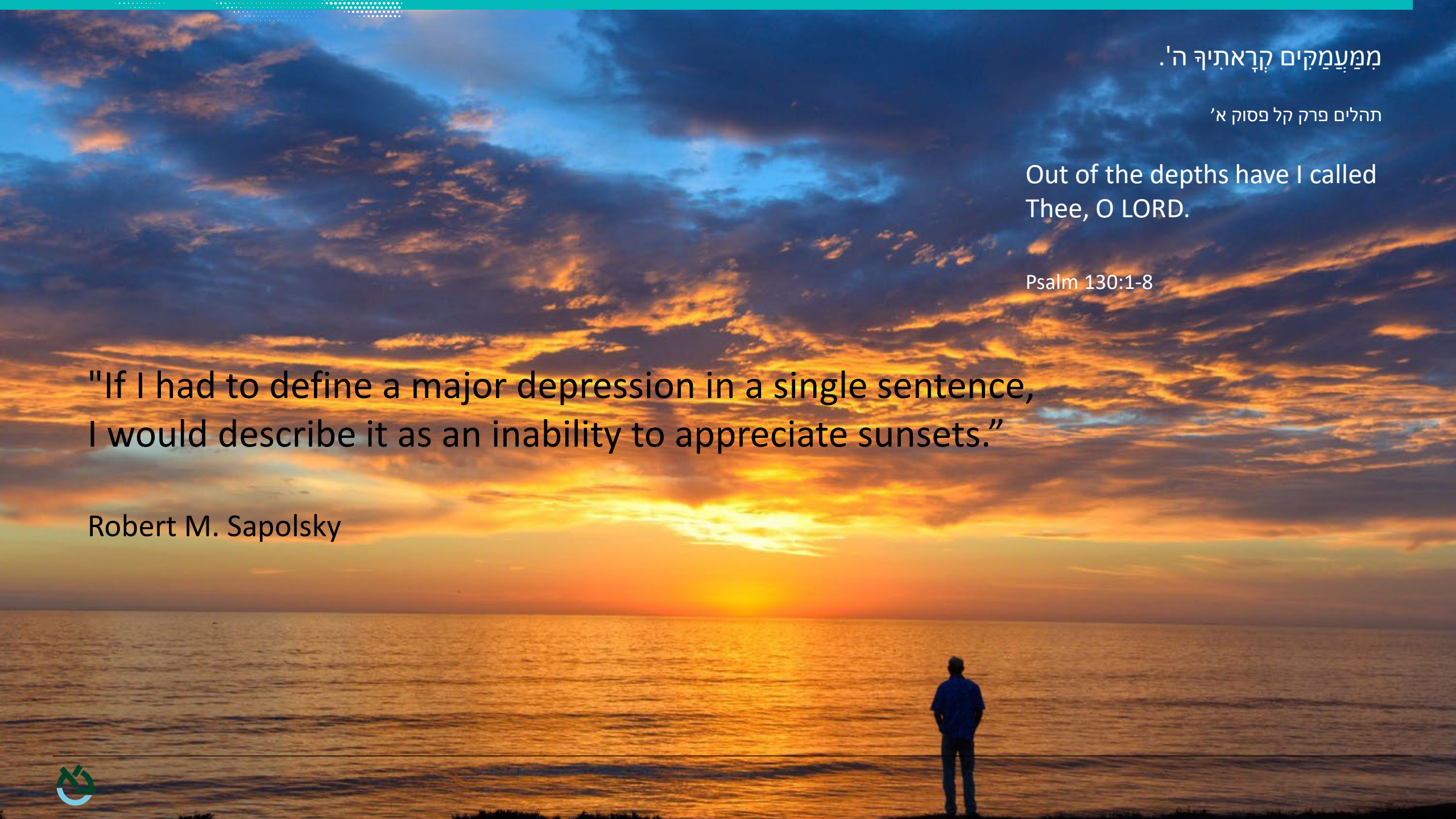
Integrating Clinical Science and Artificial Intelligence to Uncover the Dynamics Underlying Therapeutic Change

Dana Atzil-Slonim

Psychology Department

Bar-Ilan University



מִמַּעֲמַקִּים קְרָאתִיךָ ה'.

תהלים פרק קל פסוק א'

Out of the depths have I called Thee, O LORD.

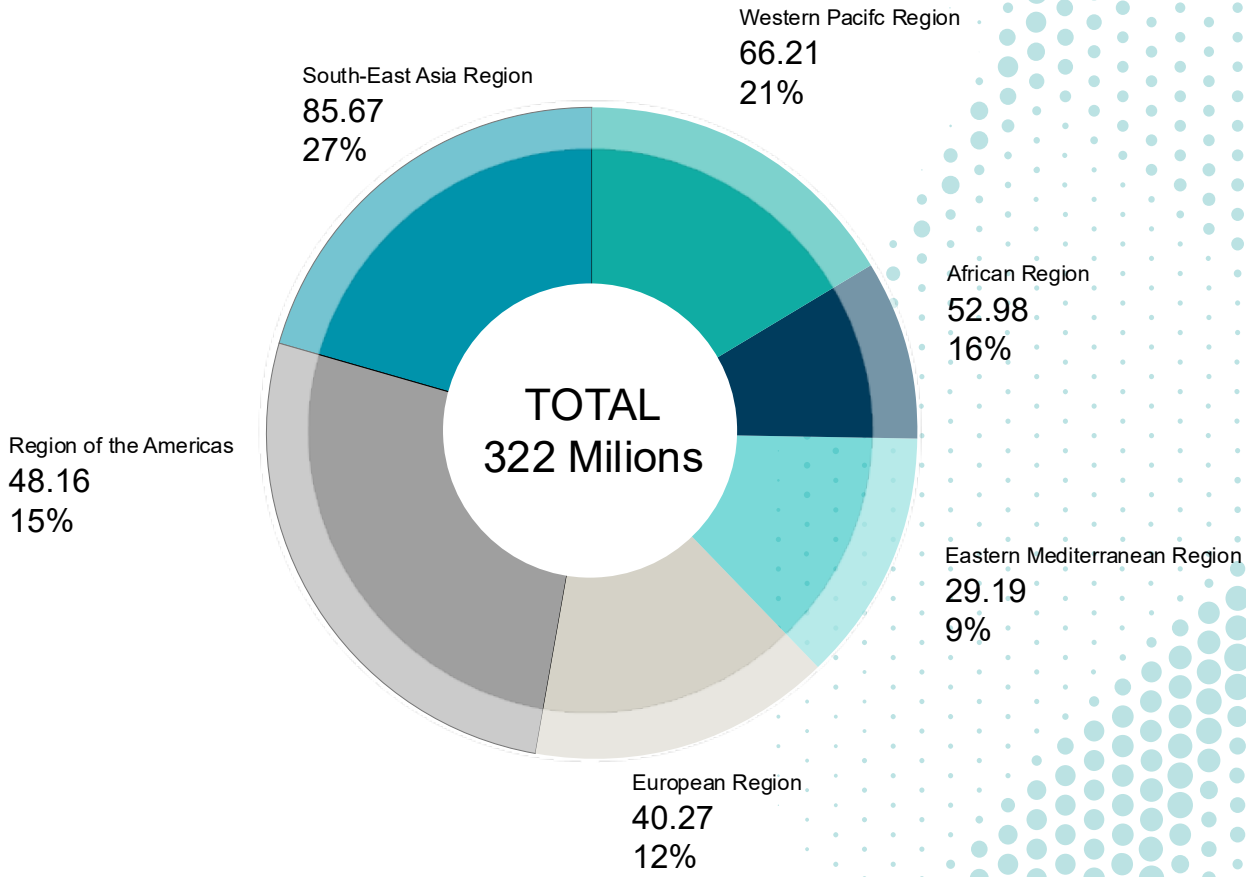
Psalm 130:1-8

"If I had to define a major depression in a single sentence,
I would describe it as an inability to appreciate sunsets."

Robert M. Sapolsky

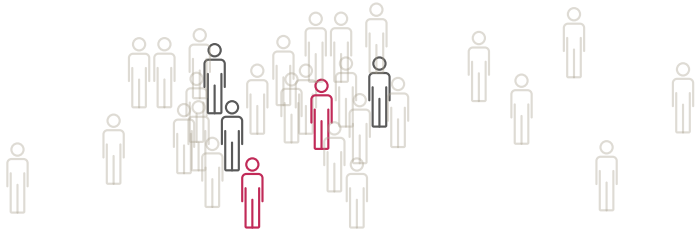
Depression is Common and Serious

- Epidemic: Globally, around **322M**
- **1:5** women, **1:10** men
- **On the rise** in industrialized societies
- **#1 global** health burden in 2030
- **1M** die by suicide every year

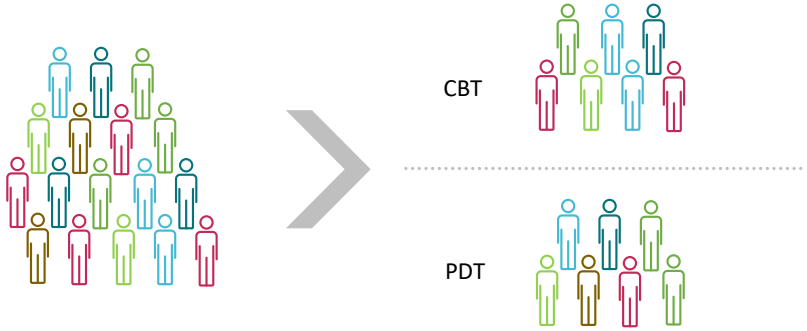


Challenges for the diagnosis and treatment of depression

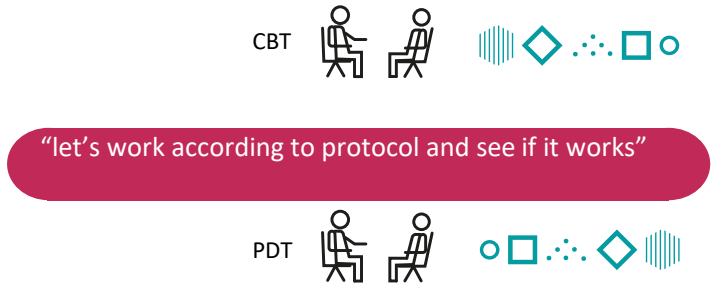
- **Undetected/ limited accessibility**



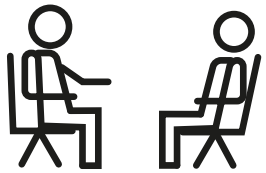
- **Sub-optimally diagnosed**



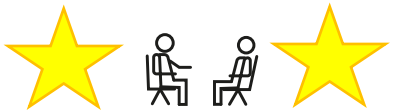
- **Sub-optimally treated**



- **Therapists differ widely in effectiveness**



50% success



80% success

Can we uncover the mechanisms serving this super shrinks?

Challenges for the research of depression

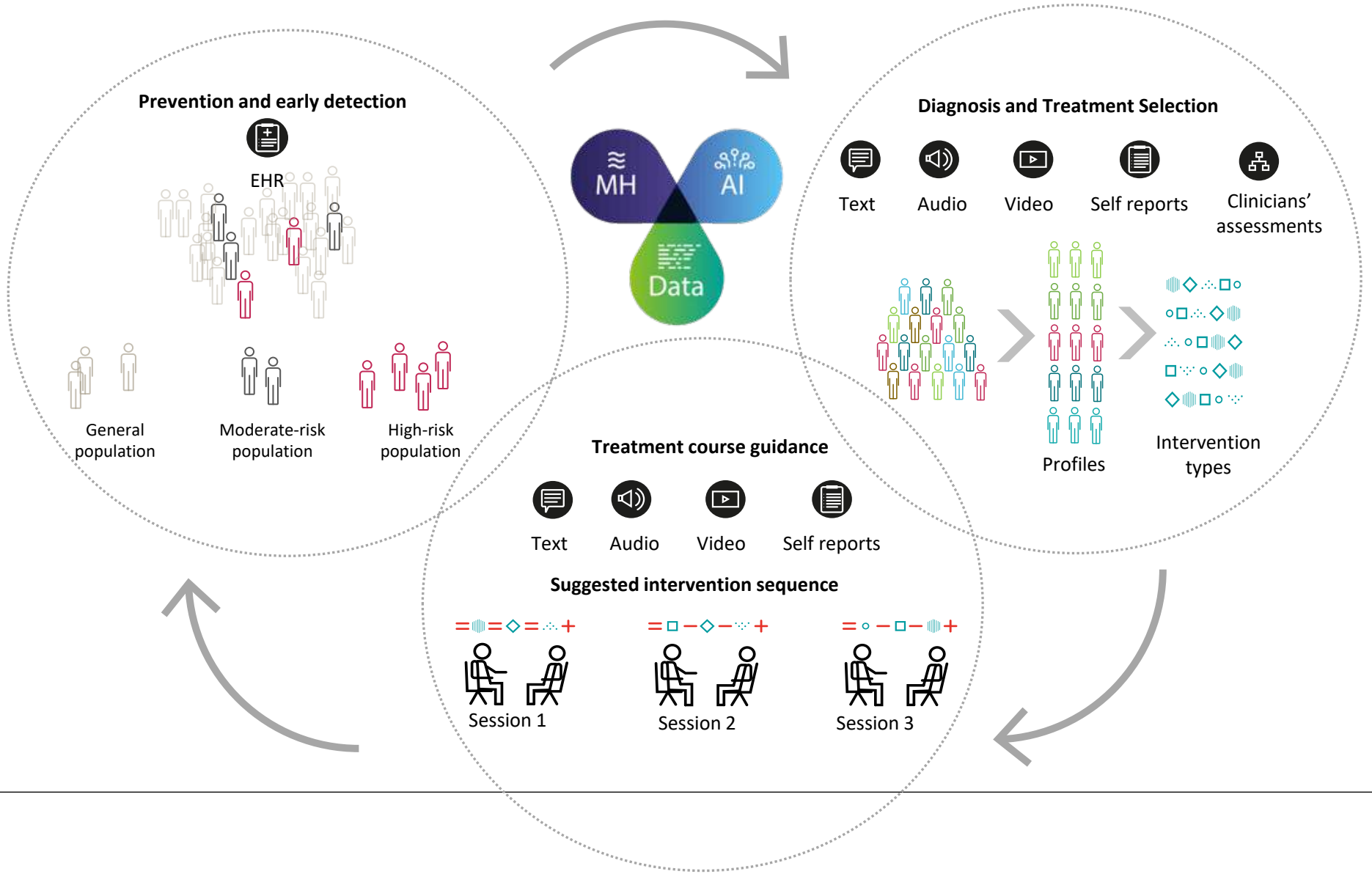
The diverse verbal and non-verbal manifestations of states, their emergence across varied contexts, and the multitude of possible therapists' responses make the expertise of "supershrinks" an elusive intuition that, to date, has been impossible to model.

Most studies use

- **Self-reports at discrete time intervals**
- **Limited number of variables**
- **Limited amount of data**

We need alternatives to the traditional way of doing research that can recognize the complexity of depression, the heterogenicity of clients and the varied options of therapeutic interventions

AI hold the potential to advance mental health in many ways

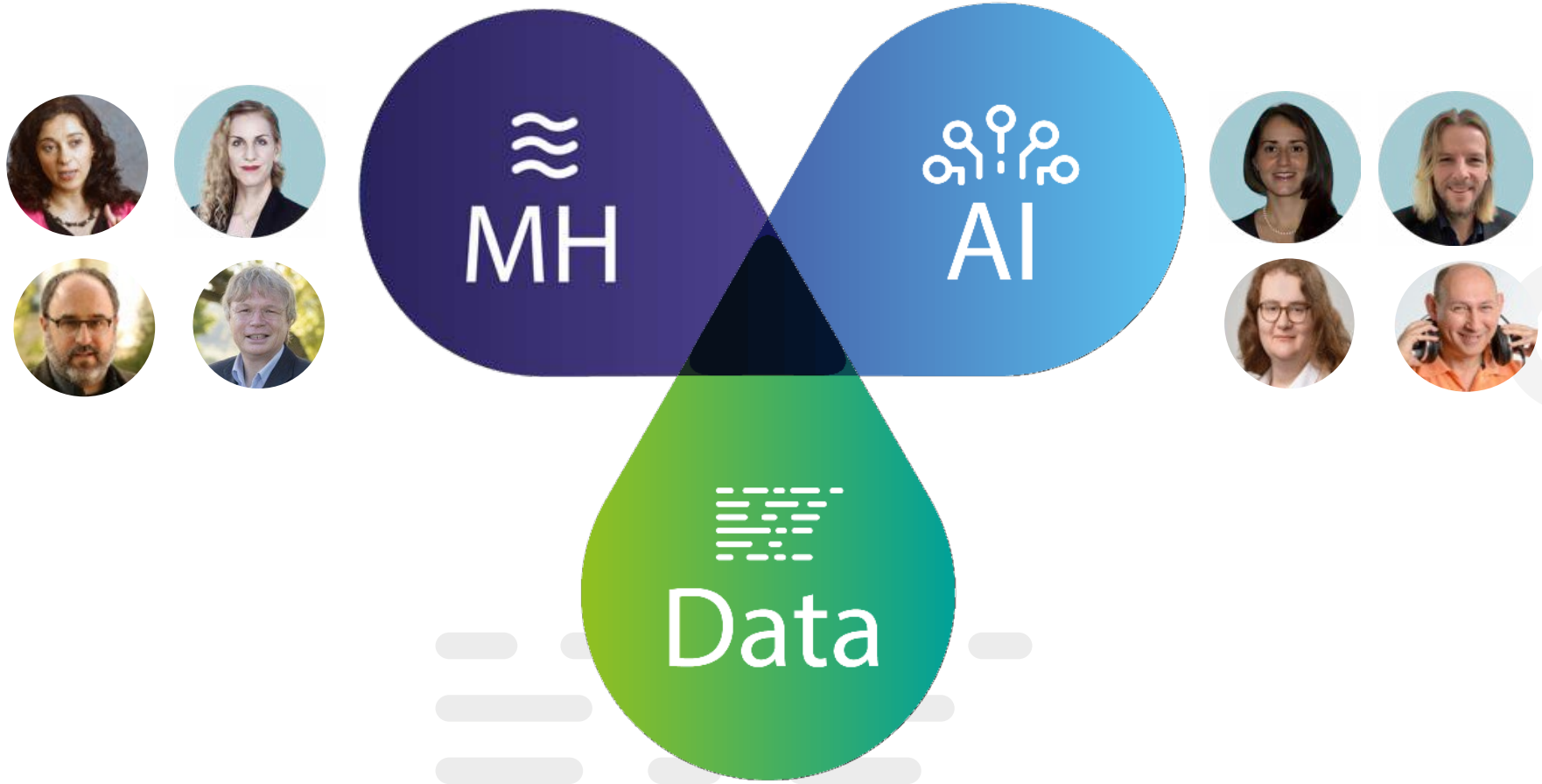


However, there are also concerns and challenges for LLMs in MH

As AI systems continue to advance in their power and capabilities, there are concerns that they may pursue objectives that do not fully align with human needs:

- **Ethical concerns (privacy, safety, responability, bias, explainability)**
- **Lack of standards for evaluation**
- **Struggle to capture temporality**
- **Lack of personalization**
- **Are not multi-modal**

The vision: Integrating novel clinical theory with breakthroughs in AI to transform depression management



Data



- Text
- Audio
- Facial expression
- Body movements
- Clinical judgements
- Subjective self report

Unique datasets

- Large
- Multimodal
- Temporal

Data

Similar datasets

- BIU dataset
- Other clinics' dataset

Other datasets

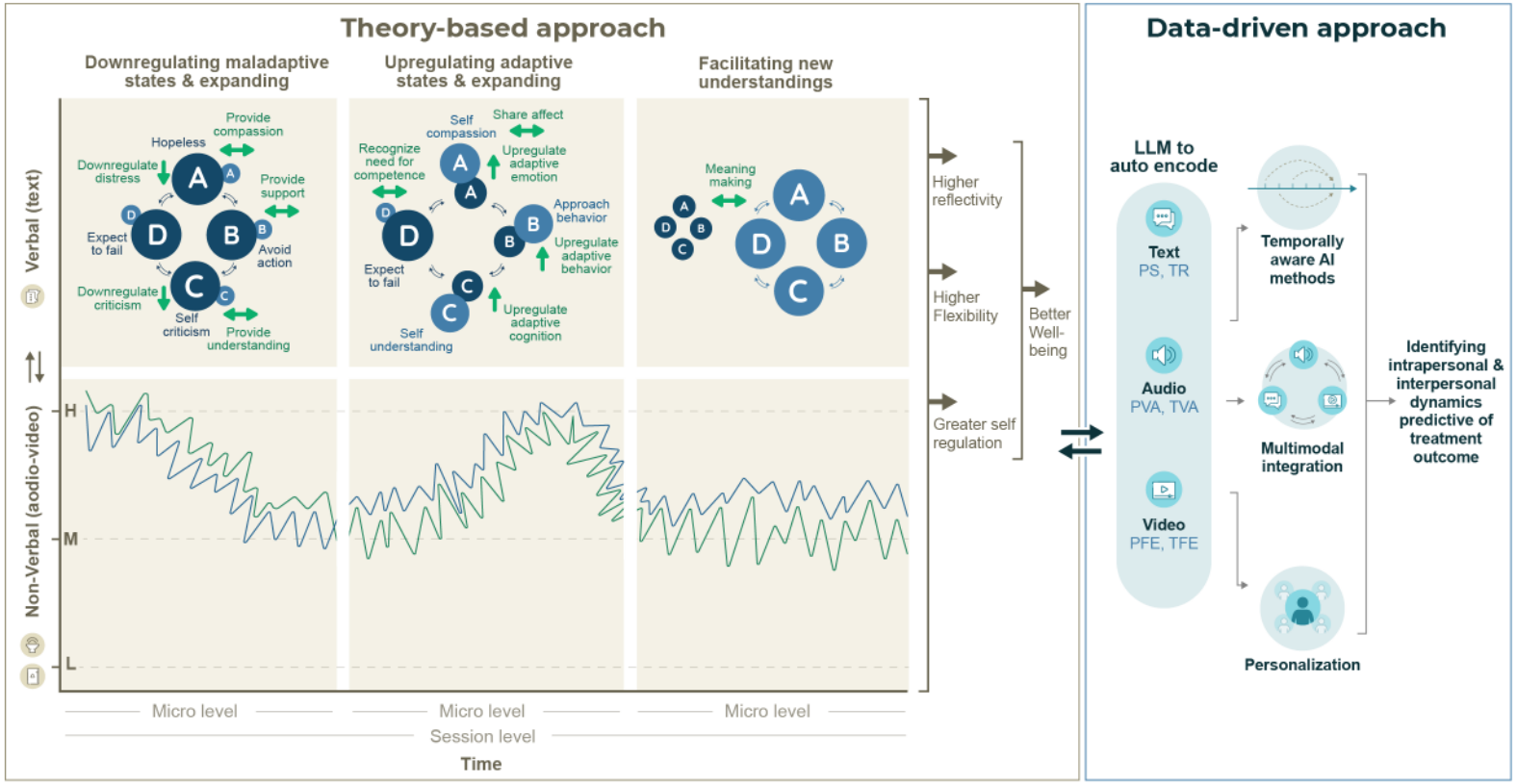
- Electronic Health Records (EHR)

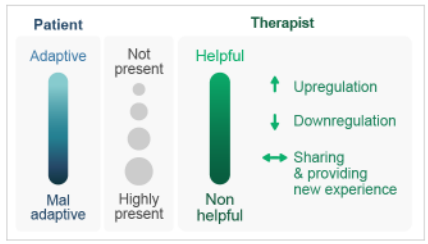
Phase I: Analysis of archival datasets

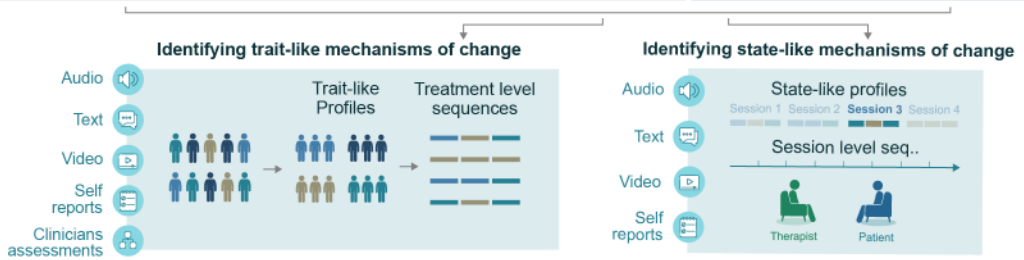
Phase II: Newly collected data from randomized controlled trials

Objective: To advance precision in the treatment of depression

- To develop a trans-theoretical framework that captures intrapersonal (within patient) and interpersonal (between patient and therapist) dynamics in psychotherapy indicative of changes in depression severity over time
- To manually annotate moment-by-moment patient, therapist and dyadic process
- To scale-up the data by automatically annotating a larger dataset
- To improve models' prediction by developing temporally aware, personalized, multimodal data models
- To drill down and explore sequences predictive of micro, session and treatment-level outcomes and continously refine the theoretical framework
- To develop AI-based support systems for patients and therapists







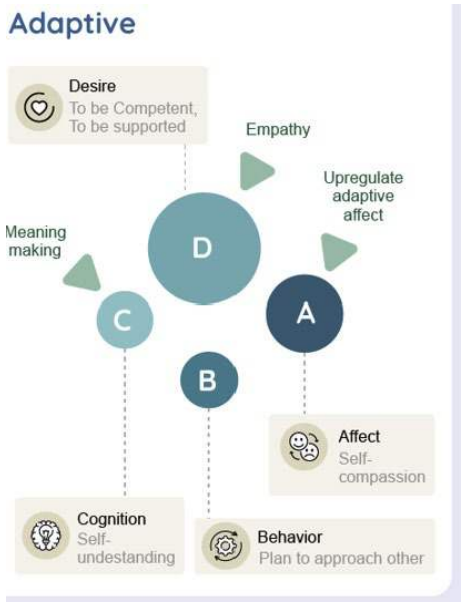
Theory: MIND – Multimodal Intrapersonal and Interpersonal Dynamics

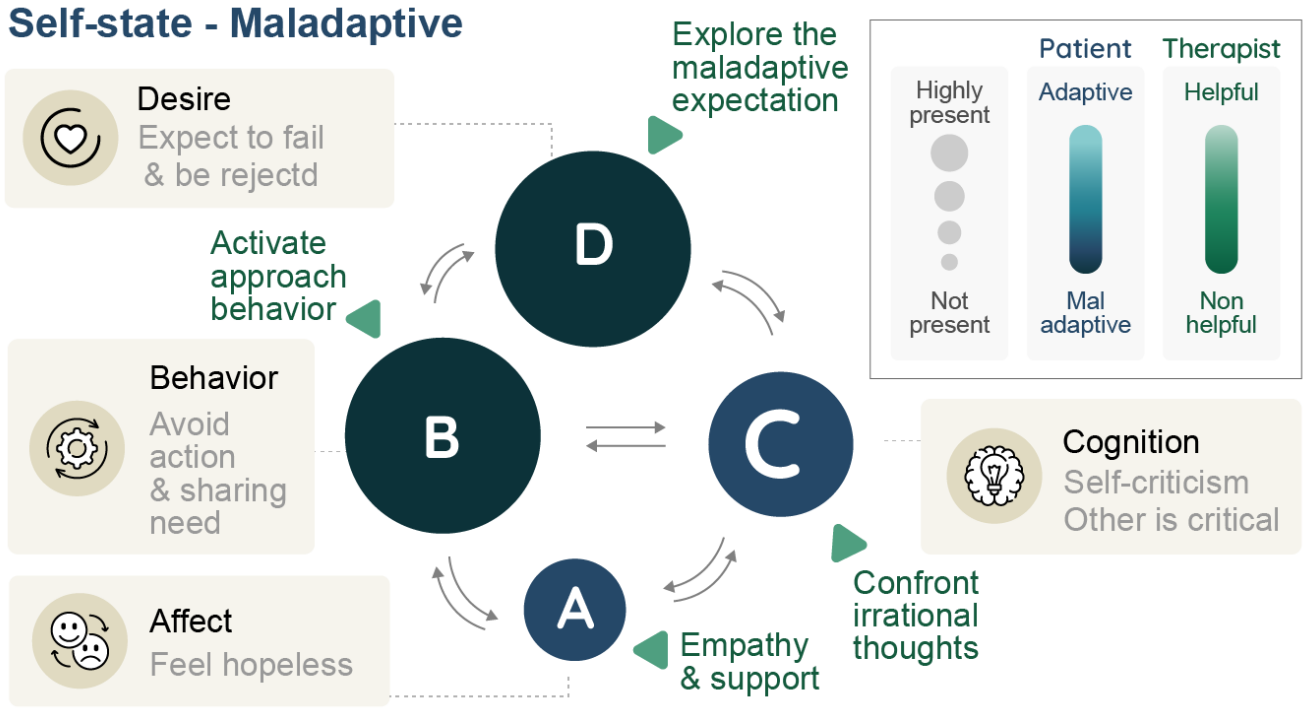


Moves the focus from

- Trait-like psychopathology to **modifiable intrapersonal dynamics**

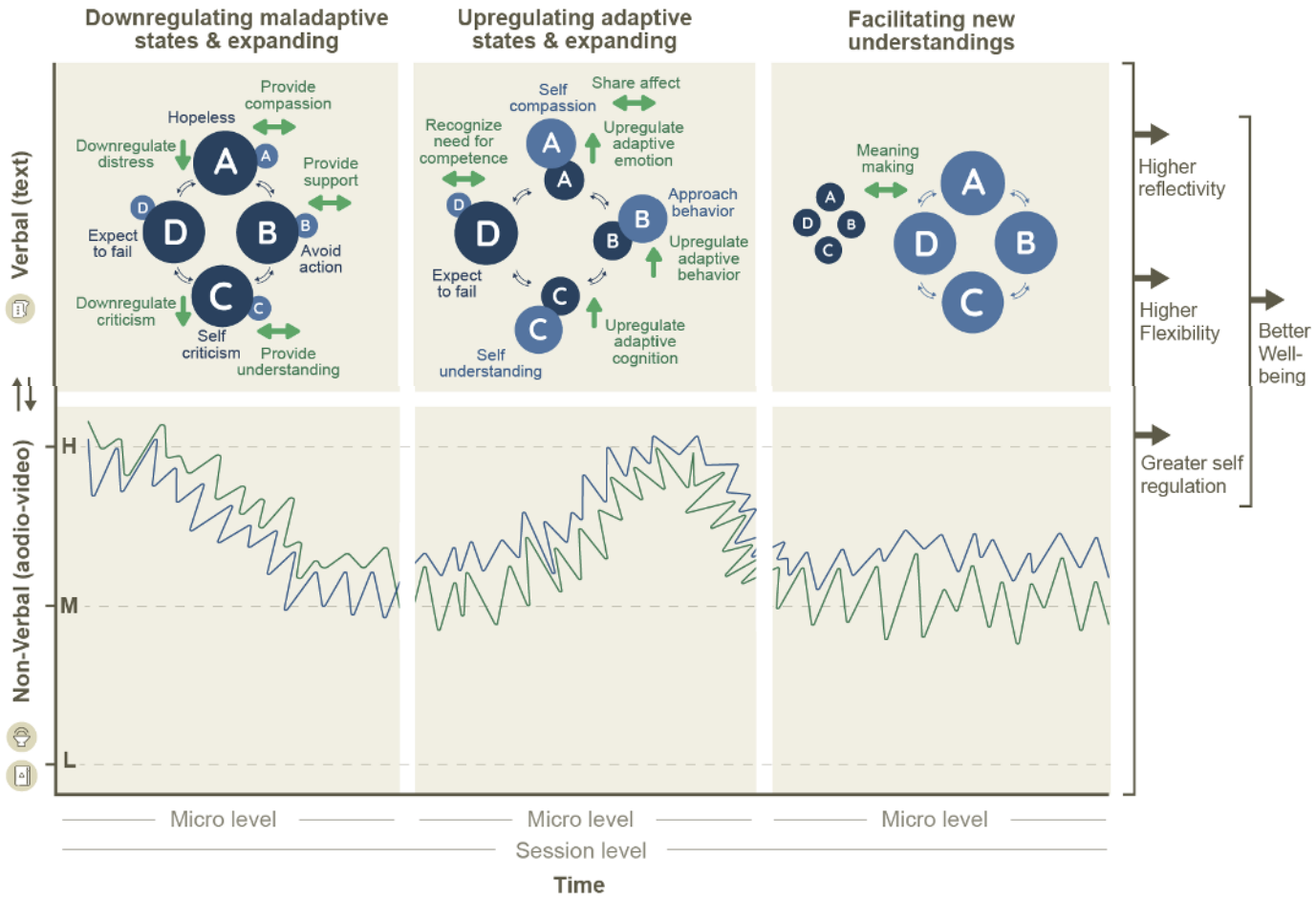
- Manualized treatments to **personalized interpersonal dynamics**



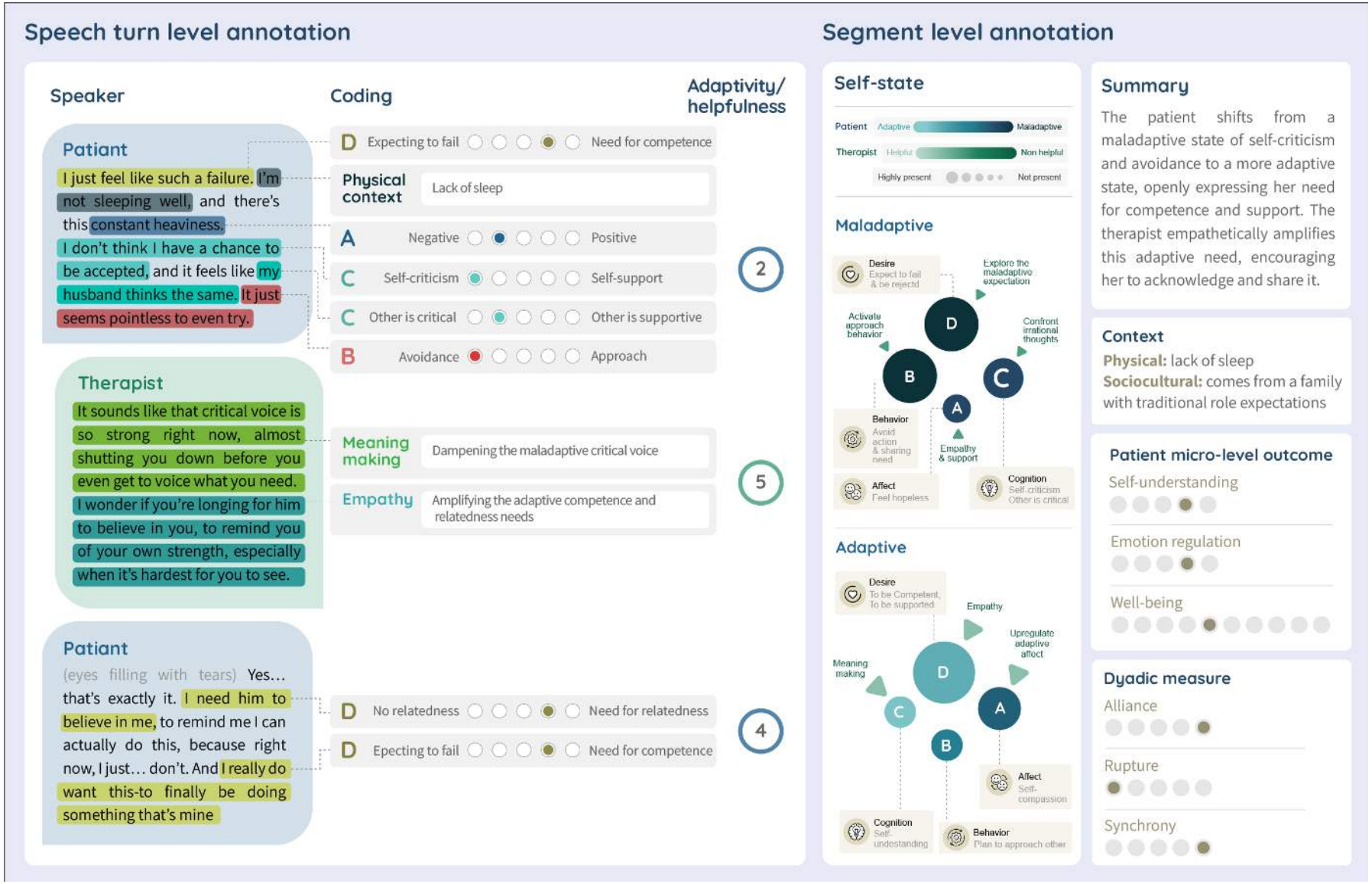


Theory: MIND – Multimodal Intrapersonal and Interpersonal Dynamics





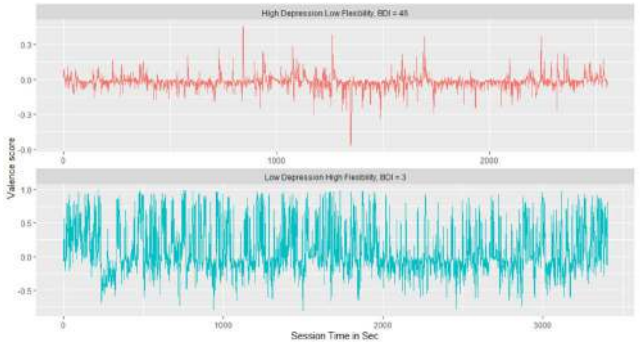
MIND aims to map adaptive and mal-adaptive intrapersonal and interpersonal dynamics



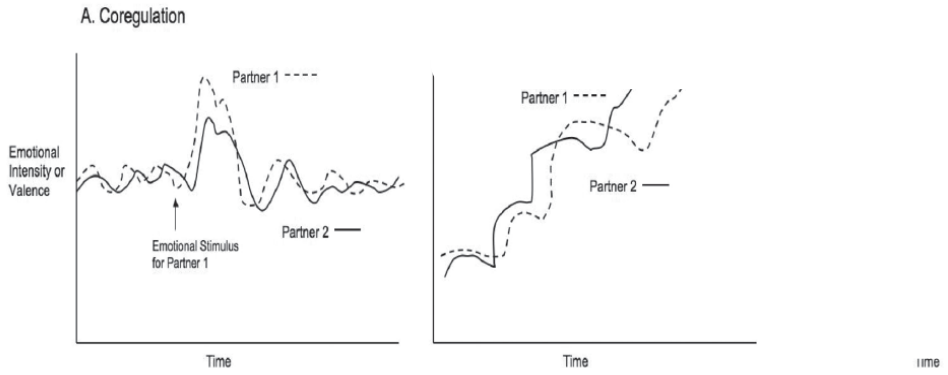
Theory

Now that we have these key building blocks, we can start exploring dynamics

Key **intrapersonal** dynamics - **flexibility and coherence** (Fosha, 2001; Greenberg, 2012)



Key **interpersonal** dynamics: **synchrony and coregulation** (Winnicott, 1971)



Text – a window into the verbal aspect of therapy

Objectives

- To automatically predict clients' emotions and therapists' interventions
- Use the labeled data to examine **Emotional Coherence** between verbally expressed and self-reported emotions
- Examine the association between emotional coherence and clients' well-being



74 dyads

872 transcriptions

Statistics	Total	Per Client	Per Session
Talk Turns	~150K	~2K	~200
Client Tokens	~4 Million	~50K	~4.5K
All Tokens	~5 Million	~70K	~6K

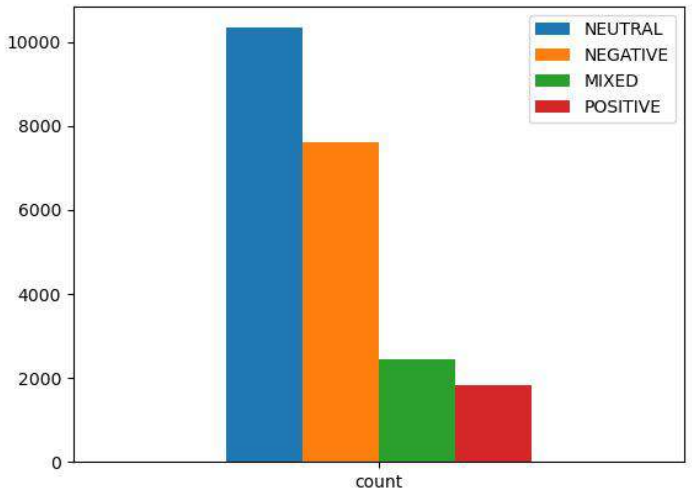
Atzil-Slonim, D., et al., (2024). Leveraging natural language processing to study emotional coherence in psychotherapy. *Psychotherapy.*

Method

A sub-sample of 196 sessions were coded speech-turn by speech-turn

Client emotions

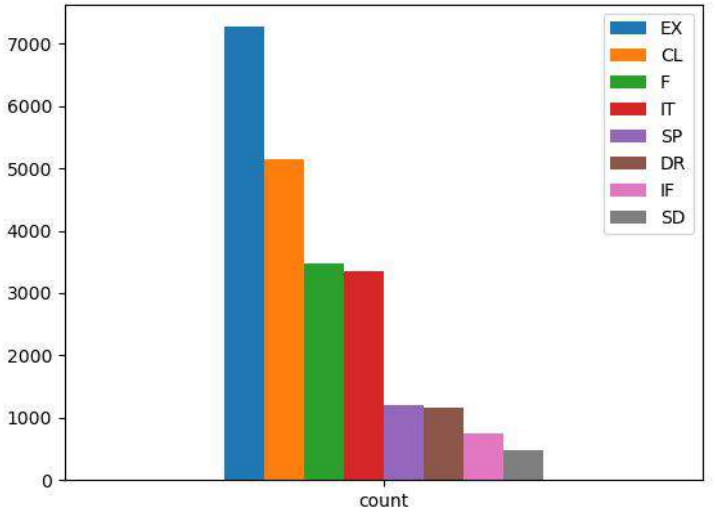
Category
Neutral
Negative
Positive
Mixed



Therapist interventions

Psychotherapy Interactional Coding system (PIC; McCullough,1988).

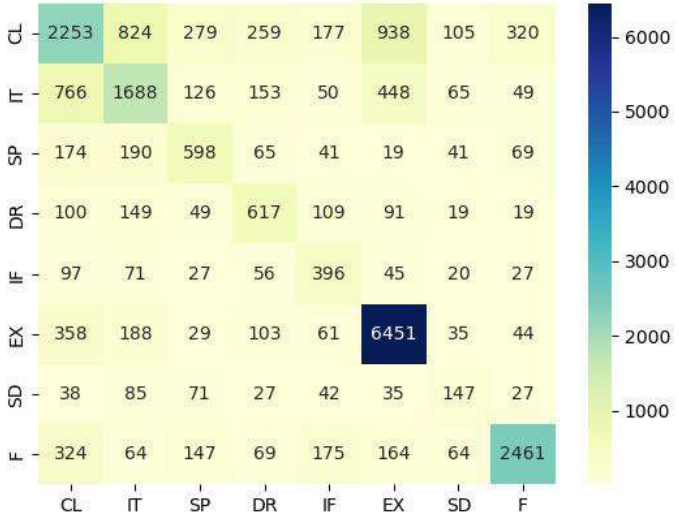
Category	Code
Clarification	CL
Interpretation	IT
Support	SP
Directive	DR
Information	IF
Expansion	EX
Self-disclosure	SD
Filler	F



Results for intervention prediction

Category	CL	IT	SP	DR	IF	EX	SD	F
Human Annotators	0.50	0.56	0.52	0.68	0.40	0.88	0.12	0.82
Language Model	0.49	0.44	0.47	0.47	0.44	0.84	0.32	0.75

Confusion matrix



Model	F1	Kappa
Annotators	0.74	0.65
XLM-ft	0.58	0.47
XLM-Adapter	0.57	0.45
HB-senti-ft	0.61	0.51
HB-Adapter	0. 60	0.50
AB-ft	**0.64**	**0.55**
AB-Adapter	0.63	0.53

Examples of errors

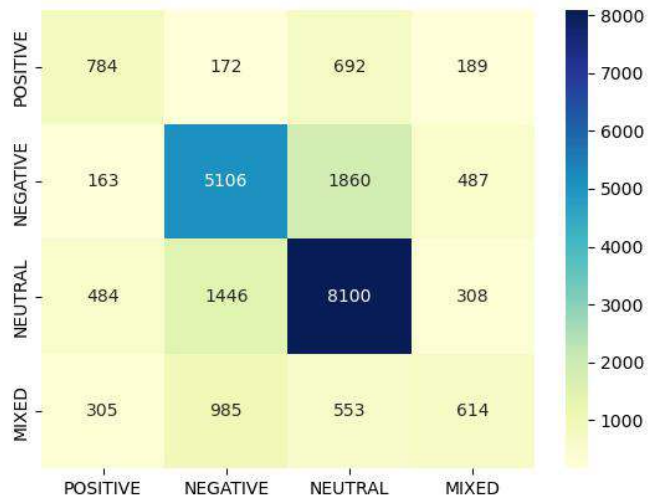
Utterance	Annotation	Prediction
So it felt like you were pushed aside, as if you are a burden?	CL	EX
It bothered you, but I feel that maybe this was not your dominant emotion	IT	SD
It sounds to me like this is your internal critical voice that was speaking now	IT	CL

Results for emotion recognition

Category	POS	NEG	NEU	MIX
Human Annotators	0.55	0.72	0.79	0.32
Language Model	0.44	0.67	0.75	0.30

Model	F1	Kappa
Annotators	0.73	0.54
XLM-ft	0.64	0.42
XLM-Adapter	0.64	0.43
HB-senti-ft	0.61	0.37
HB-Adapter	0.57	0.34
AB-ft	**0.66**	**0.46**
AB-Adapter	0.65	0.44

Confusion matrix



Examples of errors

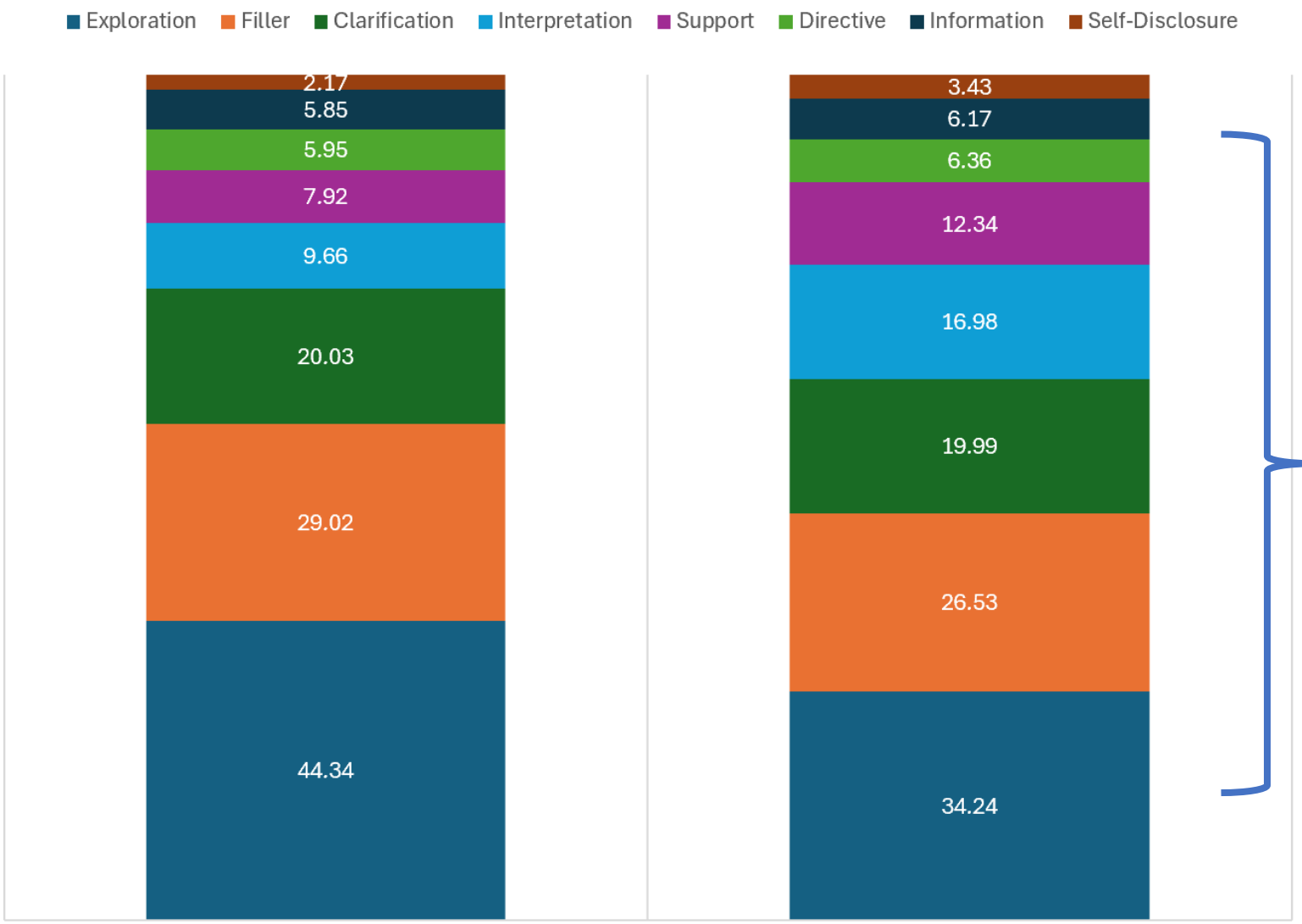
Utterance	Annotation	Prediction
I let go of myself, and say to myself I am allowed to go through difficult times	Positive	negative
I'm dying on my dad, he did it to me with good intentions	Positive	Negative
I did not have a good trip, but in the lab everyone thought that I enjoyed and had fun	Negative	Positive

Results for emotional coherence

- Significant association was found between verbally expressed and self-reported emotions
- Coherence in clients' positive emotions was associated with their well-being at the client-level

	Negative Emotions			Positive Emotions		
	Est. (SE)	CI (95%)	P	Est. (SE)	CI (95%)	P
Intercept (γ_{00})	5.43(0.20)	[5.04,5.81]	<0.001 ***	39.83(0.86)	[38.14,41.52]	<0.001 ***
VEE (γ_{10})	0.01(0.003)	[0.01,0.02]	<0.001 ***	0.09(0.02)	[0.04,0.14]	<0.001 ***

Results – therapists' technique diversity

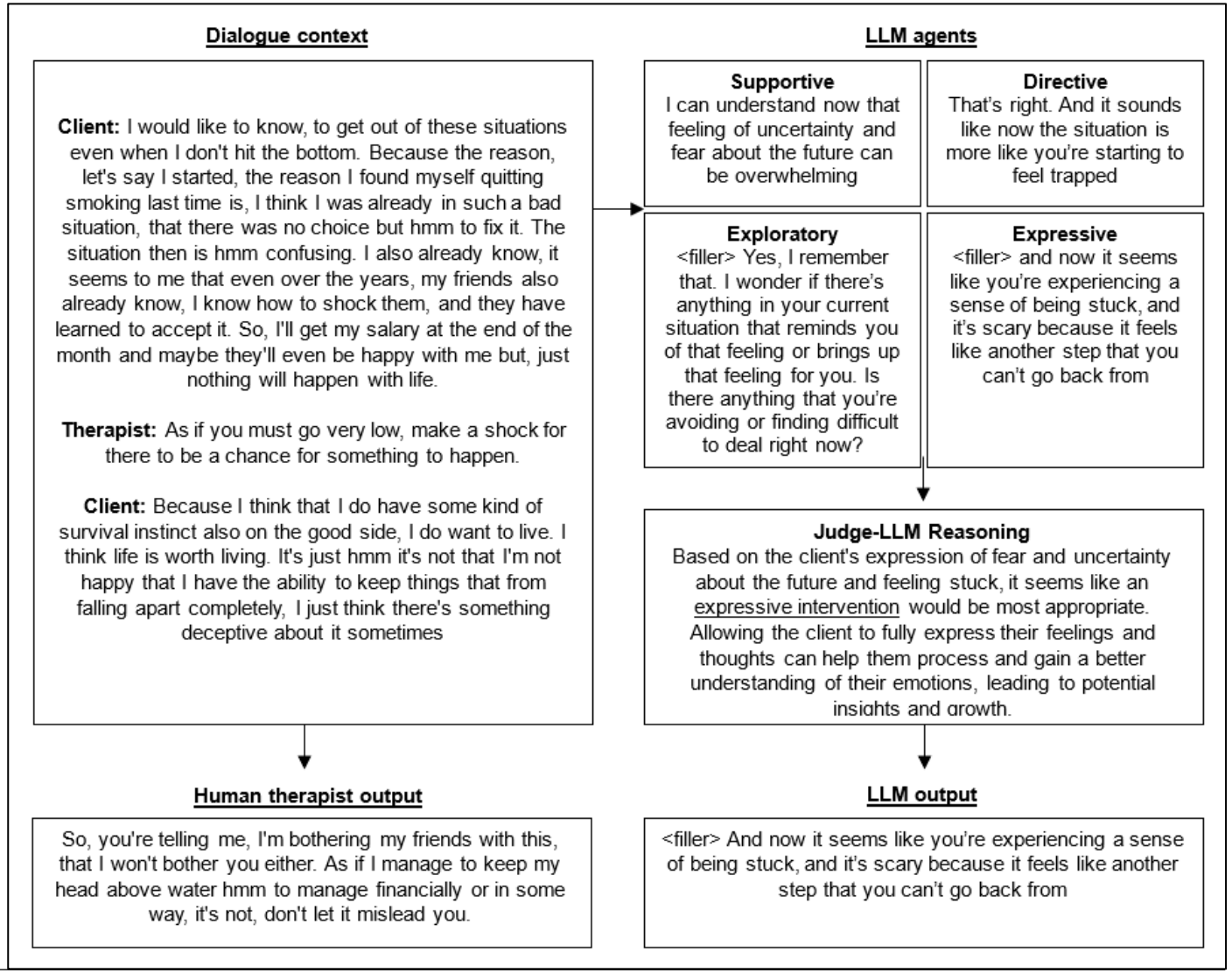


Greater therapist diversity throughout treatment/sessions was associate with stronger alliance - A decrease in the number of exploration questions and filler interventions and an increase in meaning making interventions

Evaluating the capabilities of LLMs versus human therapists in generating personalized interventions

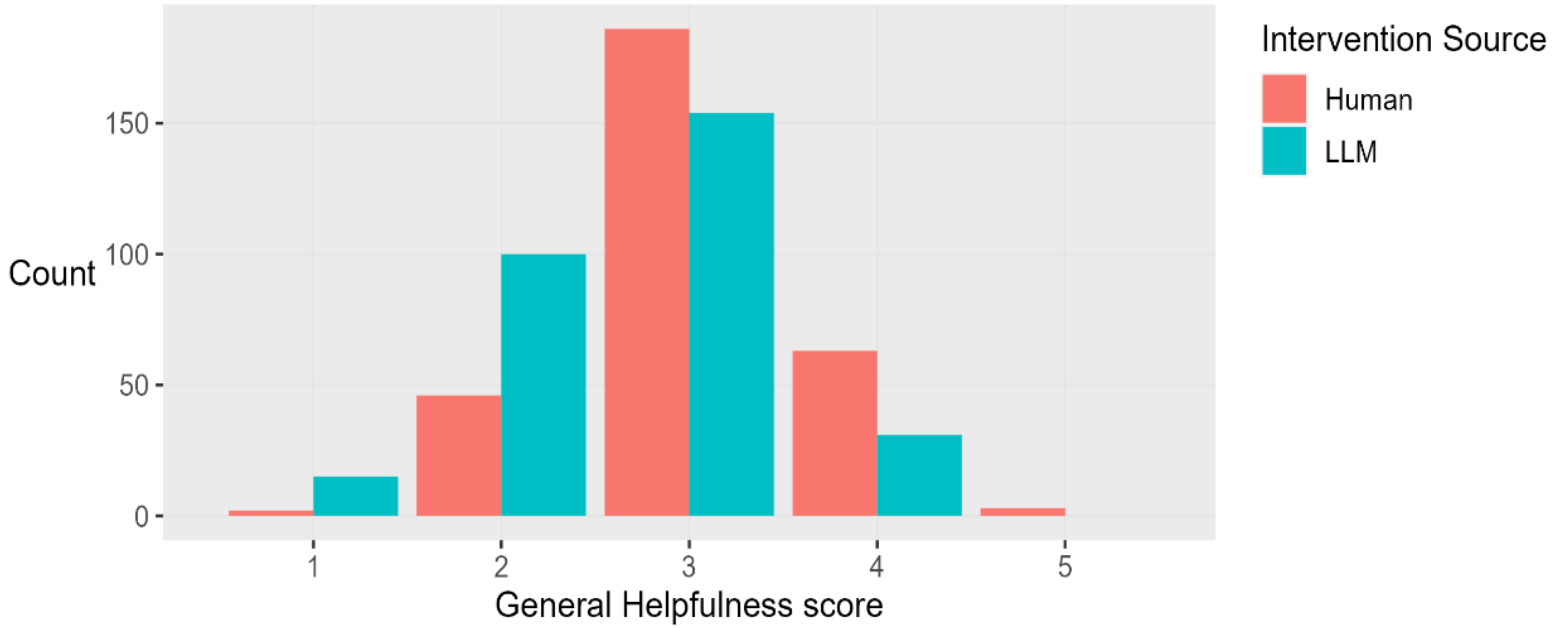
Method:

- Mistral-7B v0.3 was chosen based on clinical judges' qualitative evaluations of outputs of multiple LLMs.
- Hierarchical-LLM framework was implemented (based on Li et al., 2024):
  - 4 LLM agents generated 4 possible interventions, each prompted to focus on a specific intervention type.
  - A Judge-LLM evaluated the four responses based on contextual relevance, fluency, and alignment with best therapeutic practices, and provided the final intervention (while providing reasoning for his choice).
- 3 judges evaluated 600 interventions for general helpfulness



Results – LLMs' vs human therapists' general helpfulness

- Compared to LLM, human therapists were **generally more helpful and** provided **meaning-making interventions of higher quality**.
- **LLMs performed as well as human therapists in supportive, directive, and exploratory interventions**
- Out of 163 cases for which **low scores (1-2)** were given, 115 (70.55%) were from **LLM**.
- Out of 97 cases for which **high scores (4-5)** were given, 66 (68.04%) were from **Human therapists**.



Example A- equally good interventions by both LLM and Human therapist:

The patient is a woman in her early thirties, diagnosed with MDD. At the beginning of the dialogue, the patient shares the emotional pain caused by her partner's family around their wedding. She compares her deep trust in her mother with the distrust in her future mother-in-law, and shares that she chose to limit her interactions with her partner's family. The dialogue proceeds as follows:

Therapist: So something shifted in your relationship with your partner's family, and now there's tension right before the wedding. That's often a time when emotions run high.
Patient: Yeah, I know tensions are normal. But I think people around you are supposed to care about your happiness. That wasn't the case here. There were other motives.
Therapist: You're right to expect them to support you, to share your joy, to make you feel cared for. You felt his mother really didn't give that sense at all.
Patient: No. The wedding will be small, but I know I'll be surrounded by people I love, and that's enough for me.
Therapist: I wonder if there might come a time when you'll feel safe enough to express your feelings without getting hurt or needing to apologize. Maybe not before the wedding, though.
Patient: I honestly don't think it will ever happen.
Therapist: You don't?
Patient: No, I just don't think she's a rational person. If she were, I could talk to her, explain where she went wrong, and expect some kind of apology. But she doesn't apologize.
Therapist: That's really how it sounds.
Patient: To me, it feels like such a shallow relationship. I accepted their background even though it's different from mine- I have shown them respect. They don't seem to notice they're not doing the same. This is the life we chose- me and their son. He didn't walk away from everything, he still respects religion and where he came from. I was raised with values and traditions, too. I think next time I won't take it as hard. I know better now what I'm dealing with. I'll set some boundaries and feel stronger.

Human therapist output:
Yes, I think that throughout this whole period, you've already started figuring out how you want to relate to her, and I believe you'll continue to refine that understanding. It's really challenging.

LLM output:
Yes, you have a lot of wisdom and understanding about yourself and about the people around you. It's important to continue to trust yourself and your feelings and not let others define you or your worth.

Example B- inadequate LLM-generated intervention:

The patient is a man in his late twenties, diagnosed with MDD. At the beginning of the dialogue, the patient describes an important person in his life and the various challenges she is dealing with. He feels that, due to her struggles, she tends to deprioritize him, assuming he will remain available and supportive no matter what. The dialogue proceeds as follows:

Therapist: So, how do you feel about being pushed to the bottom of the list?
Patient: On the one hand, I'm proud and happy that I can be a stable anchor for her. But on the other hand, I'm also human. Most of the time I'm okay, but when I'm not, I really need someone. And sometimes she's there, but in the end, it's this overwhelming feeling of frustration that often leads to an argument. It just feels like I don't have that kind of anchor.
Therapist: You know, I'm thinking about what you told me happened this week. The frustrating situations sound very similar. You care so much. You go see your mom, you do a favor for a relative, helping him talk to his kid about the future. And then what you feel you get back from people is almost disregard. Like, you give a lot and end up feeling they don't appreciate it and are not there for you.
Patient: I totally agree. I often find myself thinking, I wish I cared less. Whether it's about big, meaningful things, or even something small. Maybe if I cared a little less, it would be good for me.
Therapist: It's complicated. Because you really do care, a lot. But also, on some level, you seem to care less about yourself and are willing to accept being pushed aside.
Patient: I guess it's this belief that I can take care of myself, so I don't need the attention from others.
Therapist: And you can take care of yourself. No doubt. The question is what emotional price are you paying for that?
Patient: I guess there is a price.
Therapist: You end up getting less from others. Less support, less space to lean on people.
Patient: Yeah, I think so. Honestly, that's part of why I came here. This constant feeling of carrying too much. Of always having to care, always having to perform, always needing to *do*. … All of that.

Human therapist output:
Have you ever considered whether there's a link between your need to excel and what we just discussed- the tendency to place yourself at a lower priority when it comes to what you ask from others?

LLM output:
You know, it's not just about caring, it's also about your expectations. How about trying to lower them a bit?.

Using LLMs to generate clinically meaningful summaries of social media users' timelines



- 500 user timelines
- 35 posts per timeline (on average)



Example Input Post (Highlights for adaptive and maladaptive evidence).	Summary
Someone who can help motivate me to exercise everyday and avoid binging. I have a heart issue where i have low blood pressure (it does not rise with activity either) and high heart rate. I need to exercise everyday (light exercise like walking) in order to help build strength and energy levels. When I miss a day, it is hard for me to get out of bed the next day, so it's very important for my health. I'm not overweight, but i do binge at night sometimes and would like someone who could be a support to help me avoid that. When I eat a lot of carbs/calories at a time, it makes my heart worse it feels like. I get more dizzy and a higher heart rate. So, I need to avoid that but sometimes my apathy with depression overcome me and I'm like "fuck it" and do it anyways.	The dominant self state in this post is adaptive and is centered around the writer's desire for a codependent relationship with someone else who can help her achieve her physical fitness goals. This desire is being fueled by the cognition of others as relating. She is feeling hopeful about getting better with the help of others, and behaves in ways that help promote her self-care by planning a daily exercise routine. All of the above leads her to share her needs with others, providing evidence of a relational behavioral pattern. The other self-state is maladaptive and is characterized by the expectation that her situation will not improve, and things will not get better. This expectation nurtures a feeling of depression within the writer, which in turn allows for self-harming behaviors such as binge eating, because she perceives the situation as if she has nothing to lose.

Song, J., Chim, J., Tsakalidis, A., Ive, J., Atzil-Slonim, D., & Liakata, M. (2024). Clinically meaningful timeline summarisation in social media for mental health monitoring. In proceeding of the conference of the Association for Computational Linguistics (ACL 2024).

Two steps modeling approach

1. **Evidence summary**

Highlight evidence (key phrases) and integrate into evidence summary.
Using three methods:
(a) LLaMa (baseline)
(b) Skeleton (baseline)
(c) Timeline Hierarchical VAE (TH-VAE- novel)

Example Input Post

(Highlights for adaptive and maladaptive evidence).

Someone who can help motivate me to exercise everyday and avoid binging. I have a heart issue where i have low blood pressure (it does not rise with activity either) and high heart rate. I need to exercise everyday (light exercise like walking) in order to help build strength and energy levels. When I miss a day, it is hard for me to get out of bed the next day, so it's very important for my health. I'm not overweight, but I do binge at night sometimes and would like someone who could be a support to help me avoid that. When I eat a lot of carbs/calories at a time, it makes my heart worse it feels like. I get more dizzy and a higher heart rate. So, I need to avoid that but sometimes my apathy with depression overcome me and I'm like "fuck it" and do it anyways.

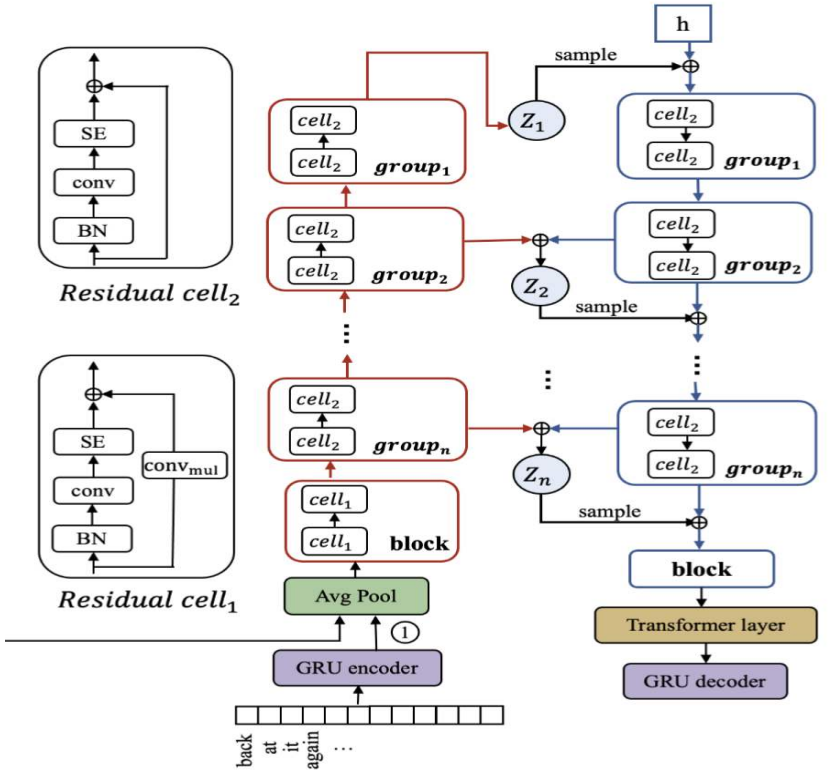
Timeline Hierarchical Variational Autoencoders (TH-VAE)

Attention mechanism – each word gets weight based on how similar it is to the key phrases

The encoding process is hierarchical – starting from the big picture, such as the structure of the text. Then, considering more specific information such as the timeline's topic

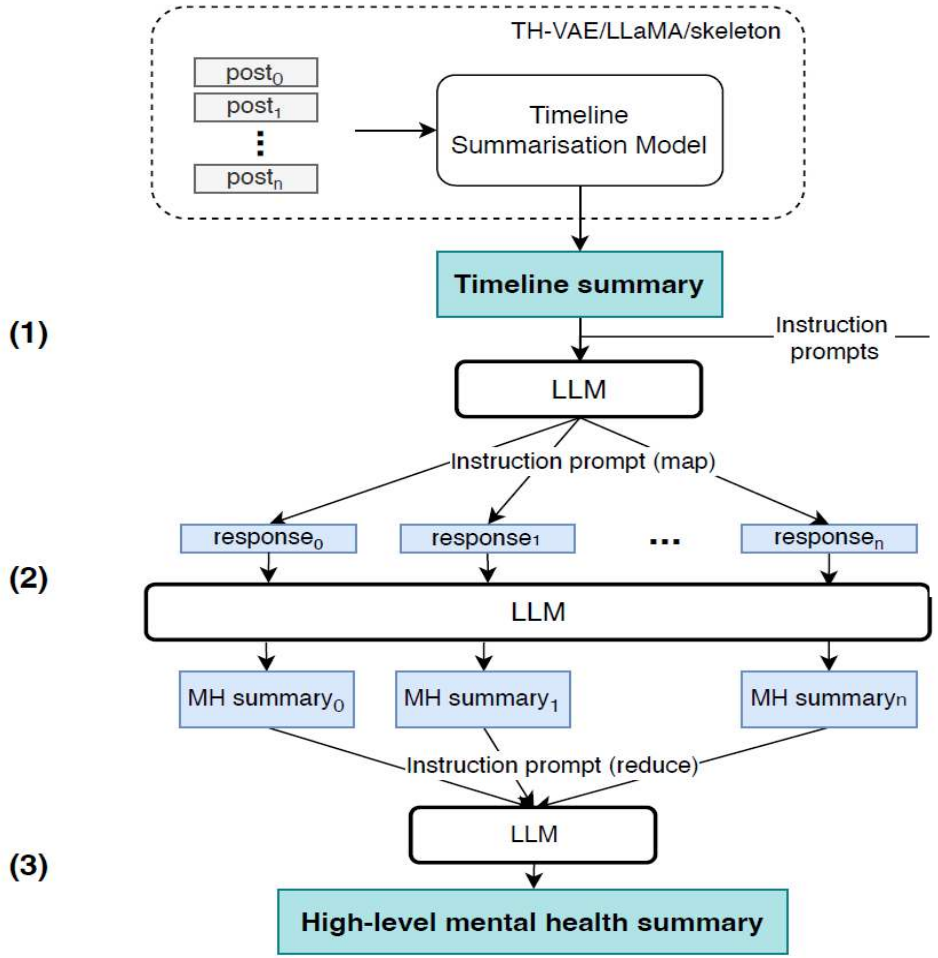
The encoder converts the words and their weight into a vector

The decoder takes this vector and produce sentences that are coherent and relevant to the original text.



2. **High level summary**

(a) The input is the evidence summary
(b) Clinicians crafted a detailed propmt based on the MIND



Summarize self-states for each timeline. Specifically, generate a summary focusing on the interplay between adaptive and maladaptive self-states along the timeline. Emphasize temporal dynamics focusing on concepts such as flexibility, rigidity, improvement, and deterioration. If applicable, describe the extent to which the dominance of the self-states changes over time and how changes in aspects (Affect, Behavior, Cognition, and Desire) contribute to these transitions

Evaluation

Comparing clinicians' generated summaries with models generated summaries (blind)

3 evaluation criteria:
1. Factual consistency
2. Salient meaning preservation
3. Personalization
4. Clinical usefulness

This adolescent complains about feeling depressed. She has low self-esteem, feeling isolated, lonely, and frustrated due to past stressful experiences. Her emotional state includes sadness, frustration, and anger.
She desires attention, connection, acceptance, and validation but tend to held back by fear of rejection and abandonment, impacting her ability to form healthy relationships. Self-criticism and focus on perceived flaws indicate a lack of self-compassion and acceptance.



The individual likely experiences depression, anxiety, and low self-esteem, feeling isolated, lonely, and frustrated due to a history of stressful life events. Their mood reflects sadness, frustration, and anger.

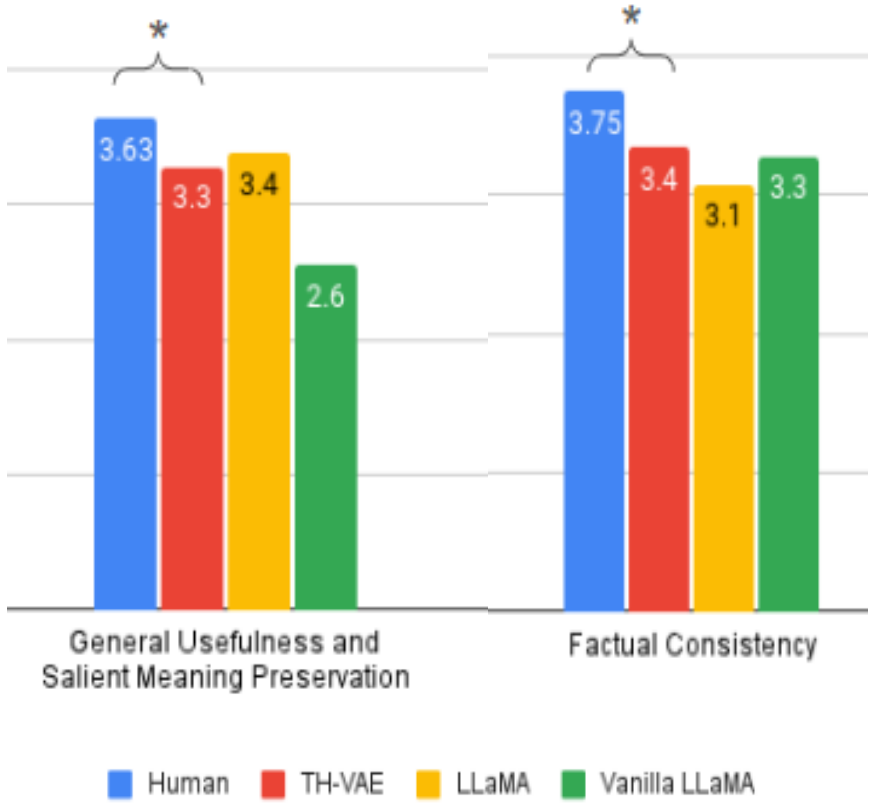
They yearn for attention, connection, acceptance, and validation but fear rejection and abandonment, hindering healthy relationships. Self-criticism and focus on perceived shortcomings reveal a lack of self-compassion and acceptance.

Results

- Humans were more precise than all models
- TH-VAE is more precise than LLaMA & skeleton
- Human summaries were more meaning preserving and useful
- Qualitative analysis reviled that human were more personalized,

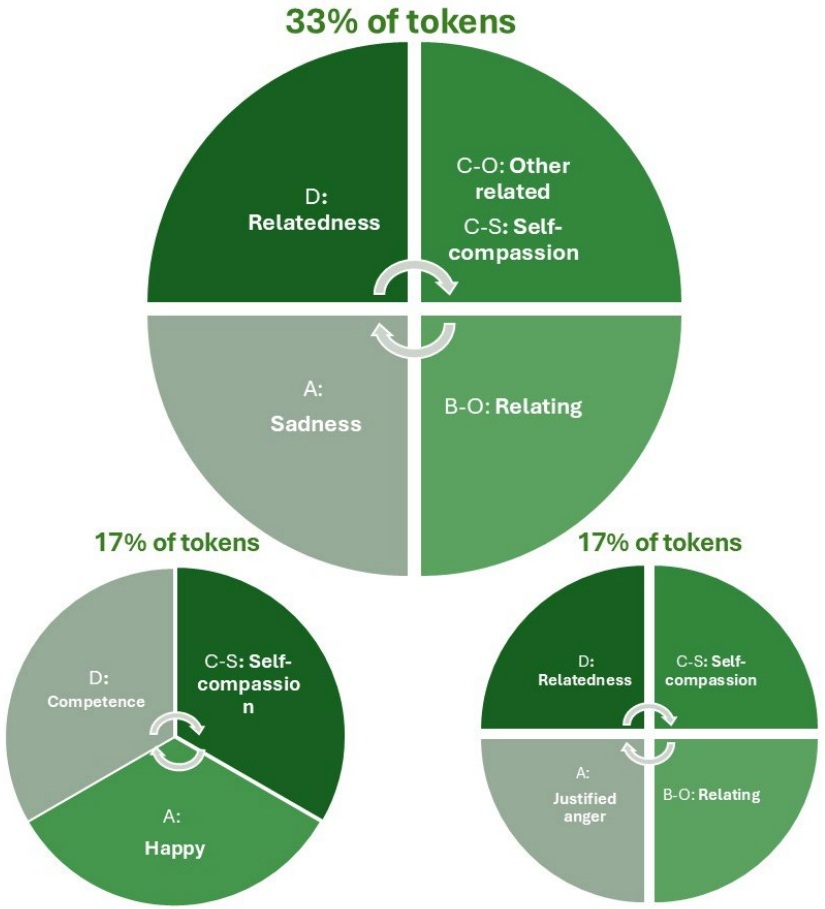
understood causality and provided more in-depth summary.

But the models provided more complete description with no shortcuts.

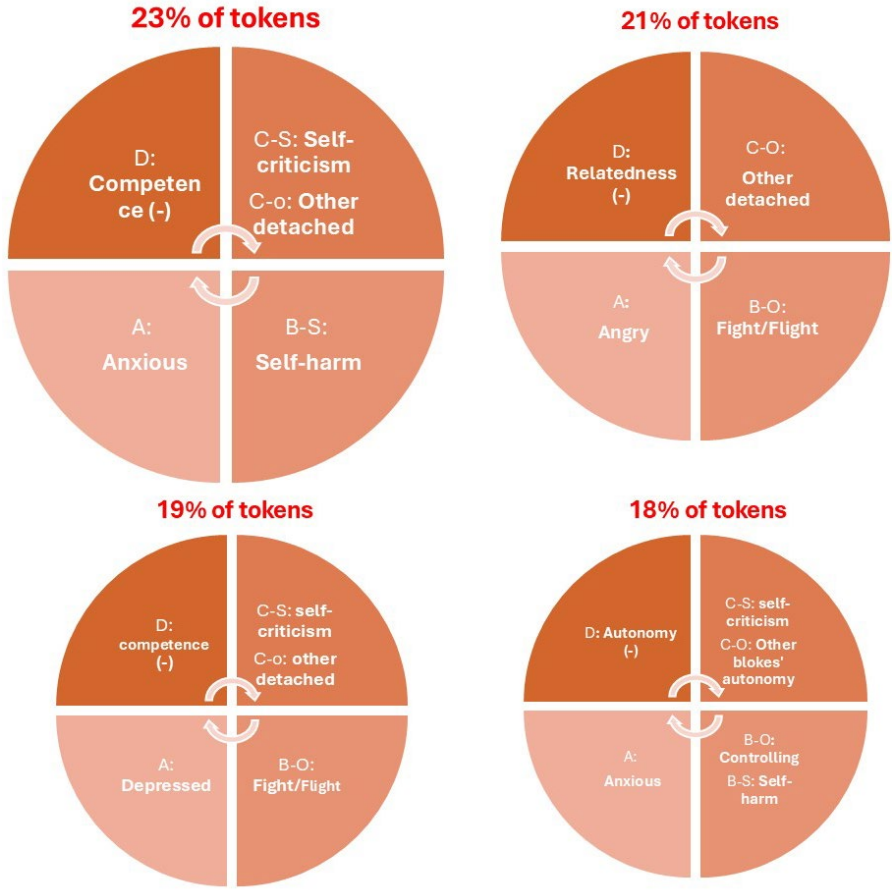


Identifying prototypical self-states among depressed patients

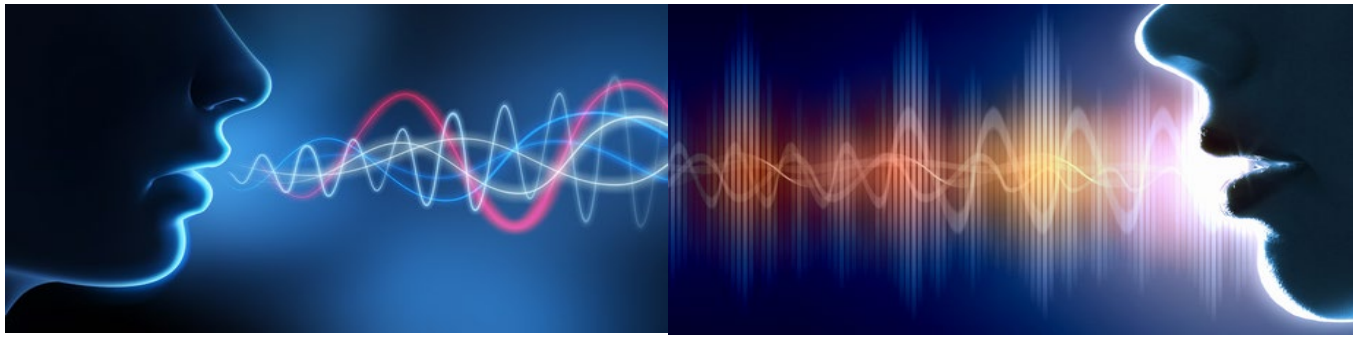
Adaptive Self-States

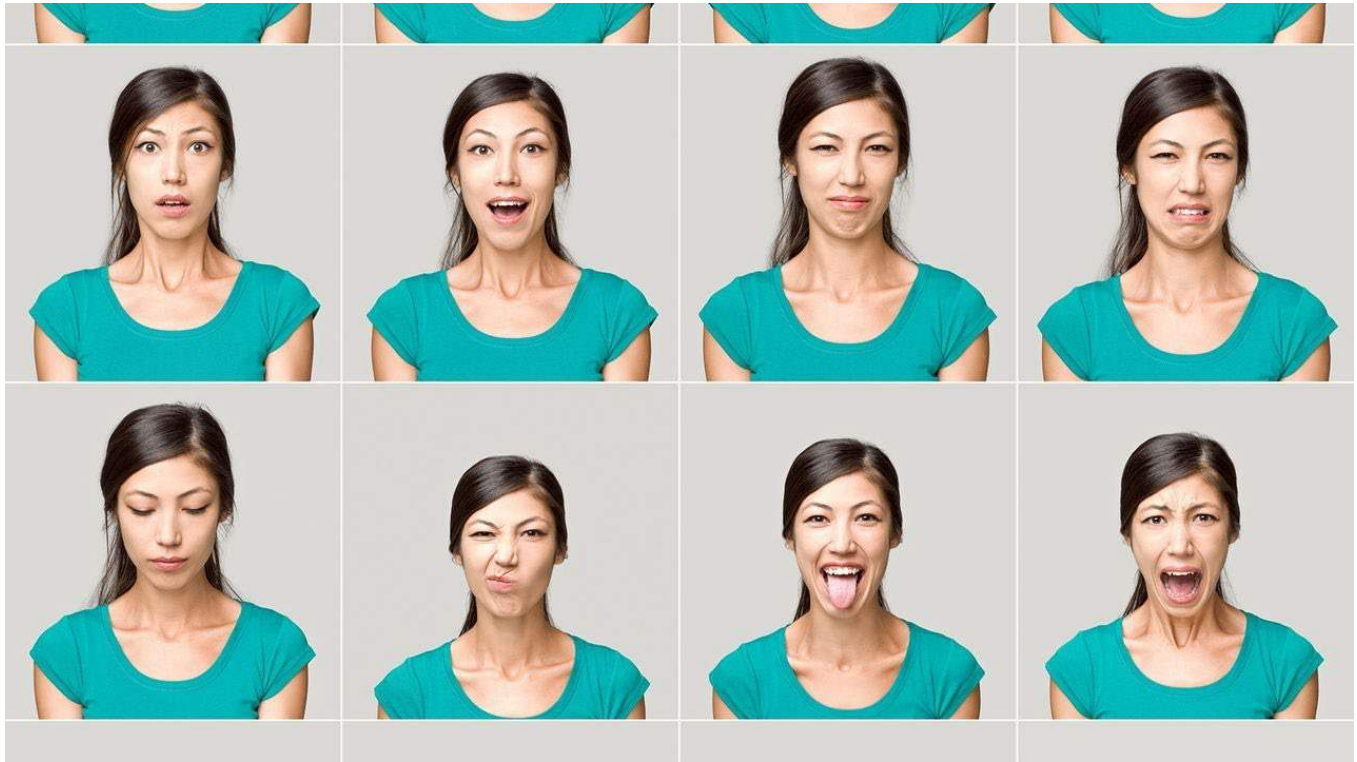


Maladaptive Self-States



Voice/ facial expression–
a window into the non-verbal aspect of therapy





Emotional flexibility in the treatment of depression

Objectives

- To examine whether **Emotional Flexibility** is associated with depression levels at the session and treatment levels

Method

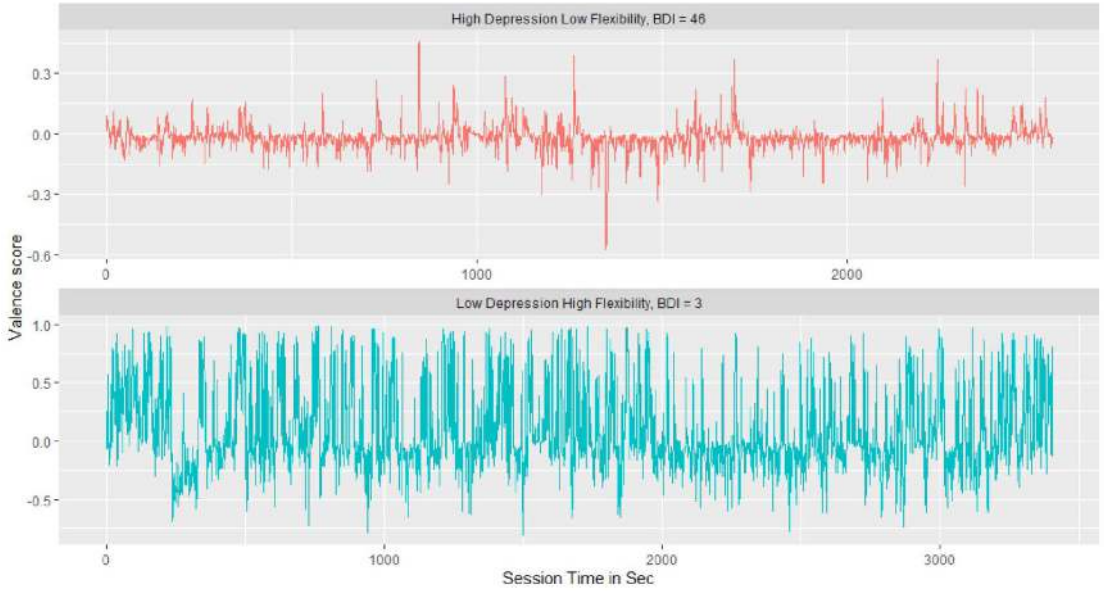
- 58 patients
- 283 sessions

- Emotional flexibility was measured using automated facial expression emotion recognition system (indexed as SD)



Results

- Higher emotional flexibility was associated with lower depressive symptoms at the session and treatment levels
- Similar results in other channels (voice, physiology, self reports)



Atzil-Slonim, D. A., Yehezkel, I., Paz, A., Bar-Kalifa, E., Wolff, M., Dar, A., & Gilboa-Schechtman, E. (2024). Facing change: using automated facial expression analysis to examine emotional flexibility in the treatment of depression. *Administration and Policy in Mental Health and Mental Health Services Research*, *51*(4), 501-508.

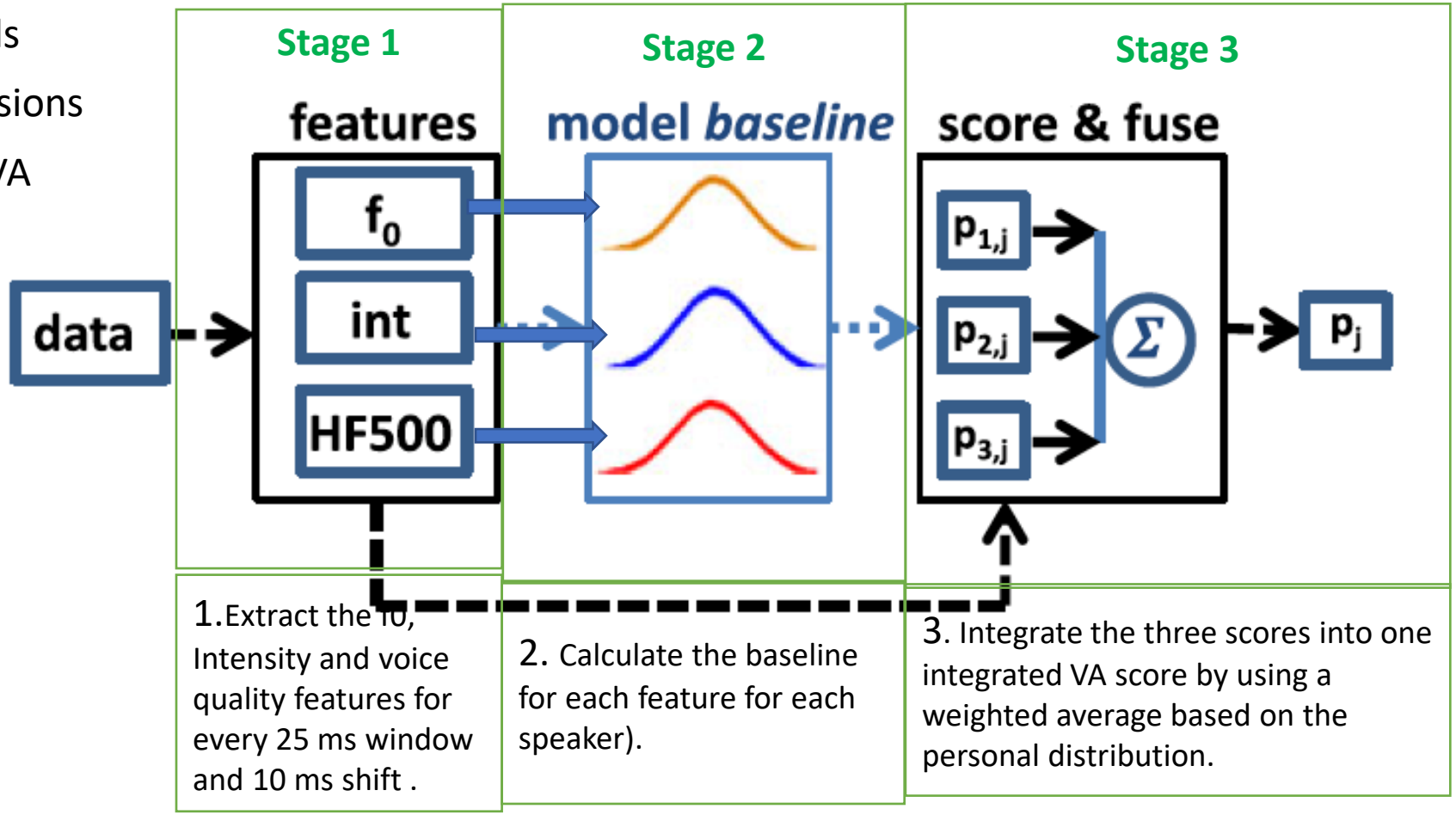
Using Vocal arousal to study co-regulation

Objectives

- To examine the ways in which **coregulation** facilitates treatment outcome

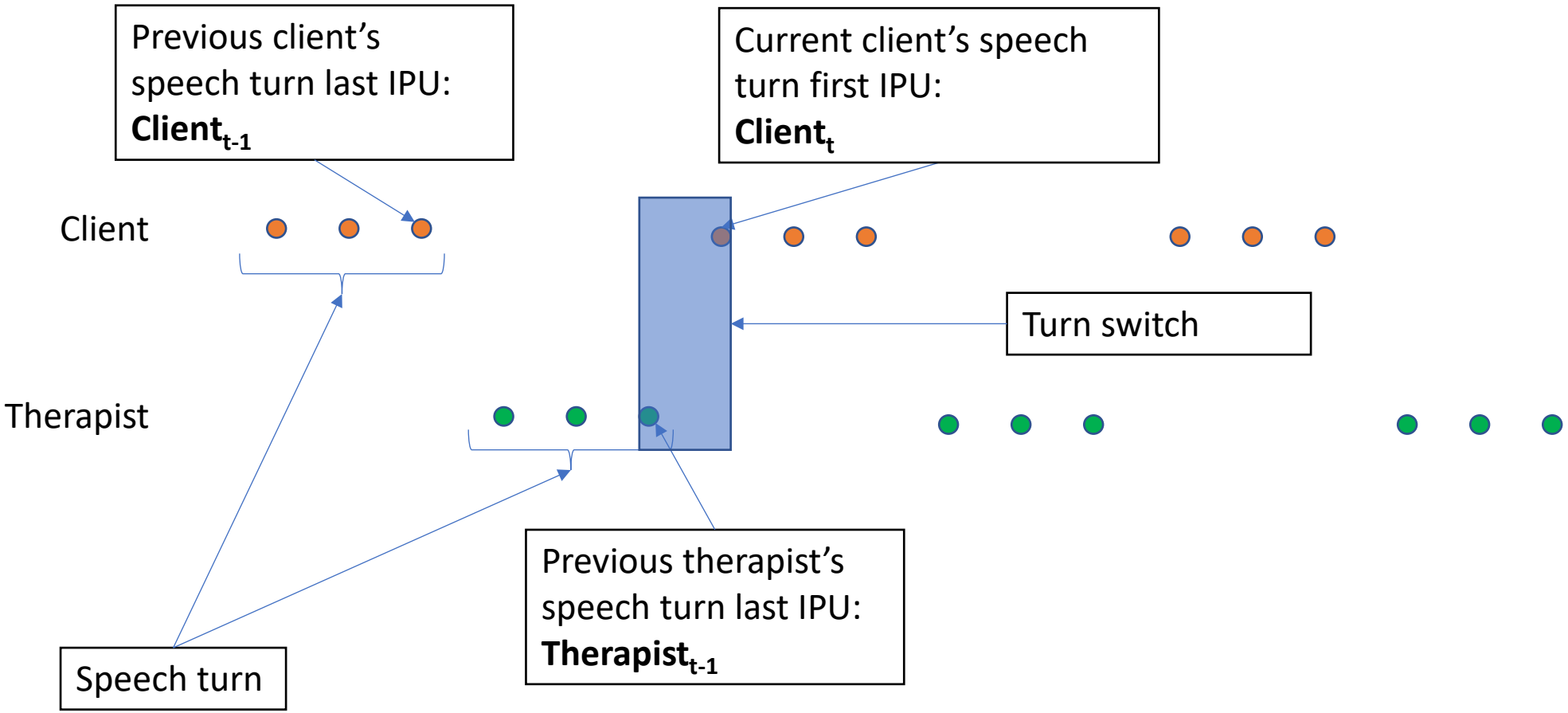
Method

- 30 dyads
- 279 sessions
- 21133 VA



Paz, A., Rafaeli, E., Bar-Kalifa, E., Gilboa-Schectman, E., Gannot, S., Laufer-Goldshtein, B., ... & Atzil-Slonim, D. (2021). Intrapersonal and interpersonal vocal affect dynamics during psychotherapy. Journal of Consulting and Clinical Psychology, 89(3)

- Focusing on speech turn switches.
- Inter-Pausal Unit (IPU)-The continuous speech period by a single speaker between speech breaks (Hirschberg, 2011)



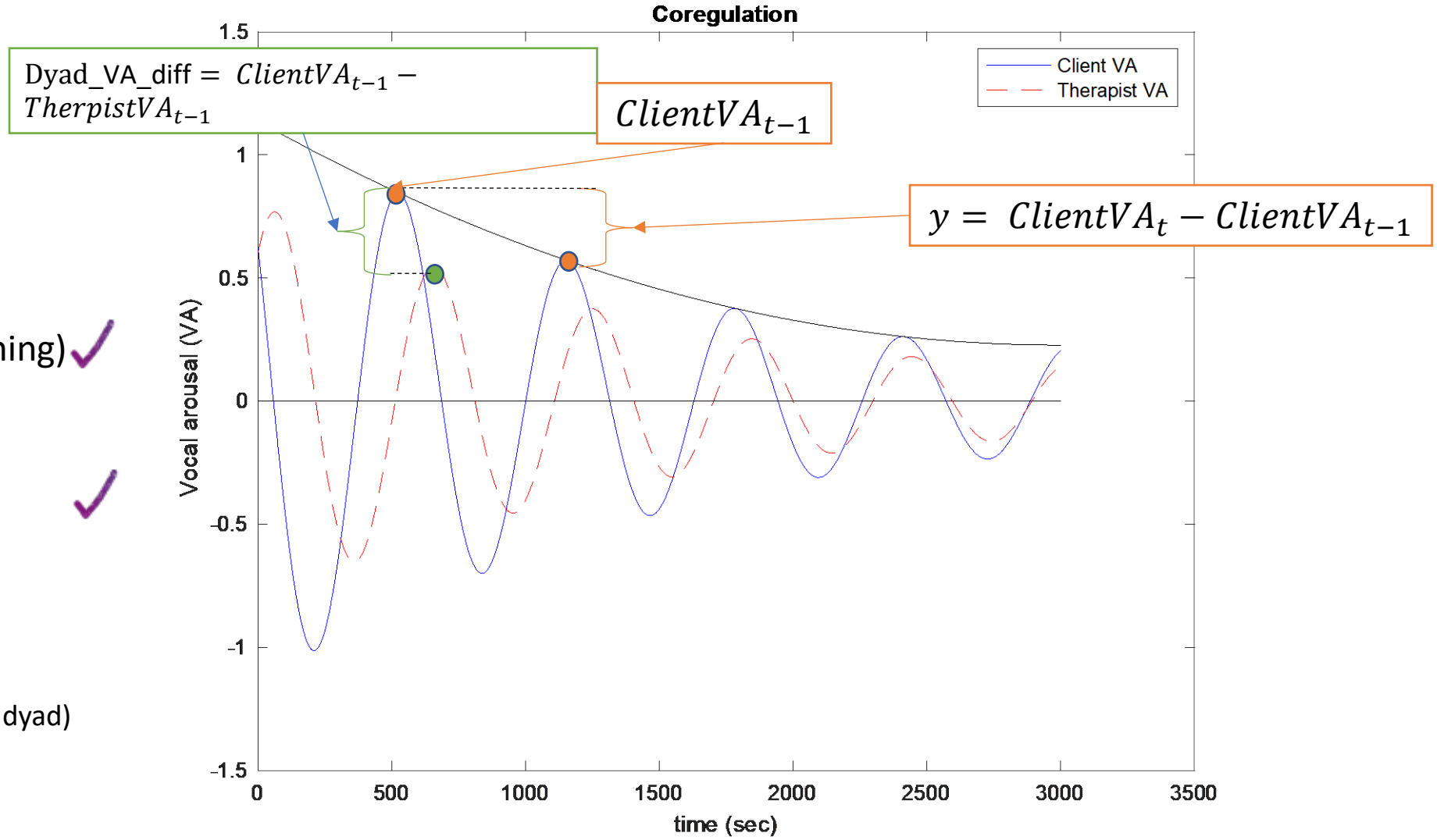
Results

Evidence for the two forces:

- Self-regulation (Dampening) ✓
- Co-regulation (Dampening) ✓

Three level MLM

(IPU nested in session, nested in dyad)



Client VA change

$$ClientVA_{(t)sd} - ClientVA_{(t-1)sd} = bsd_0 + b_{sd1} * ClientVA_{(t-1)sd} + b_{sd2} * (ClientVA_{(t-1)sd} - TherpistVA_{(t-1)sd}) + \text{err}$$

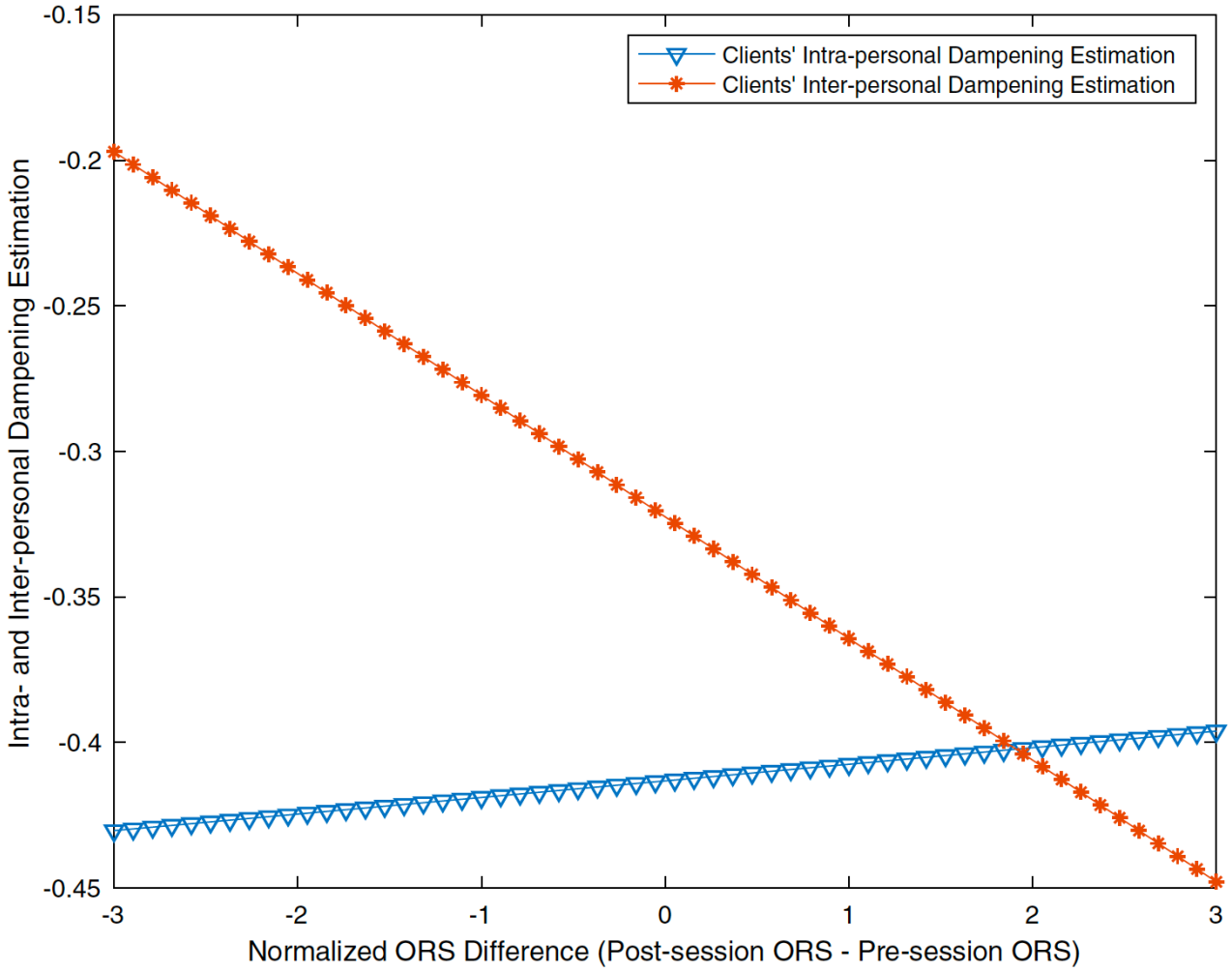
Dyad_VA_diff = VA difference between client and therapist

Results

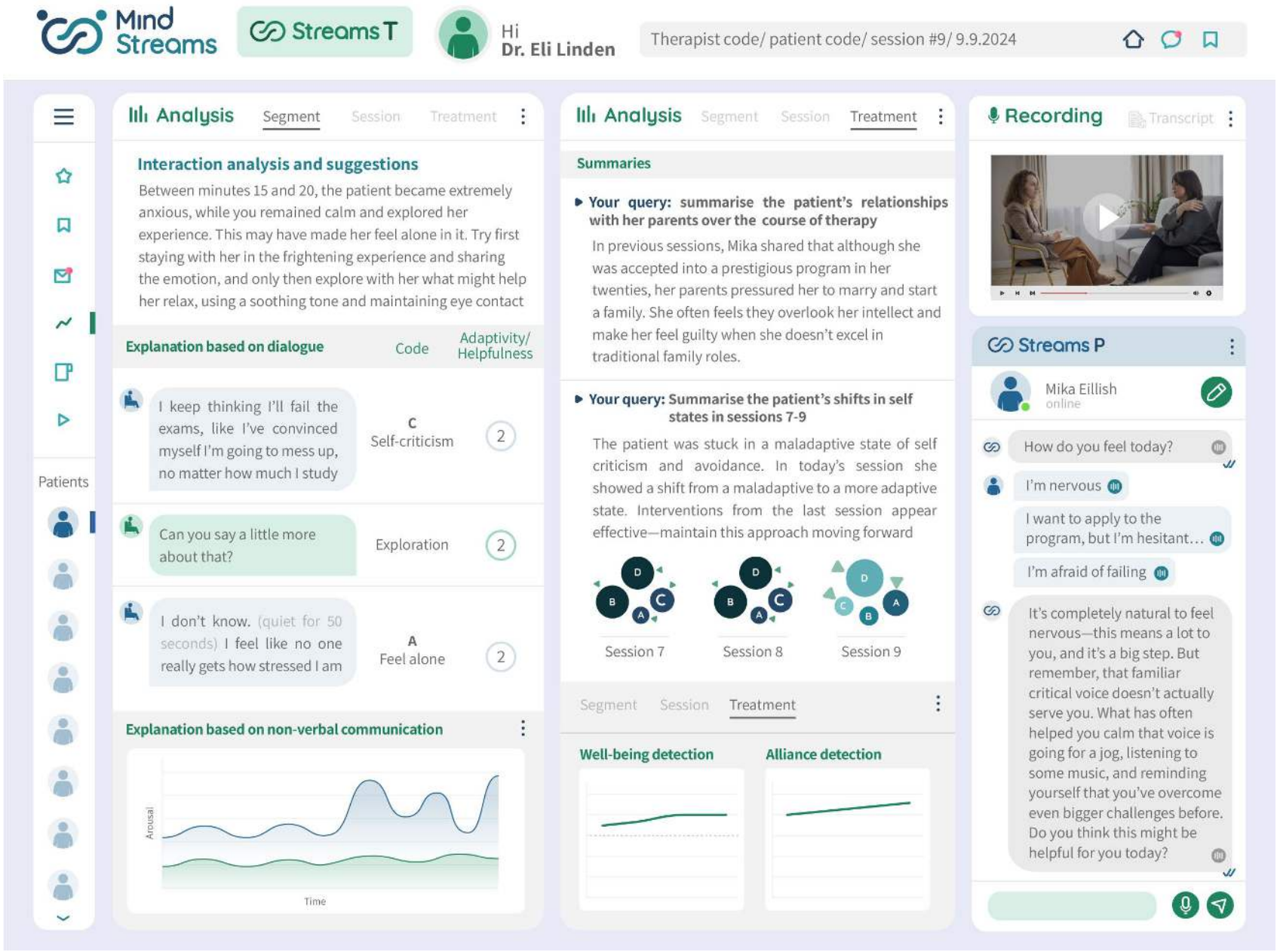
- Association between co-regulation and treatment outcome

Figure 4

The Association Between Session Outcome and Clients' Intrapersonal and Interpersonal Dampening

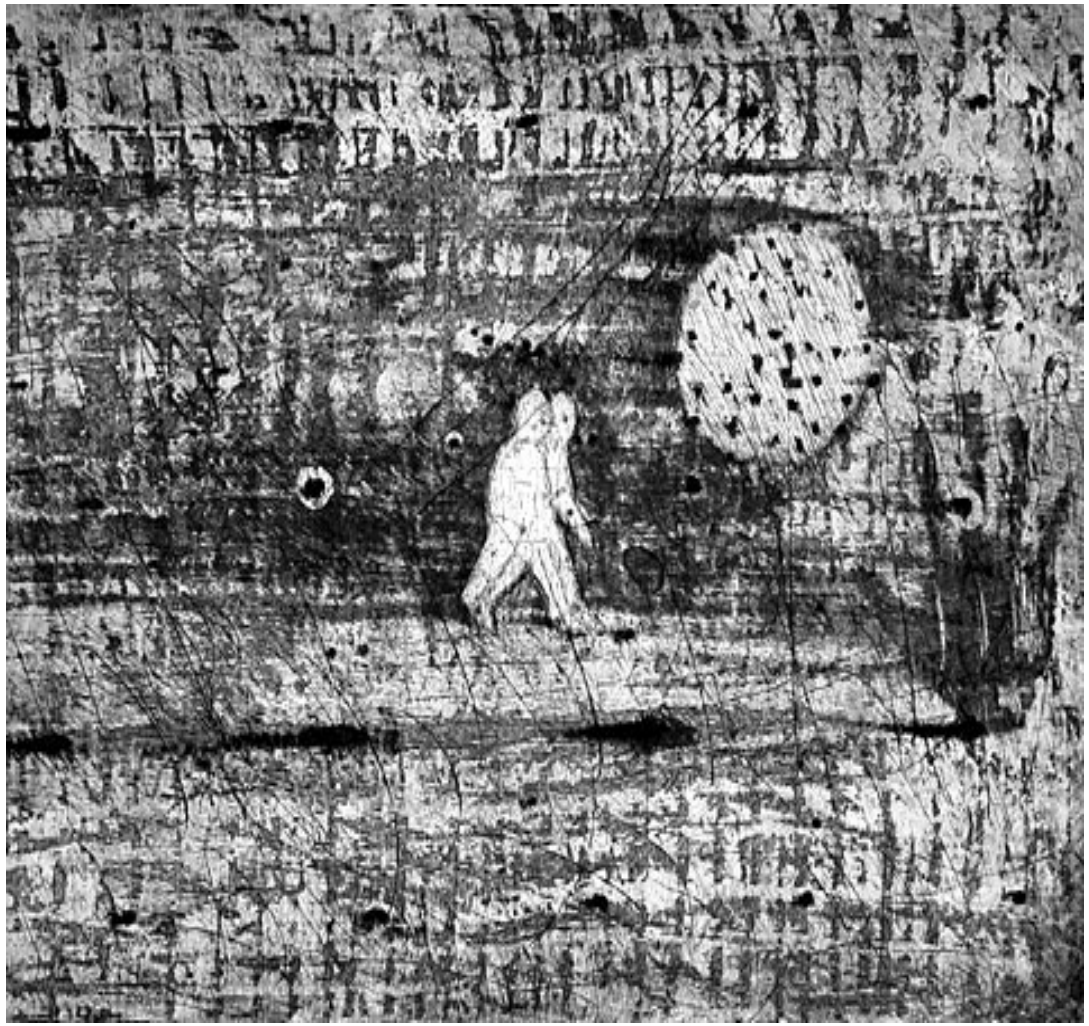


Next steps


Mind Streams
Streams T
Hi
Dr. Eli Linden
Therapist code/ patient code/ session #9/ 9.9.2024
Analysis
Segment
Session
Treatment
Interaction analysis and suggestions
Between minutes 15 and 20, the patient became extremely anxious, while you remained calm and explored her experience. This may have made her feel alone in it. Try first staying with her in the frightening experience and sharing the emotion, and only then explore with her what might help her relax, using a soothing tone and maintaining eye contact
Explanation based on dialogue
Code
Adaptivity/ Helpfulness
I keep thinking I'll fail the exams, like I've convinced myself I'm going to mess up, no matter how much I study
C
Self-criticism
2
Can you say a little more about that?
Exploration
2
I don't know. (quiet for 50 seconds) I feel like no one really gets how stressed I am
A
Feel alone
2
Explanation based on non-verbal communication
Arousal
Time
Patients
Analysis
Segment
Session
Treatment
Summaries
Your query: summarise the patient's relationships with her parents over the course of therapy
In previous sessions, Mika shared that although she was accepted into a prestigious program in her twenties, her parents pressured her to marry and start a family. She often feels they overlook her intellect and make her feel guilty when she doesn't excel in traditional family roles.
Your query: Summarise the patient's shifts in self states in sessions 7-9
The patient was stuck in a maladaptive state of self criticism and avoidance. In today's session she showed a shift from a maladaptive to a more adaptive state. Interventions from the last session appear effective—maintain this approach moving forward
Session 7
Session 8
Session 9
Segment
Session
Treatment
Well-being detection
Alliance detection
Recording
Transcript
Streams P
Mika Eillish
online
How do you feel today?
I'm nervous
I want to apply to the program, but I'm hesitant…
I'm afraid of failing
It's completely natural to feel nervous—this means a lot to you, and it's a big step. But remember, that familiar critical voice doesn't actually serve you. What has often helped you calm that voice is going for a jog, listening to some music, and reminding yourself that you've overcome even bigger challenges before. Do you think this might be helpful for you today?

Back to practice





Thanks!תודה

dana.slonim@gmail.com

Research: Flexibly in synchrony

Objective

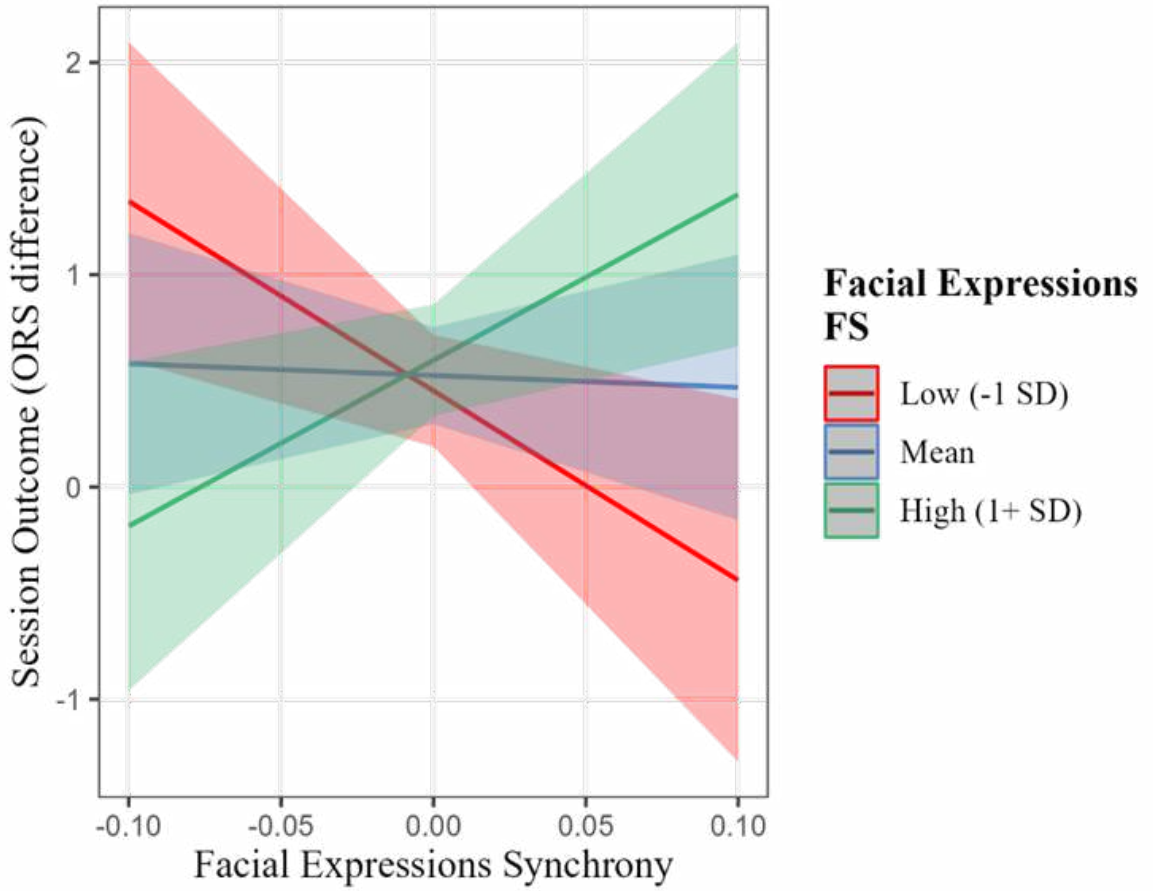
- To examine whether flexibility in synchrony moderates the association between synchrony and outcomes

Method

- 58 dyads
- Facial expression synchrony was assesses using cross-correlations (lags -5 to +5 seconds) in non-overlapping 30-second windows
- Flexibility in synchrony (FS) was indexed as the SD of the windows' synchrony values over the session

Results

- Flexibility in synchrony moderated the association between synchrony and outcome such that higher synchrony was associated with better session outcome when FS was high, but with poorer outcome when FS was low



Sayda, D. Bar-Kalifa, E., Atzil-Slonim, D., (work in progress). Can flexibility explain mixed finding on synchrony outcome association?