Understanding the Management of Rape Trauma with AI and Neuroimaging

Samruddhi Kamble

Undergraduate Researcher
kamblesamruddhi85@gmail.com

Abstract

Sexual trauma leaves wounds that science cannot see, yet survivors live with them every day. Traditional tools rely on words or self-reports, often forcing survivors to "bleed in silence" when their pain is doubted or dismissed. Trauma, however, is not one-dimensional. It disrupts multiple brain networks and produces states fear, vigilance, detachment that cannot be captured by words alone. This creates the need for approaches that reveal trauma's complexity in ways that are both objective and interpretable. Our framework that combines fMRI, EEG, and DINO to investigate machine learning models can identify neural or psychological patterns associated with trauma responses. Instead of producing abstract scans or opaque predictions, the system will generate exploratory measures of trauma response that support therapists' understanding while guiding future research. These measures will be presented through a simple dashboard that summarizes three indices (TPI, DI, RBS) alongside heatmaps and plain language notes. By turning complex data into clear, anonymized session snapshots, the dashboard provides researchers with a output that can be compared across participants and refined in future work.

Introduction and Methodology

I am drawn to the intersection of AI, data science, and computational neuroscience because it offers a way to make invisible suffering visible. Survivors of sexual trauma often carry pain that words cannot fully express, and the need to describe or relive their experiences during therapy can itself be retraumatizing. This research seeks to change that to understand survivors' neural patterns without requiring them to verbalize their pain. By combining functional MRI and EEG with an interpretable self-supervised AI model, DINO (Distillation with No Labels) (Caron et al. 2021), this work explores how the brain encodes fear, dissociation, and hypervigilance after trauma (Shymanskaya et al. 2023). Here, interpretation does not mean subjective inference, but the translation of complex brain activity into transparent, clinically relevant indices that reflect emotional and regulatory states. As someone who has seen firsthand how silence and misinterpretation can shape diagnosis, I am deeply motivated to build systems that bridge this gap. My goal is to transform DINO into a system that listens to what the brain communicates when words fall short, and to explore how the human brain perceives the world after trauma. By uniting spatial (fMRI) and temporal (EEG) signals through ethically grounded AI, this framework aims to make invisible suffering visible-offering clinicians trustworthy insights into survivors' neural responses. This approach does not replace clinical judgment, it complements it. Bridging data with empathy and algorithms with the human experience of pain. This study proposes a five-stage framework for exploring trauma-related brain patterns using multimodal neuroimaging and AI:

1. **Stage 1: Data Processing:** Preprocess fMRI and EEG datasets using standard methods (motion correction, ICA, filtering).
2. **Stage 2: Interpretability:** Apply explanation methods to identify active brain regions during trauma responses.
3. **Stage 3: DINO Modeling:** Train self-supervised model on fMRI–EEG data to detect patterns of fear, dissociation, hypervigilance, and shame.
4. **Stage 4: Dashboard Design:** Generate three indices (TPI, DI, RBS) with heatmaps and plain-language notes.
5. **Stage 5: Validation:** Compare indices against controls and clinical measures (CAPS-5) to assess feasibility.

Stage 1: Multimodal Acquisition with fMRI and EEG & Data Processing

The first stage tests the framework using multimodal sexual-trauma datasets containing both fMRI and EEG recordings. Together, these modalities capture where and when trauma alters brain dynamics fMRI providing spatial precision across the Salience (SN), Default Mode (DMN), Fronto Parietal (FPN), and Dorsal Attention (DAN) networks, and EEG revealing millisecond-level timing of fear, hypervigilance, and dissociative states. Resting-state data establish chronic connectivity patterns, while task-based sessions with trauma-related and neutral narratives expose transient reactivation of memories. To analyze these effects, brain activity will be represented as graph-based connectivity networks, linking spatial and temporal signals into a unified model. Standard preprocessing ensures data quality, but the focus remains on how trauma reorganizes communication between neural systems. The resulting multimodal

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

maps provide the biological foundation for later AI stages objective markers that complement subjective reports and help visualize the hidden architecture of traumatic experience (Shymanskaya et al. 2023).

Stage 2: Neuroimaging AI for Interpretability

A major limitation of current neuroimaging AI is opacity: most models act as black boxes, producing outputs without explaining them. In trauma research, that's more than a flaw it risks dehumanizing survivors' experiences. Stage 2 therefore focuses on adding interpretability safeguards before any results are shared. We apply SmoothGrad, an explanation technique that highlights which brain regions contribute most to the model's prediction (Khatri and Kwon 2023). Unlike GradCAM or LRP, SmoothGrad averages over multiple noisy samples to yield clearer, more stable relevance maps. These are transformed into Trauma Response Maps-visual summaries linking patterns of activity to trauma states such as fear, hypervigilance, or dissociation. Each map carries a Trust Score, reflecting accuracy, stability, and error bounds. Stage 2 will be helpful in showing how AI can be made more transparent and safe for trauma research. By adding trauma specific checks and basic reliability measures, this stage ensures that results are easier to interpret for clinicians and more respectful toward survivors.

Stage 3: DINO Training

This stage adapts DINO (Distillation with No Labels) (Caron et al. 2021), a self-supervised vision transformer, to learn trauma-related brain patterns without large labeled datasets (Gaziv et al. 2022). Multimodal fMRI-EEG inputs capture spatial and temporal dynamics, guided by key networks (Salience, Default Mode, Fronto-Parietal, Dorsal Attention) governing perception, emotion, and control (**?**). DINO's embeddings are modeled as statistical predictors of three neural indices: Threat Perception (TPI), Dissociation (DI), and Regulation Balance (RBS). The model learns through probabilistic supervision using weak cues from EEG rhythms (alpha $\rightarrow$ dissociation, theta $\rightarrow$ recall), physiological peaks (heart rate, skin conductance), and narrative timing. We extend DINO with four lightweight "emotion heads" that estimate continuous intensities for fear, hypervigilance, shame, and dissociation, transforming these into interpretable indices: TPI quantifies threat intensity, DI measures freeze-like states, and RBS captures the balance between cognitive regulation and self-blame.

To ensure clinical validity, these indices are benchmarked against the Clinician-Administered PTSD Scale (CAPS-5) the gold standard structured interview assessing 20 DSM-5 symptoms. To ensure clinical validity, these indices are benchmarked against the Clinician-Administered PTSD Scale (CAPS-5). Expected relationships include: TPI correlating with arousal and intrusion symptoms (hypervigilance, startle response); DI reflecting avoidance and detachment (emotional numbing, withdrawal); and RBS (inverse) mapping to negative mood and self-blame (guilt, shame, low affect). This mapping anchors neural indices to clinical constructs, ensuring interpretability. DINO's representations aligned with CAPS-5 transform brain activity into measurable evidence of trauma, revealing the hidden neurodynamics of pain and recovery.

Stage 4: Output Interpretation

After training, DINO's outputs will be summarized into three research indices (TPI, DI, RBS), an anonymized activation map, and a brief explanatory note for research validation. All fMRI and EEG data will be processed offline, generating indices benchmarked against normative ranges, spatial activation maps, and plain-language notes (e.g., 'This pattern shows the brain staying in threat mode'). Full clinical dashboard development will be deferred to future work pending validation of index reliability.

Background Work

Recent advances show that self-supervised models like DINO can learn meaningful neural representations without large labeled datasets (Caron et al. 2021; Gaziv et al. 2022; Khatri and Kwon 2023). However, their application to trauma-related brain states remains unexplored. Clinical research identifies trauma as a network-level disorder disrupting the Salience, Default Mode, Fronto-Parietal, and Dorsal Attention networks (Shymanskaya et al. 2023). Yet existing analyses fail to integrate spatial (fMRI) and temporal (EEG) dimensions in ways that are both rigorous and clinically interpretable. Building on prior work in interpretable AI for mental-health prediction, this research integrates DINO with multimodal neuroimaging to translate neurobiological traces of trauma into interpretable indices (TPI, DI, RBS). By grounding AI in both empathy and evidence, this study bridges computational neuroscience and mental-health research, setting the stage for more ethical, interpretable trauma science.

Conclusion

Sexual trauma leaves deep neurological imprints that are difficult to quantify through self-reports or standard assessments. This work integrates fMRI and EEG with DINO's self-supervised and interpretable learning framework to uncover measurable patterns of fear, dissociation, and regulation. These are summarized through three indices Threat Perception (TPI), Dissociation (DI), and Regulation Balance (RBS) that make complex brain responses both interpretable and ethically usable. The framework aims to provide measures that researchers and clinicians can study more clearly, without overstepping into diagnostic claims. The ultimate goal is to turn silent suffering into measurable, interpretable science that opens new directions for trauma research in neuroscience and AI.

Ethics and Participant Safety: All data collection will follow trauma-informed protocols with full IRB, HIPAA, and GDPR compliance. This work will be conducted at the Computer Neuroscience Lab with proper ethical supervision. All data will remain within authorized lab systems.

Acknowledgments

I would like to express my sincere gratitude to Dr.Vincent Lostanlen for his guidance, insightful feedback, and encouragement throughout this project.

This work is dedicated to every survivor whose pain was never heard. I hope science one day gives words to what silence could not.

References

Caron, M.; Touvron, H.; Misra, I.; Jégou, H.; Mairal, J.; Bojanowski, P.; and Joulin, A. 2021. Emerging Properties in Self-Supervised Vision Transformers. *arXiv preprint arXiv:2104.14294*.

Gaziv, G.; Beliy, R.; Granot, N.; Hoogi, A.; Strappini, F.; Golan, T.; and Irani, M. 2022. Self-Supervised Natural Image Reconstruction and Large-Scale Semantic Classification from Brain Activity. *NeuroImage*, 254: 119121.

Khatri, U.; and Kwon, G.-R. 2023. Explainable Vision Transformer with Self-Supervised Learning to Predict Alzheimer's Disease Progression Using 18F-FDG PET. *Bioengineering*, 10(10): 1225.

Shymanskaya, A.; Kohn, N.; Habel, U.; and Wagels, L. 2023. Brain Network Changes in Adult Victims of Violence. *Frontiers in Psychiatry*, 14: 1040861.