**Annotation Process and Evaluation**

Identification of Posts Related to Depression Symptoms

# 1. Recruiting Process

We recruited annotators via direct email invitations sent to psychiatrists affiliated with well-known hospitals. Email addresses were obtained from publicly available hospital staff listings. A total of **15 invitation emails** were sent, and **5 responses** were received.

**Two psychiatrists consented to do the annotation**, each with more than 10 years of clinical practice and extensive experience diagnosing depression and related mood disorders. Both were provided with the Annotation Protocol and Instruction document and completed a qualification test consisting of 50 labeled examples. Both annotators achieved ≥95% accuracy compared with gold-standard labels.

## 1.1 Recruitment Email

**Subject:** Invitation to Participate in Annotation of Depression-Related Social Media Posts

**Dear Dr. [Surname],**

We are conducting a research project aimed at identifying and categorizing depression-related symptoms expressed in social media posts. The purpose of this study is to improve the detection and understanding of depressive symptoms in online contexts, which may inform future clinical interventions and public mental health strategies.

We are seeking experienced psychiatrists to serve as expert annotators for a dataset of anonymized social media posts. The task will involve:

* Completing an initial qualification test (50 labeled examples)
* Reviewing posts and determining whether they describe or imply depressive symptoms.

**Eligibility Criteria:**

* Licensed psychiatrist with at least 10 years of clinical experience in psychiatry.
* Prior experience of diagnosing depression and related mood disorders.
* Willingness to annotate approximately 5,000 posts over a period of 2 weeks.

**Compensation and Benefits:**

* **Honorarium**: RMB ¥500 upon successful completion of the task.
* **Data Access**: Full access to the finalized annotated dataset for academic research.
* **Collaboration Opportunities**: Possibility to join our research group in future collaborative projects in mental health and computational psychiatry.

If you are interested in participating, please reply to this email within two weeks.

We look forward to the possibility of working together.

Best regards,

Research Team

# 2. Annotation Process and Evaluation

From the full dataset, 5,000 posts were randomly sampled. Each post was annotated independently by the two psychiatrists according to the Annotation Protocol and Instruction. Annotation was conducted using a shared Excel file containing post IDs, content, and columns for each annotator’s label (“Yes” or “No”).

We computed Cohen’s Kappa to assess the inter-rater agreement before consensus discussion. The results are summarized below:

|  | Annotator B: Yes | Annotator B: No |
| --- | --- | --- |
| Annotator A: Yes | 295 | 8 |
| Annotator A: No | 10 | 4,687 |

* **Observed Agreement (Po)** = (295 + 4,687) / 5,000 = 0.9964
* **Expected Agreement (Pe)** = [(305 × 303) + (4,695 × 4,697)] / (5,000²) ≈ 0.9936
* **Cohen’s Kappa** = (Po − Pe) / (1 − Pe) ≈ **0.93**

Following the initial annotation, the annotators reviewed the **18** posts with conflicting labels. Disagreements were resolved through discussion. The final dataset consisted of:

* **303** symptom-related posts
* **4,697** non-symptom-related posts

These finalized labels were then used for subsequent analysis in the study.