class 2: Care Handoff; description: alteration in medication and/or healthcare professional, Modification in admission status (release, relocation, etc.); tokens/keywords: Release,Enrollment,Medication alteration,Relocation

class 3: Impediments to healthcare; description: challenges in communication, individuals with intellectual disabilities,Difficulties with transportation, absence of trust or connection; tokens/ keywords: Unintelligible speech,Transportation challenges,communication difficulties

class 4: Instability in Finance; description: Financial challenges, employment difficulties, impoverished conditions; tokens/keywords: Jobless, impoverished, joblessness, reintegration, reemployment

class 5: Housing insecurity; description: Housing problems, housing concerns; tokens/keywords: Unsheltered,Displacement,Houseless

class 6: Food scarcity; description: Poor dietary intake and/or nutrition, inadequate access to nutritious meals, reliance on food assistance programs or resources such as food charities, food vouchers, or food stamps; tokens/keywords: starving, food storage,deprivation,meal coupon

class 7: Brutality; Description: Availability and/or acquisition of deadly resources; harassment; intimate partner violence; any mistreatment/abuse/trauma; racial discrimination; thoughts of committing homicide; feeling fearful/vulnerable; tokens/keywords: Guns, aggression, battery, armament, mistreatment, murderous, racial discrimination

class 8: Legal challenges; description: Incarceration; legal proceedings; confinements; punitive measures; protection orders; encounters with the justice system; criminal allegations; infractions of the law; tokens/keywords: Incarceration, conditional release, indictable offense, under arrest, jail, probe etc

class 9: Drug abuse; description: Substance use disorder; alcohol abuse; alcohol use disorder; dependency; drug overdose; tokens/keywords: Liquor, smoking products, Narcotic, stimulant, Tobacco use, excessive dose

class 10: Psychological symptoms; description: Despair, diminished social engagement, cognitive problem-solving impediments, sleep deprivation, abnormal eating patterns, Mental health hospitalization, Reference to a mental health condition; tokens/keywords: clinical depression, chronic sleep problems,nervousness,hallucinations and delusions,Psychotic disorder,traumatic stress syndrome

class 11: Excruciation; description: Pain physically; tokens/keywords: Agony, Suffer, Pangs, Discomfort, Anguish, Ache

class 12: Patient Impairment; description: Dependency on compensation for disabilities, Service-connected disability ratings,Dependence on assistive devices; tokens/keywords: Visually impaired, Mobility aid, Handicapped, auditory impairment

class 13: Suicide fatality; description: Suicidal behavior, Contemplation of self-harm; tokens/keywords: Life not worth living,profound lack of will to live,shoot myself

Here is the text example to classify according to the above strategy

"Life was terribly unstructured for me. I was struggling with addiction, mental issues, homelessness, debt, and my world was so toxic. I felt like I was going around an endless cycle of dead ends. I had no direction, no hope, and the people around me didn't know how to support me nor had the time to as most of them were facing similar battles."

ChatGPT Response

Based on the provided text, ChatGPT would classify it as follows:

Binary classification:

- Class: 1 (One or more social determinants of mental health exist)
- Explanation: The text indicates the presence of multiple social determinants of mental health such as addiction, mental issues, homelessness, debt, and toxic surroundings. These factors suggest that the individual is experiencing challenges that can impact their mental well-being.

Further classification into specific classes of social determinants of mental health:

- Class: 2 (Care Handoff)
- Explanation: The text mentions struggling with addiction and mental issues, indicating a need for alteration in medication and healthcare professionals' involvement.
- Class: 4 (Instability in Finance)
- Explanation: The text mentions being homeless and facing debt, indicating financial challenges and employment difficulties.
- Class: 5 (Housing insecurity)
- Explanation: The text explicitly mentions being homeless, indicating housing problems and concerns.
- Class: 9 (Drug abuse)
- Explanation: The text mentions struggling with addiction, indicating substance use disorder and dependency.

It is important to note that the classification is based on the provided information and relies on the interpretation of the text by ChatGPT. The actual classification may vary depending on the context and additional information present in the text.

Figure 6: (Appendix) Our proposed Human-LLM Interaction for Annotation (HLIA) framework example interactions and generated text responses