Supplemental Material

Cognitive Distortion Definitions

Below are definitions of cognitive distortion types used in the study. Included are examples of labeled cognitive distortions identified in client text messages.

Mental Filter: focusing on one detail of a situation exclusively so that all other details are taken out of context or skewed to reflect the feeling associated with that one detail. Also referred to as selective abstraction.

Examples:

- "I can't say I have anything to be grateful for today."
- "I'm doing better now, but it won't last."

Jumping to Conclusions: Making an interpretation of a situation without definite facts or supporting evidence to justify that interpretation. This includes instances of predicting how someone else feels/thinks/reacts without evidence (i.e., mind reading) and predicting how future events will turn out with limited supporting evidence (i.e., fortune telling).

Examples:

- "My mom thinks I have no talent and won't help me."
- "They weren't acting normal today, I'm not sure it was them."

Catastrophizing (Magnification/Minimization): Inappropriately exaggerating or discounting the importance of an event in one's life.

Examples:

- "I just feel so emotional right now, my life is falling apart."
- "I'm low on my meds and need more. If I don't get more, I won' be able to do anything and it will be ruined".

Should Statements: trying to motivate oneself with absolute expectations of behavior or holding others to expectations driven by absolutist expectations (i.e., should, must, ought).

Examples:

- "The team is stopping by so I feel like I should have to have my shit together. Things need to look good."
- "I should be farther along, I can't be in this program anymore."

Overgeneralization: Extending a single occurrence or isolated incident as evidence of an ongoing or never-ending pattern.

Examples:

- "Its always hard to depend on people, they never follow through."
- "I'm never comfortable around others. I don't like people."

Supplemental Material

Supplemental table 1, table 2, and table 3 present the tuned model parameters for the three prediction frameworks, LR, *SVM*, and *BERT* respectively. To address the infrequency of the cognitive distortions (positive class), experimentation included two different loss functions: 1) *uniform* weighting across positive and negative classes, and *balanced* weighting where loss weights are inversely proportional to the class frequency.

Table 1. LR configuration

Parameter	Iteration				
	#1	**#2**	**#3**	**#4**	**#5**
regularization strength, C	1.0	1.0	1.0	1.0	1.0
n values for n-grams	1	1	1, 2	1, 2	1, 2
positive class loss weight	balanced				
regularization type	L2				
solver	Limited-memory BFGS				

Table 2. SVM configuration

Parameter	Iteration				
	#1	**#2**	**#3**	**#4**	**#5**
regularization strength, C	100	10	10	10000	10000
positive class loss weight	uniform	uniform	balanced	balanced	uniform
kernel	Radial Basis Function				

Table 3. BERT configuration

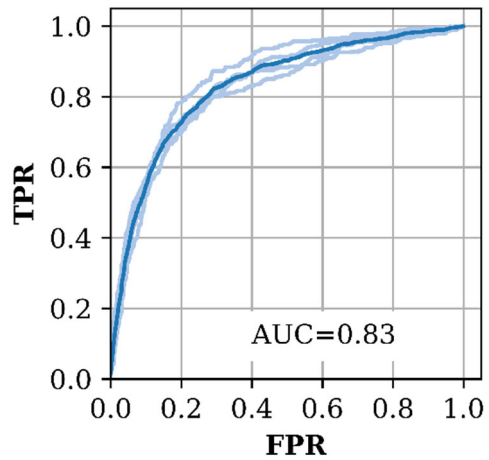| Parameter	Iteration				
	#1	**#2**	**#3**	**#4**	**#5**
number of epochs	14	14	10	14	8
dropout	0.2	0.3	0.1	0.2	0.2
optimizer	AdamW				
learning rate	5e-5				
number of warm-up steps	0				
maximum gradient norm	1.0				
maximum sequence	120 word pieces				
batch size	50				
positive class loss weight	balanced				
pre-trained model	BERT-base-uncased				

Supplemental Material

Figure 1. Receiver operating characteristic the "any distortion" label achieved using the BERT model. The light blue curves or associated with the individual iterations, and the dark blue curves is the average across the iterations.

Supplemental Material

Mean distribution of distortions by type in patient-therapist conversations.

metric	MF	JC	C	SM	O	AD
mean	17.1	16.1	29.5	5.2	7.1	48.5
std	24.0	35.9	90.1	10.8	13.3	106.1
min	0.0	0.0	0.0	0.0	0.0	0.0
25%	1.0	1.0	1.0	0.0	0.0	5.0
50%	6.0	2.0	5.0	0.0	1.0	11.0
75%	26.0	19.0	24.0	5.0	10.0	62.0
max	89.0	206.0	546.0	52.0	67.0	633.0

Supplemental Material

Distortion	SHAP Explainer
mental filter	Good morning how are you? I couldn't fall asleep last night until 1am. And this morning when I got up I got very dizzy almost like I was going to pass out. I passed out once when I was 11. I remember how I felt right before and that is how I felt this morning. It was very disconcerting, so I don't think I'm doing to do much today.
jumping to conclusions	Maybe they really don't like me.
catastrophizing	Feeling unsure of myself right now. Desperate. Idk if I can handle my money or tackle my goals of saving for a rainy day.
should statements	Yes, but I've got to be able to make better decisions for myself, and not let people use my issues for their personal shit
Over-generalizing	I've never been able to rest in my life the way I wanted to.

Figure 2. SHAP feature interpretability plot for each of cognitive distortion types.