# Discrete‐Latent Video Modeling of BOLD Dynamics

## 1. Why?

• Dimensionality reduction: fMRI produces long sequences of high-dimensional 3D volumes; discretizing each volume into a compact set of tokens makes temporal modeling tractable.  
• Spatiotemporal prediction: Leveraging video-modeling advances can reveal how brain activity patterns evolve, offering insights into normal cognition and early warning for pathological events.  
• Unified framework: A single pipeline can forecast future volumes, detect anomalies (e.g., pre-ictal states), and support stimulus decoding.

## 2. What?

1. Discrete Encoding  
 - Use a 3D VQ-VAE to learn a finite codebook of volume “tokens,” representing each fMRI snapshot as a grid of discrete symbols.  
  
2. Temporal Modeling  
 - Fit a causal transformer over sequences of these token grids to predict the next timepoint’s codes.  
  
3. Downstream Applications  
 - Forecasting: Anticipate upcoming brain states in resting-state or task scans.  
 - Anomaly Detection: Flag large deviations between predicted and observed tokens as potential markers of seizures or other events.  
 - Stimulus Decoding: Invert the model to infer unobserved stimuli or cognitive states from predicted codes.

## 3. How?

- Data & Preprocessing: Select appropriate fMRI datasets, apply standard corrections and normalization.  
- VQ-VAE Training: Optimize reconstruction and quantization objectives to build the discrete codebook.  
- Transformer Training: Teach an autoregressive model to forecast token sequences based on past context.  
- Integration: Chain encoder→transformer→decoder for end-to-end prediction and anomaly scoring.

## 4. Impact

• Efficiency: Dramatically shrinks the data footprint of each volume for faster sequence learning.  
• Novelty: First application of discrete-latent video modeling to BOLD dynamics.  
• Utility: Supports multiple research and clinical goals—prediction, detection, and decoding—within one cohesive framework.