ReDSM5: A Reddit Dataset for DSM-5 Depression Detection

Eliseo Bao, Anxo Pérez, Javier Parapar

IRLab, CITIC, Universidade da Coruña, Spain

Problem & Contribution

Problem

- **280 million** people worldwide affected by depression
- Social media detection lacks **DSM-5 grounding**
- **Binary classification** ignores specific **symptom patterns**
- Models lack **explainability** and **clinical trustworthiness**
- No **sentence-level datasets** with **expert rationales**

Contribution

- **Reddit dataset** with **DSM-5 annotations** at sentence level
- **1484 posts** labeled by a **professional psychologist**
- Each label includes **DSM-5 clinical rationale**
- Enables **detection** and **explainable reasoning** for symptoms

Dataset Overview

1484 Posts

1.39 Symp./Post

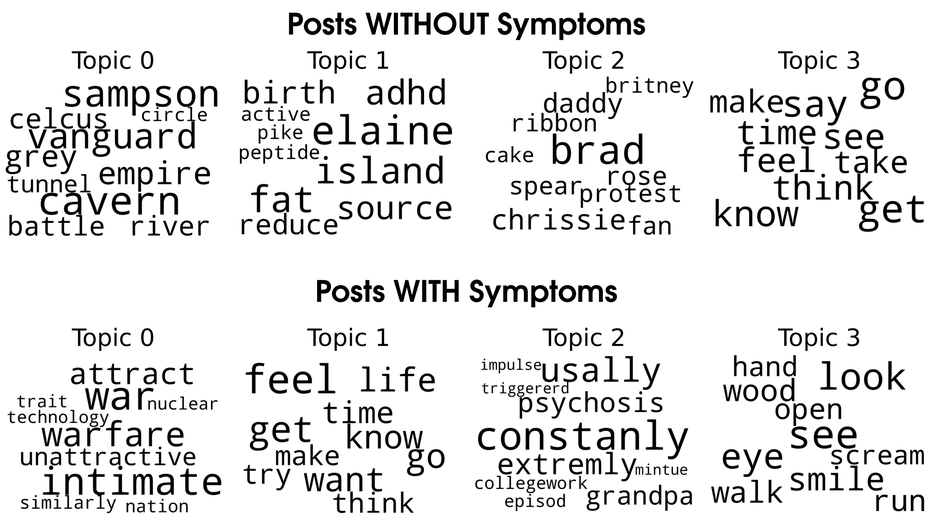
Most Frequent Symptoms

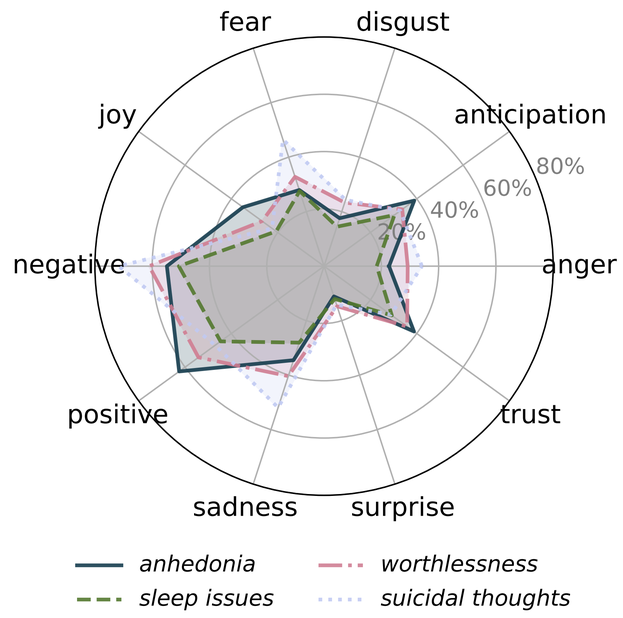
Depressed mood	**328**
Worthlessness	**311**
Suicidal thoughts	**165**
Anhedonia	124
Fatigue	124

392 posts with **no symptoms** (*hard negatives*)

Post length: **2 to 6990 words** (avg. 295)

How Depression Appears in Text





→ **Suicidal thoughts** show **highest fear and negative sentiment** across all symptoms

→ **Sleep issues** have most **TIME entity references**, reflecting detailed disruption descriptions

→ No-symptom posts use ∼50% **past tense** vs. symptom posts use >55% **present tense**

→ **DATE entities dominant** in depressed mood, anhedonia, suicidal thoughts (clinical temporal grounding)

→ **Worthlessness**: 67.5% **first-person singular**, showing strong self-critical focus

→ Symptom posts use **introspective verbs** (feel, know, want, think) vs. **gaming/hobby nouns**

Data Sample

Post excerpt:

"I feel like crying constantly, even during activities I used to enjoy. I find myself sleeping throughout the day just to avoid the overwhelming sadness. I feel completely worthless."

Sentence-level Annotations:

- **Depressed Mood**: "feel like crying constantly"
  Rationale: Persistent sadness with frequent crying episodes, indicative of depressed mood in major depressive disorder
- **Sleep Issues**: "sleeping throughout the day"
  Rationale: Excessive daytime sleep as avoidance behavior, consistent with hypersomnia pattern
- **Worthlessness**: "feel completely worthless"
  Rationale: Negative self-perception and feelings of inadequacy, meeting DSM-5 criterion 7

Baseline Performance

Multi-label Symptom Classification

Method	$F1_{micro}$	Acc.
SVM	0.39	0.22
CNN	0.25	0.14
BERT	0.51	0.41
Llama 3.2	**0.54**	**0.45**

Explanation Generation

0.78
Embedding Similarity
(to expert rationales)

0.62
Clinical Quality Score
(accuracy + coverage + clarity)

CIKM 2025, The 34th ACM International Conference on Information and Knowledge Management. November 10-14, 2025. Seoul, Republic of Korea.