Thousand Voices of Trauma:
A Large-Scale Synthetic Dataset for Modeling Prolonged Exposure Therapy Conversations

Suhas BN, Andrew M. Sherrill, Rosa I. Arriaga,
Chris W. Wiese, Saeed Abdullah

Mental Healthcare Faces Rising Demand & Clinician Shortages

Current Landscape

- **Global rise in mental health issues** demanding work, declining quality of life, reduced physical activity, and digital dependencies [6-9]
- WHO: Mental health is a fundamental human right [19]
- **Critical shortage**: Limited availability of trained clinicians
- **Quality concerns**: Variability in clinician training and adherence to evidence-based practices

58 million U.S. adults with mental illness - fewer than half receive care [36]

350,000+ projected shortfall of mental health professionals by 2030 [37]

[6] Makwana, N. (2019). Disaster and its impact on mental health: A narrative review. *Journal of Family Medicine and Primary Care*, *8*(10), 3090–3095. https://doi.org/
[7] Green, B. L. (1996). Traumatic stress and disaster: Mental health effects and factors influencing adaptation. *International Review of Psychiatry*, *8*(2), 177–210.
[8] Lazar, S. G. (2014). The mental health needs of military service members and veterans. *Psychodynamic Psychiatry*, *42*(3), 459–478.
[9] Johnson, E. M., & Possemato, K. (2019). Correlates and predictors of mental health care utilization for veterans with PTSD: A systematic review. *Psychological Trauma: Theory, Research, Practice, and Policy*, *11*(8), 851–860.
[19] World Health Organization. (2022). *Mental health: Strengthening our response*. Retrieved October 18, 2023, from https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response
[36] Kaiser Permanente Institute for Health Policy. (2023). *Addressing behavioral health workforce shortages*. Retrieved May 1, 2025, from https://www.kpihp.org/wp-content/uploads/2023/06/0501_MH_Workforce_infoflyer_060823_ADA.pdf
[37] Mental Health America. (2024). *Mental health statistics 2024*. The Zebra. Retrieved May 1, 2025, from https://www.thezebra.com/resources/research/mental-health-statistics/

What is Prolonged Exposure (PE) Therapy?

“…based on the idea that PTSD and trauma-related symptoms result from avoidance of traumatic memories, PE gradually exposes you to those memories, helping your brain to accept them as safe…”

The solution - Dataset Overview

Dataset Specifications

- **3,000 simulated therapy conversations**
- **500 unique clinical cases**
- **6 conversational perspectives per case:**

1. Orientation to Imaginal Exposure
2. Imaginal Exposure Duration
3. Monitoring SUDS Ratings
4. Reinforcing Comments
5. Eliciting Thoughts and Feelings
6. Processing the Imaginal

Generation Model:

> **Claude Sonnet 3.5** [208] selected based on coherent, contextually grounded dialogue (see Appendix for details)

> **Customized prompts** based on Emory University therapist feedback

> **Systematic methodology** ensuring clinical plausibility

[208] Anthropic, "Claude 3.5 Sonnet," Anthropic Blog, 2024. [Online]. Available: https://www.anthropic.com/news/claude-3-5-sonnet

Dataset Diversity

Client Profile Generation (N=500)

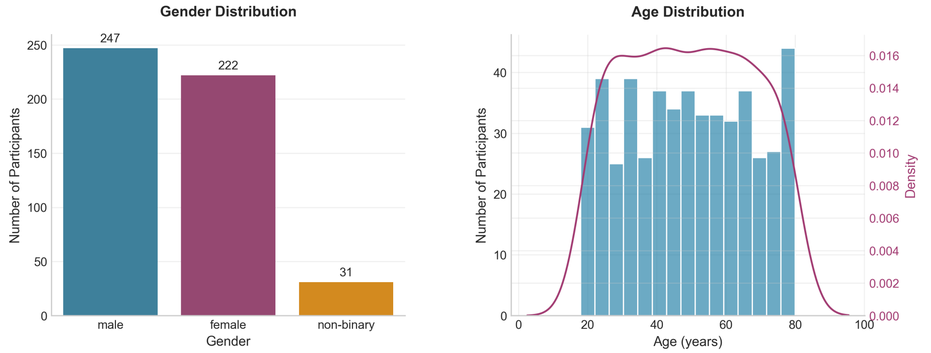
- **Age**: 18-80 years (M = 49.3), balanced across groups
- **Gender**: 49.4% male, 44.4% female, 6.2% non-binary - US Census [1]
- **Ethnicity**: 8 global regions represented
- **Relationship status**: Married (36.0%), Single (29.8%), In relationship (14.0%)

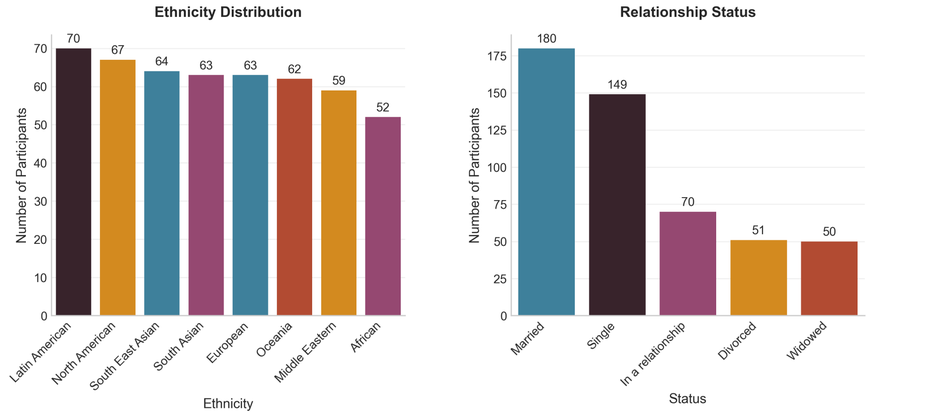
Co-occurring Conditions

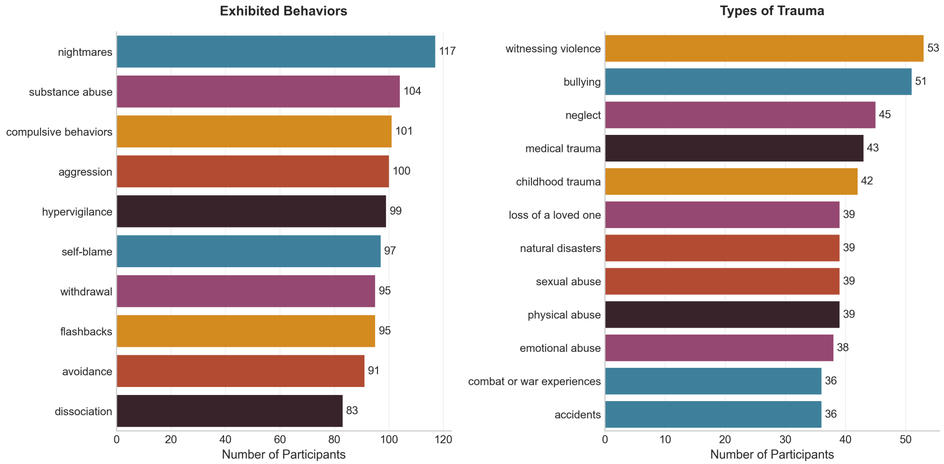
- None (25%), Anxiety (25%), Depression (30%) [209-211]
- Substance Use Disorder (10%), Chronic Pain (10%)

Trauma-Related Behaviors

- 1-3 behaviors per profile from **10 behavioral patterns**
- Most common: Nightmares (23.4%), Substance abuse (20.8%), Compulsive behaviors (20.2%)







[1] U.S. Census Bureau. (2025). QuickFacts United States. Retrieved February 18, 2025, from https://www.census.gov/quickfacts/fact/table/US/LFE046223
[209] E. Bilevicius, J. L. Sommer, M. T. Keough, and R. El-Gabalawy, "An examination of comorbid generalized anxiety disorder and chronic pain on substance misuse in a Canadian population-based survey," The Canadian Journal of Psychiatry, vol. 65, no. 6, pp. 418–425, 2020.
[210] S. Jennifer, B. R. Brady, M. M. Ibrahim, K. E. Herder, J. S. Wallace, A. R. Padilla, and T. W. Vanderah, "Co-occurrence of chronic pain and anxiety/depression symptoms in US adults: prevalence, functional impacts, and opportunities," Pain, vol. 165, no. 3, pp. 666–673, 2024.
[211] K. Hagiwara, Y. Mochizuki, C. Chen, H. Lei, M. Hirotsu, T. Matsubara, and S. Nakagawa, "Nonlinear probability weighting in depression and anxiety: insights from healthy young adults," Frontiers in Psychiatry, vol. 13, p. 810867, 2022.

Dataset Development & Validation

Study Design

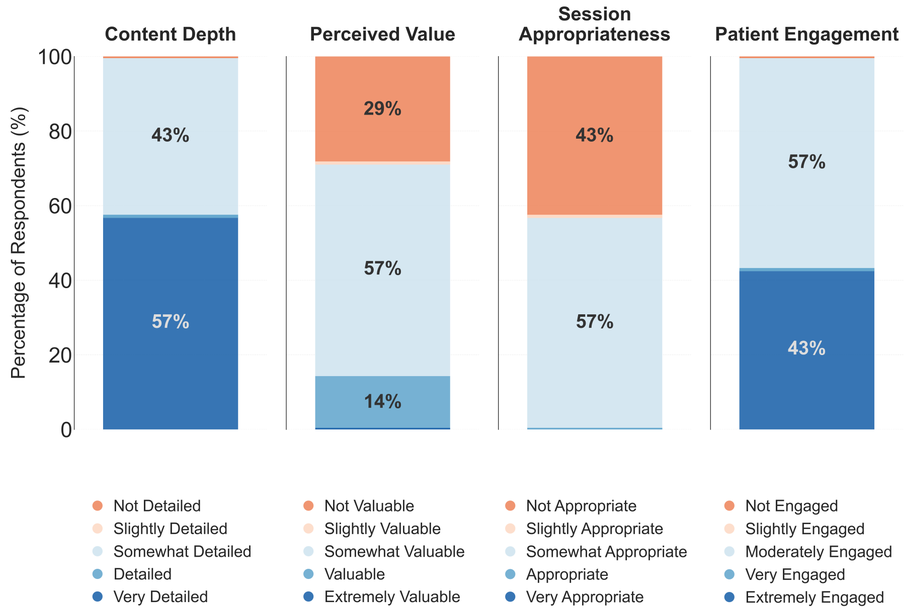
- **7 expert therapists** (6-30 years experience)
- Diverse backgrounds: Clinical practice, research, education, VA/military
- **4 evaluation dimensions**

Assessment Results

- **Content depth**: “Somewhat” to “Very detailed” (consistent strength)
- **Patient engagement**: “Moderately” to “Extremely” engaged
- **Therapist skill level**: “Novice” to “Competent”

Key Strengths Identified

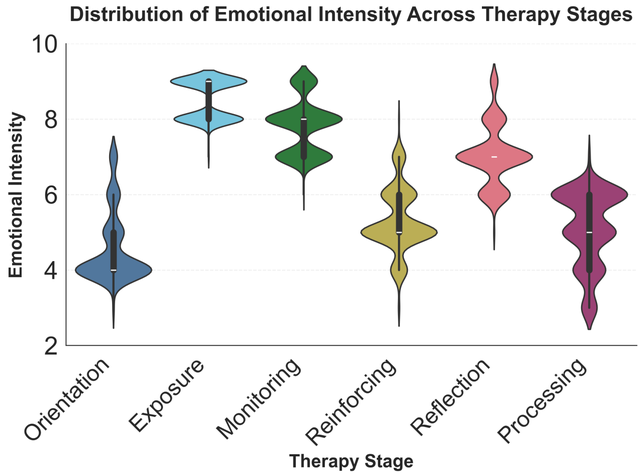
- Detailed trauma narratives ✓
- Emotional depth ✓
- Realistic therapist behavior ✓



Emotional Trajectory Benchmark

Design

- **6 therapy phases**: Orientation → Exposure → Processing
- **Expected pattern**: Initial anxiety → Peak distress → Gradual reduction
- 3 metrics: Sequence Similarity (Pearson Correlation), Pattern Accuracy (Dynamic Time Warping), and Phase Consistency (RMSE)



Results

- **Best model**: Mistral Large (S_{abs} = 0.74)
- Key findings:
  - Larger models perform better
  - Instruction tuning more critical than raw parameter count
  - Range: 0.59-0.74

(see Appendix for calculation)

Model	N*	Pearson ↑ (Avg ± S.D)	DTW ↓ (Avg ± S.D)	RMSE ↓ (Avg ± S.D)	S_{abs} ↑
Mistral Large	500	0.80 ± 0.14	2.38 ± 0.69	1.07 ± 0.33	0.74
Amazon Nova Pro	500	0.74 ± 0.16	2.63 ± 0.73	1.24 ± 0.35	0.69
Llama 3 70B Instruct	489	0.73 ± 0.16	2.61 ± 0.75	1.28 ± 0.36	0.69
Llama 3.1 70B Instruct	500	0.70 ± 0.17	2.80 ± 0.73	1.29 ± 0.35	0.67
Llama 3 8B Instruct	489	0.64 ± 0.23	3.24 ± 0.84	1.61 ± 0.43	0.61
Llama 3.1 8B Instruct	500	0.63 ± 0.23	2.91 ± 0.70	1.44 ± 0.37	0.63
Mistral 7B Instruct	500	0.62 ± 0.21	2.88 ± 0.75	1.49 ± 0.38	0.63
Mistral Small	500	0.61 ± 0.20	3.30 ± 0.94	1.70 ± 0.42	0.59

*N=489 for original Llama 3 v1 models due to limited 8k context window limit exceeded by some samples.

Thank you! Questions?

TL; DL (Too Long; Didn't Listen)

- Created “Thousand Voices of Trauma,” a large-scale synthetic dataset of Prolonged Exposure (PE) therapy sessions, with clinical validation, and a unified evaluation benchmark.

- Access:

Paper

Dataset

More Questions/ Want to collab?
Reach out to: hi@bnsuhas.com

Acknowledgements: This work was supported by the National Science Foundation (NSF) under Grant No. 2326144.