



Task Overview

Psychological defenses are the “immune system” of the mind, shaping what speakers disclose and how they accept or resist help. Despite their critical role in mental health and counseling, defensive functioning remains largely unmodeled in current emotional support conversation systems.

This shared task invites participants to bridge the gap between clinical theory and NLP by analyzing the PSYDEFCONV dataset. Participants will work with multi-turn dialogues to identify the specific defense level of a target utterance given its context. The goal is to develop models that can recognize subtle, context-dependent defensive maneuvers—ranging from adaptive coping to immature distortion.

Example

Input (dialogue excerpt)

Supporter: Hello, how are you today?

Seeker: yea good but lil lazy could you please help me to be concentrated on work and new job preparation to get a job.

Supporter: Sounds like you need a little help staying on track and feeling more productive?

Seeker (target utterance): yes. i am so attracted to baby... i think nobody will take care of baby as much as i can.
So unable to concentrate on job preparation.

Output (defense label)

Level 1 Action Defenses

The model reads the dialogue plus the target utterance and returns one label from the table above. In this case the seeker diverts frustration about job preparation into protective, action-oriented statements about the baby, so the correct prediction remains *Action Defenses*.

Data and Labels

This task utilizes **PSYDEFCONV**, the first conversational dataset annotated with defense levels based on the DMRS. The dataset is constructed from a stratified subset of the ESConv corpus to ensure diverse coverage of problem types and emotions. The corpus contains 200 dialogues and 4,709 total utterances, including 2,336 help-seeker turns annotated for defense levels.

Participants must classify utterances into 9 categories, comprising seven hierarchical levels of defensive maturity and two auxiliary labels.

LEVEL	MECHANISMS	DEFINITION
0 · No Defenses	N/A	Functional utterances that maintain conversational flow without engaging conflict.
1 · Action Defenses	Passive Aggression, Help-Rejecting Complaining, Acting Out	Distress is released by acting on the environment instead of reflecting.
2 · Major Image-Distorting	Splitting (self/other), Projective Identification	Reduces anxiety via all-good/all-bad distortions of self or other.
3 · Disavowal Defenses	Denial, Rationalization, Projection, Autistic Fantasy	Rejects threatening reality by denying, excusing, blaming, or fantasizing.
4 · Minor Image-Distorting	Devaluation/Idealization (self or other), Omnipotence	Softer distortions temporarily inflate or deflate self-esteem.
5 · Neurotic Defenses	Repression, Dissociation, Reaction Formation, Displacement	Keeps unacceptable motives out of awareness; feelings surface indirectly.
6 · Obsessional Defenses	Isolation of Affect, Intellectualization, Undoing	Uses excessive logic or symbolic acts to separate feelings from events.
7 · High-Adaptive Defenses	Affiliation, Altruism, Anticipation, Humor, Self-Assertion, Self-Observation, Sublimation, Suppression	Mature coping that integrates emotion and thought to channel affect constructively.
8 · Needs More Information	N/A	Label used when an utterance is too ambiguous or lacks context.

For a detailed definition, check out our [Annotation Handbook!](#)

Baseline and Evaluation Metrics

Baseline runs and official metrics are published on our [CodaBench Page](#).