Brain LM: A Foundation Model for Brain Activity Recordings

MLBBQ June 21 2024

Self Supervised Learning

Conventional supervised learning methods depend significantly on the quantity of labeled training data.

Self Supervised Learning - eliminates the need for annotations, allows models to learn with the data itself providing the supervision.

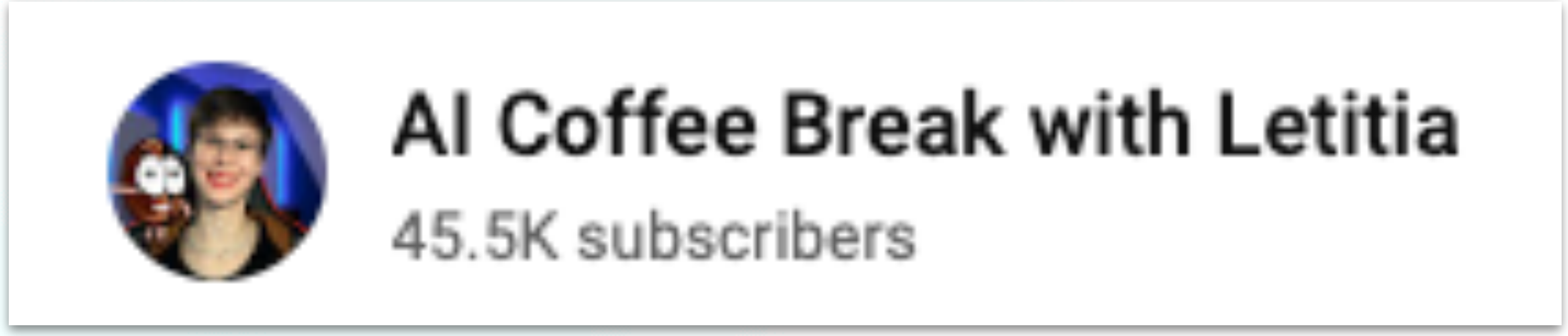
Masked Prediction Training - enables models to predict missing parts of data, improving their understanding of the dataset.

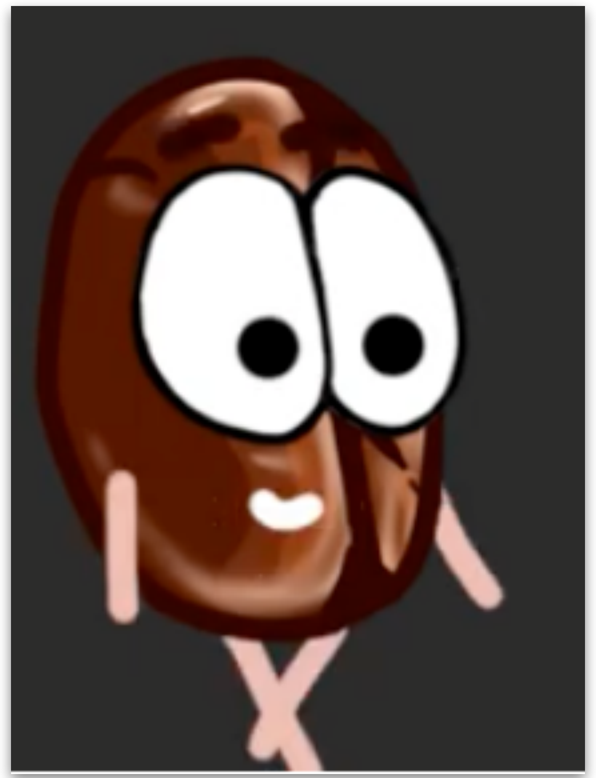
AI Coffee Break :)

Part of this presentation is heavily influenced by the AI Coffee Break with Letitia YouTube Channel…

…where Mrs. Coffee Bean explains AI concepts to you!

Make sure to check out her channel!

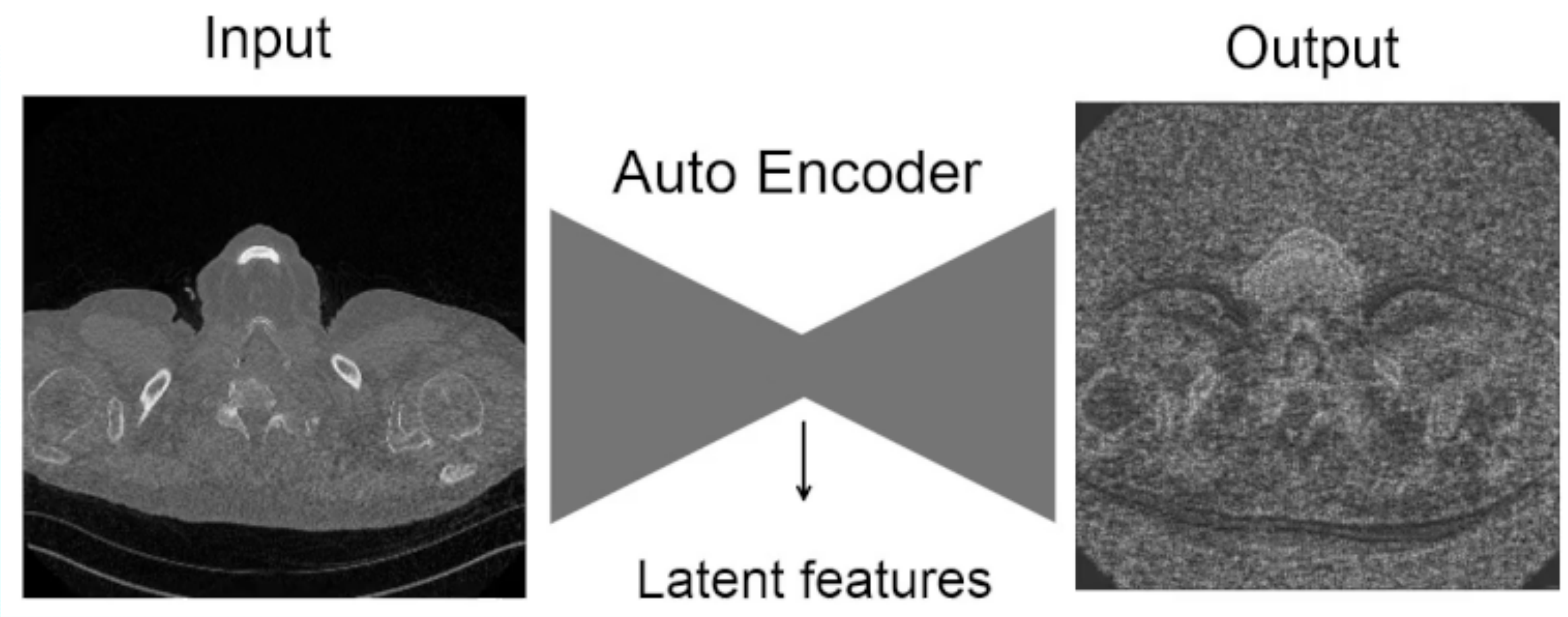


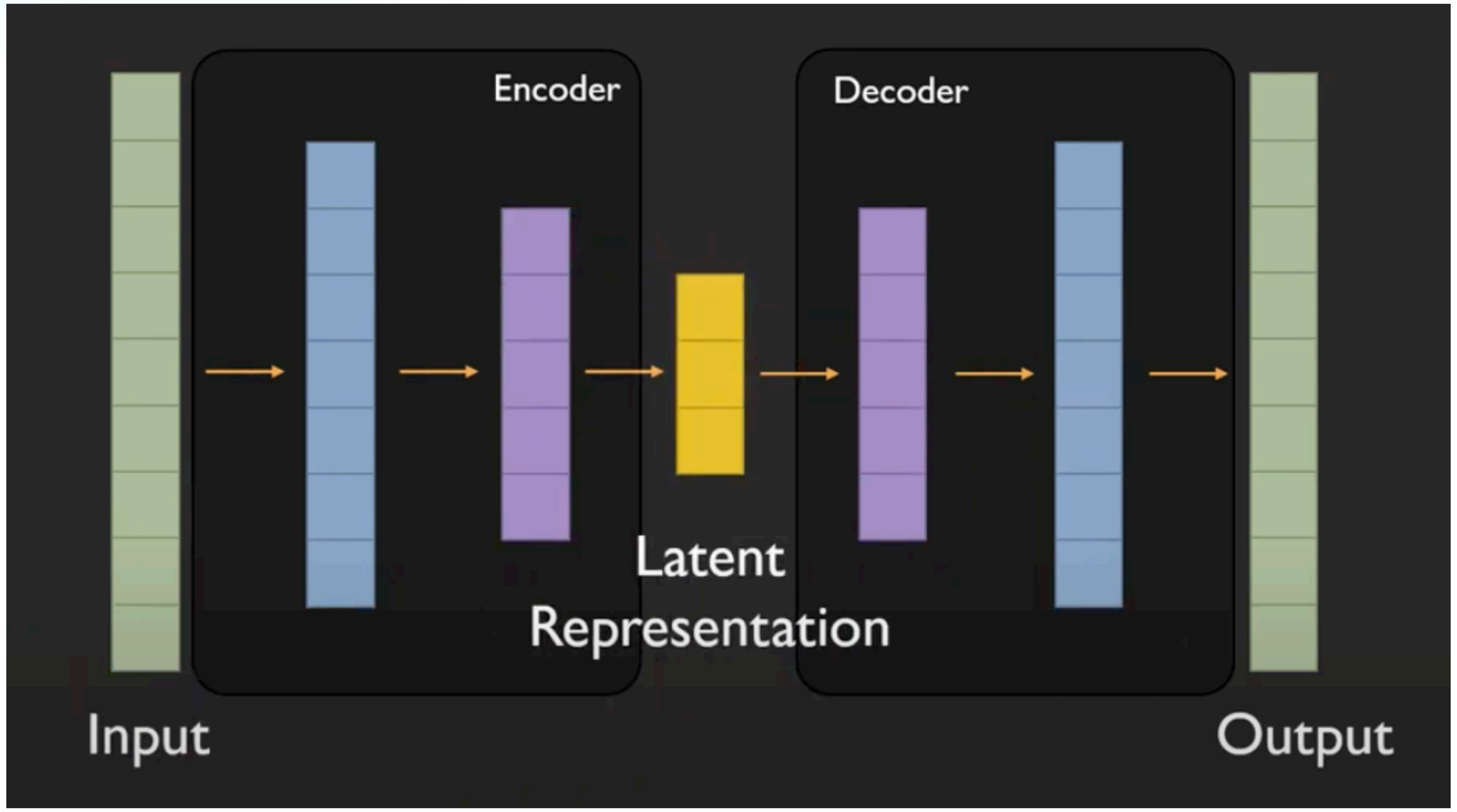


Autoencoders

An artificial neural network that reconstructs the input while being constrained to encode the input to a lower dimensional space then reconstruct the input again.

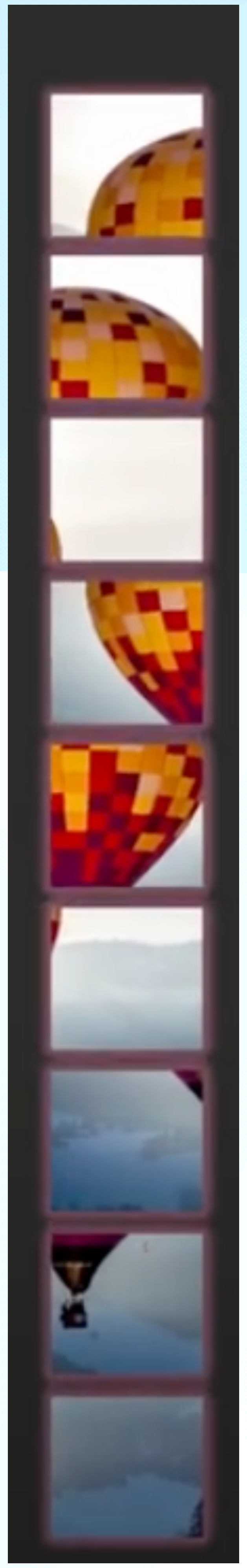
Enforced by L2 loss, simple distance metric between input and output.

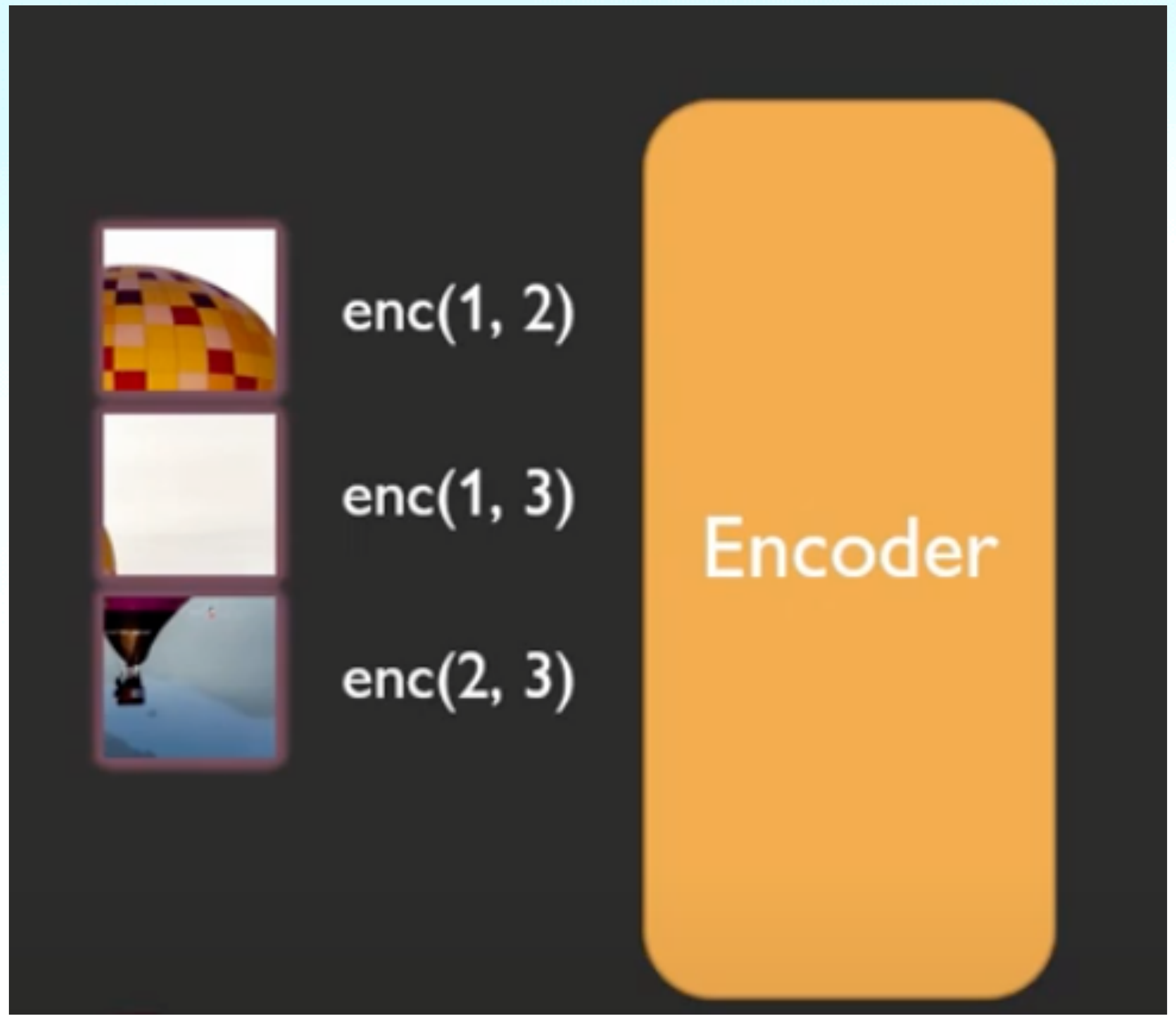


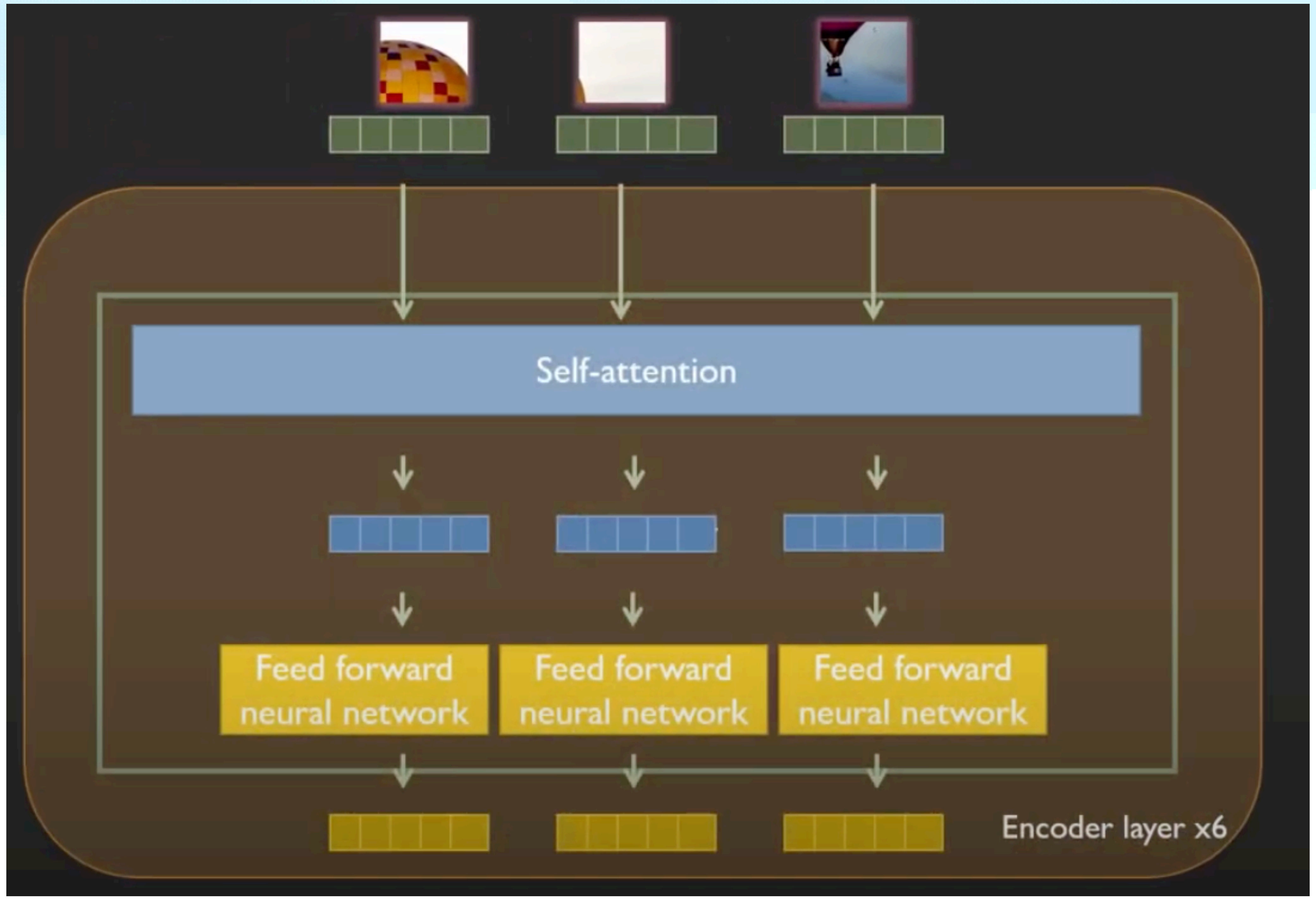


Masked Autoencoders: Encoder

Segments image into patches, adds positional embedding, encodes patches into latent space.







Masked Autoencoders: Decoder

Takes tokens from latent space and computes original image patches.

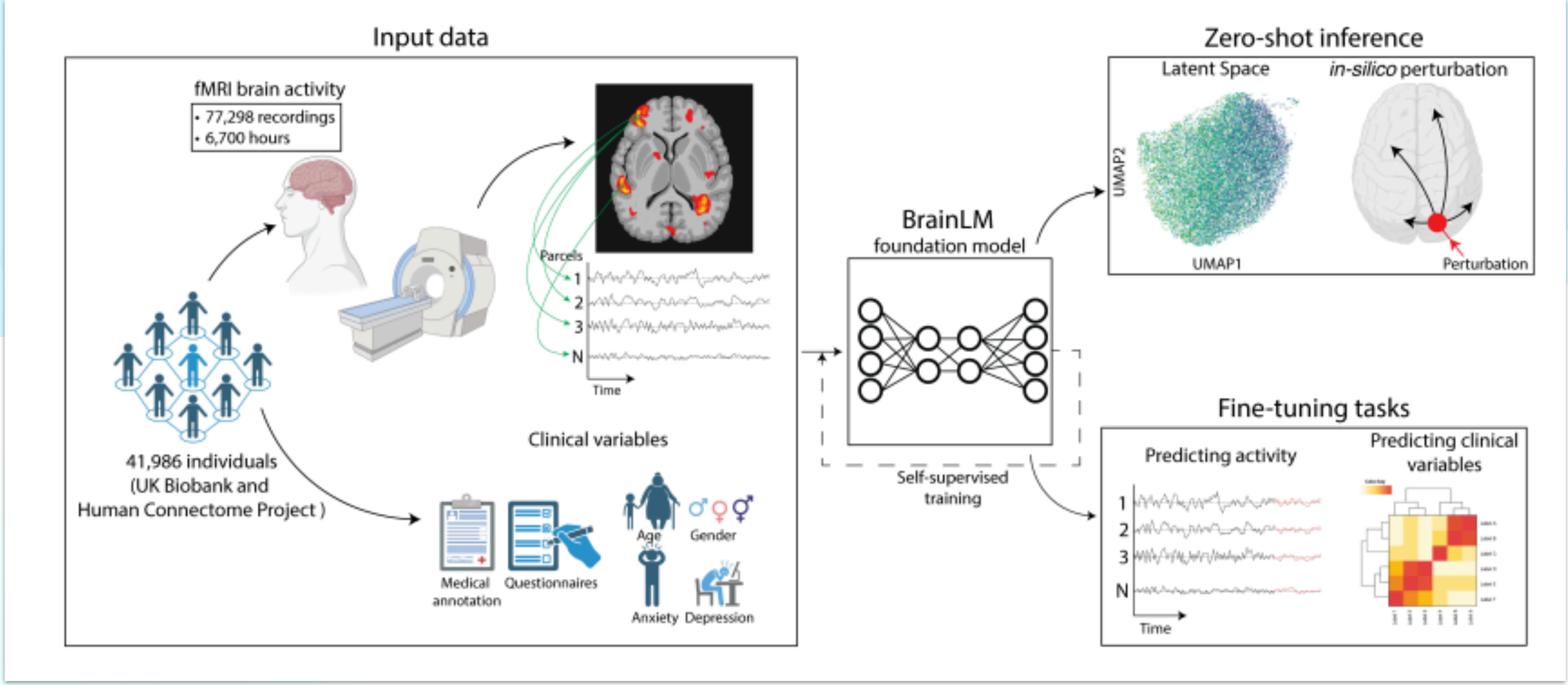
Brain LM Overview

- **Trained on 6,700 hours of fMRI data** - uses self-supervised masked-prediction training.

- **Excels in fine-tuning** - for predicting clinical variables and zero-shot inference for identifying functional networks.

- **Novel prompting technique** - enables BrainLM to simulate brain activity responses, enhancing analysis of large-scale brain data.

Architecture Overview



Introduction

- **Understanding Cognition** from brain activity
  - Using fMRI -> spatio temporal dynamics

- **Non-Linear Interactions** are challenging for some models
  - GLM, ICA, Seed-based, ML models w/ linear assumptions

- **Utilizing scale and complexity** of fMRI data
  - Need adaptable frameworks to fully utilize large datasets

Related Work

- Early fMRI analysis employed SVM and neural networks for task-specific decoding.

- Recent methods focus on training autoencoders and mapping techniques for more versatile fMRI representations.

- Limited dataset sizes continue to constrain the robustness and transferability of these approaches.

Datasets and Preprocessing

- **UK Biobank** (UKB) and **Human Connectome Project** (HCP)
  - fMRI, 40k+ subjects

- **Model training**
  - 80% of UKB (61k recordings)
  - Validated on remaining UKB & HCP dataset

- **AAL- 424 brain atlas** parcellation
  - 424 regions
  - Each region has 1 time course

Model Architecture

- **Transformer based architecture** like BERT and Vision Transformer, employs a masked autoencoder approach

- **Multi-headed self-attention** in the encoder
  - process visible patches

- **Decoder reconstructs** encoded patches
  - both visible and masked patches

- Timeseries **embedded and masked**
  - parcel time series segmented
  - partially masked

Training Procedure

- **fMRI recordings segmented**: into random 200-timestep subsequences, each containing 10 segments of 20 timesteps.

- **Segments embedded**: 512-dimensional space and masked at ratios of 20%, 50%, or 75%.

- **Transformer encoder**: 4 layers, 4 heads. Decoder: 2 layers. Trained 100 epochs, batch size 512, Adam optimizer. Minimized MSE for downstream tasks.

Clinical Variable Prediction

- Fine-tuned BrainLM for clinical variable prediction: appended MLP head, trained for age, neuroticism, PTSD, and anxiety.

- Applied normalization and scaling: Z-score for age, min-max for neuroticism, log transformation plus min-max for PTSD and GAD7.

- Evaluated regression on unseen data: used UK Biobank test set, compared performance with baseline models using raw data and pretrained embeddings. Applied 40% dropout during fine-tuning.

In Silico Perturbation Simulation

- Introduced in silico perturbation prompting to explore BrainLM's understanding of brain dynamics computationally.

- Defined perturbation function $G(X\_input, \theta)$ modifying fMRI sequence X~ to simulate cognitive conditions.

- Optimized θ to minimize distance between predicted (X^) and target (XT) states, revealing BrainLM's sensitivity to task-driven changes in visual cortical areas.

Overview of in silico perturbation approach

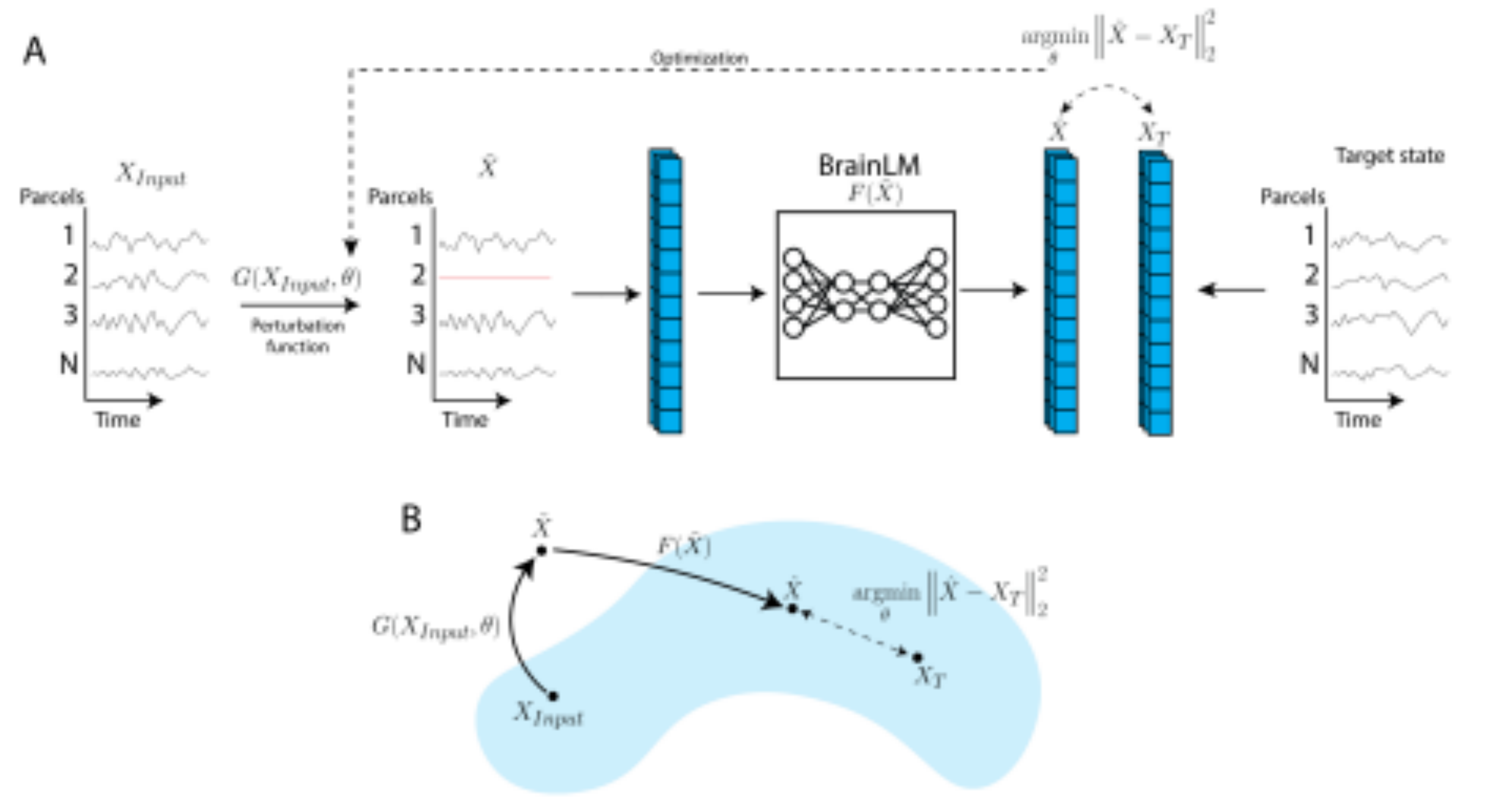


Figure 6: Overview of in silico perturbation approach. The goal of this procedure is to use the pre-trained BrainLM foundation model to simulate the effect of perturbations and find the modifications that result in a desired brain state. To achieve this, A) we learn a perturbation function G, which given an input state X_{Input} and parameters θ produces a perturbed state $\tilde{X}$. To find the optimal θ, we encode and decode the output of G and minimize the loss between the decoded perturbed state $\hat{X}$ with the target state X_T. B) Schematic of the *in silico* brain perturbation algorithm in the model manifold space.

- Figure 6 illustrates the in silico perturbation approach using BrainLM.
- The procedure involves learning a perturbation function G to modify brain states.
- Optimization minimizes the difference between the perturbed state Xˆ and the target state XT in the model manifold space.

Model Generalization

- Evaluated BrainLM on UK Biobank (UKB) test data and HCP dataset.

- Achieved strong generalization with an average R^2 score of 0.402 on UKB test set, influenced by training data size and masking ratio.

- Generalized well to HCP dataset with an R^2 score of 0.316, demonstrating robust representation learning across different datasets (see Table 1 and Figure 3).

Prediction of Clinical Variables

- Evaluated BrainLM's clinical prediction abilities by fine-tuning on metadata variables from UK Biobank.
- Appended MLP head to pretrained encoder for age, neuroticism, PTSD, and anxiety disorder regression.
- Achieved lower mean squared error compared to SVMs on raw fMRI data and BrainLM's embeddings, showcasing BrainLM's robust representation learning and transferability. (Table Below)

Table 2: Results for the regression of the clinical features. The values show the MSE (mean ± std)

	Age	Post traumatic stress disorder (PCL)	General Anxiety disorder (GAD7)	Neuroticism
Raw data	2.0 ± 0.2219	0.034 ± 0.0027	0.172 ± 0.0066	0.160 ± 0.0137
BrainLM Pretrained	0.857 ± 0.1135	0.022 ± 0.0019	0.094 ± 0.0079	0.086 ± 0.0047
BrainLM fine-tuned	$\mathbf{0.485 \pm 0.0252}$	$\mathbf{0.018 \pm 0.0008}$	$\mathbf{0.074 \pm 0.0053}$	$\mathbf{0.072 \pm 0.0049}$

Interpretability via Attention Analysis

- **Enhanced Interpretability**
  - visualized self-attention weights
  - reveals representations
- **Task-based fMRI**
  - heightened attention in the visual cortex
  - compared to resting states
- **Mild vs severe depression**
  - Attention patterns differ
  - Funct. variations relevant to clinical conditions

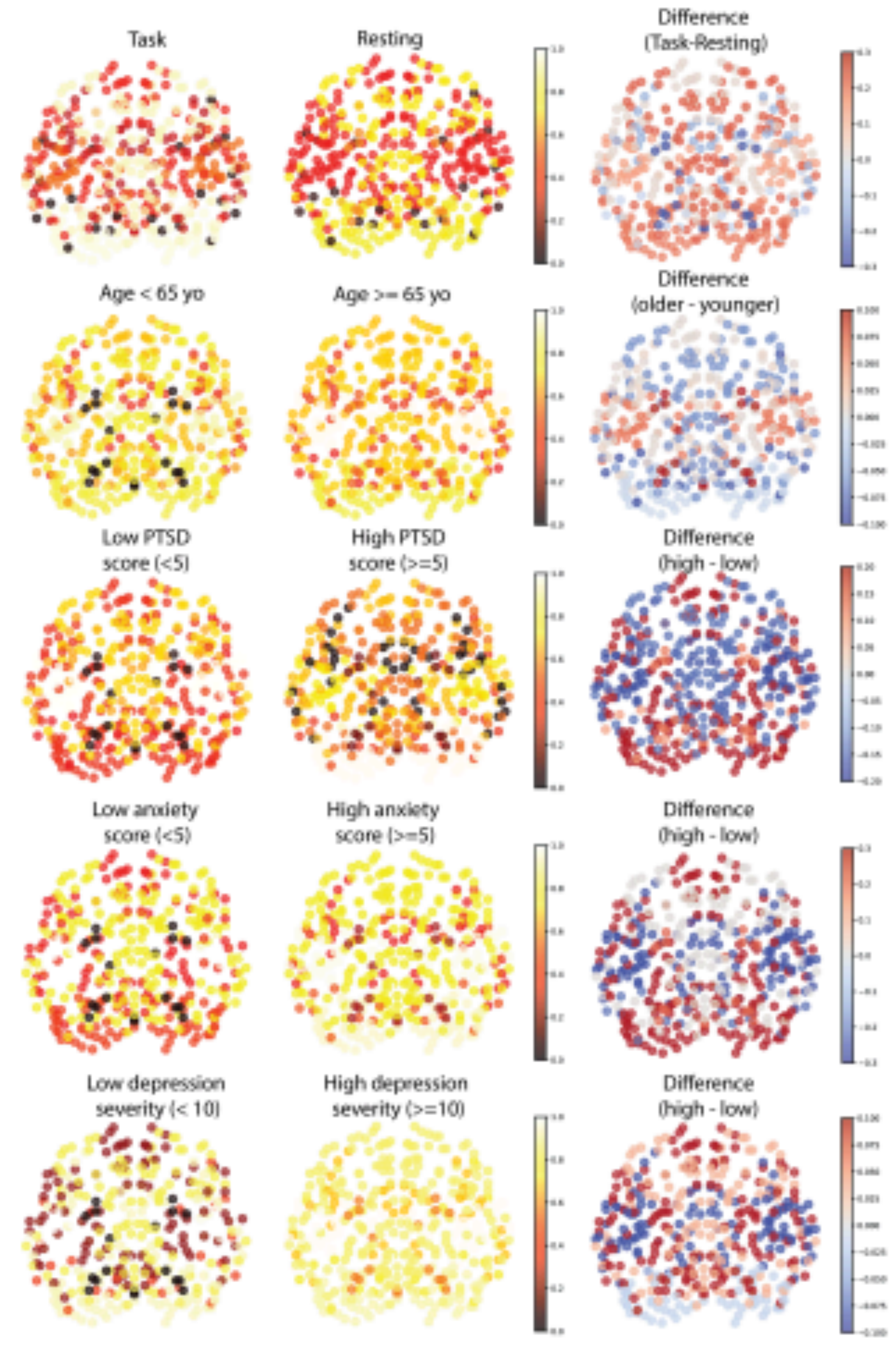
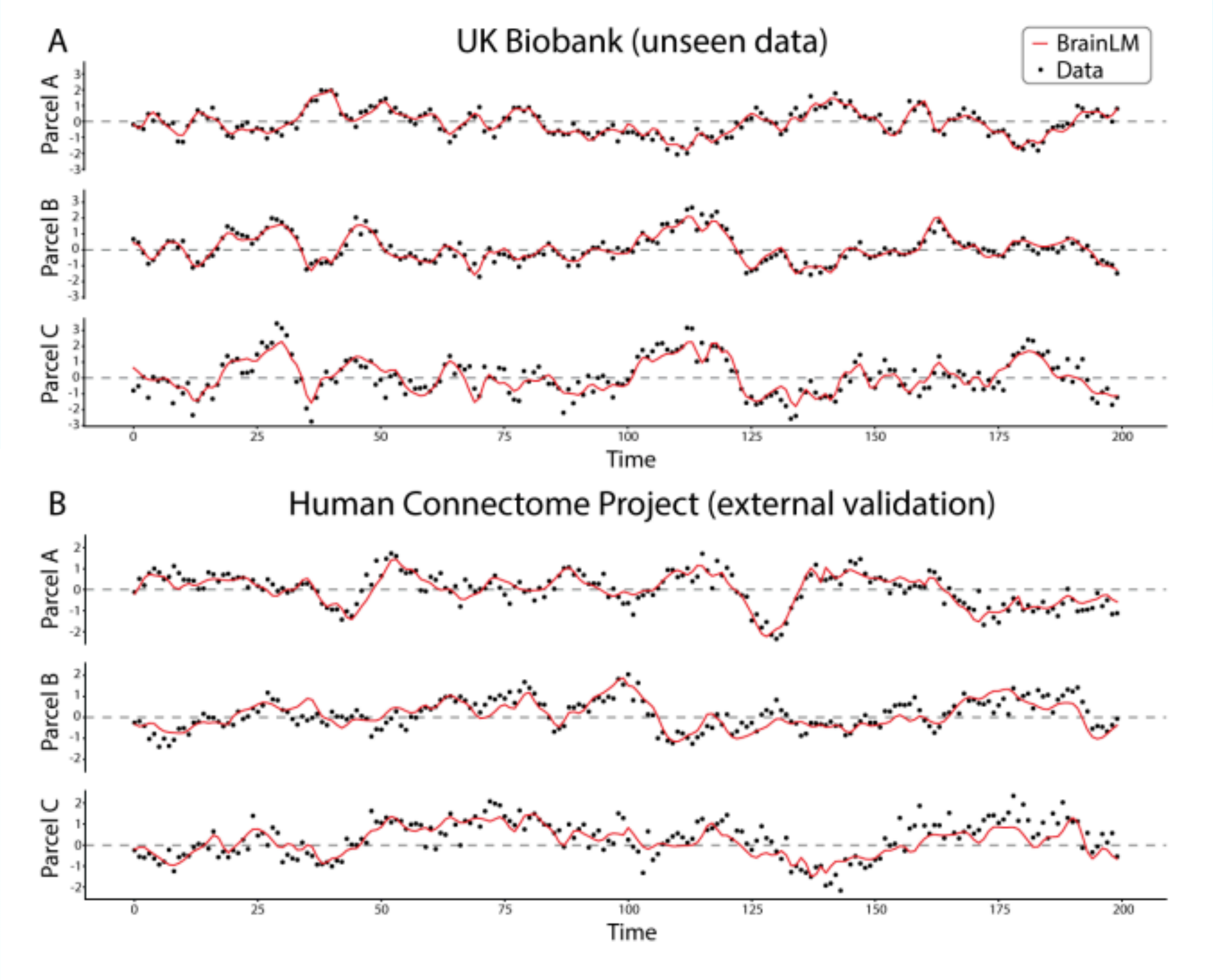


Figure 9: Attention visualization across 424 parcels. Each parcel is localized by its X and Y position and it is colored by the attention intensity with respect to the CLS token.

Results: Reconstruction performance on held-out data



- Accurately reconstructs data from both its training cohort (A, UKB) and an external cohort (B, HCP), demonstrating strong generalization.
- Model's ability to match predictions with ground truth showcases its robust performance across different datasets and subjects.

In silico Perturbation Analysis Reveals Functional Connectivity

- **Used pre-trained representations**
  - perturbation analysis **->** revealed functional brain networks

- **Optimized perturbations**
  - b/w resting-state & task-based fMRI
  - revealed key regional differences **->** visual cortex

- **BrainLM's able to uncover**
  - intrinsic knowledge of functional connections
  - across different contexts and age groups
  - without additional tuning

In silico perturbation of resting state to match task-based recordings reveals functional changes.

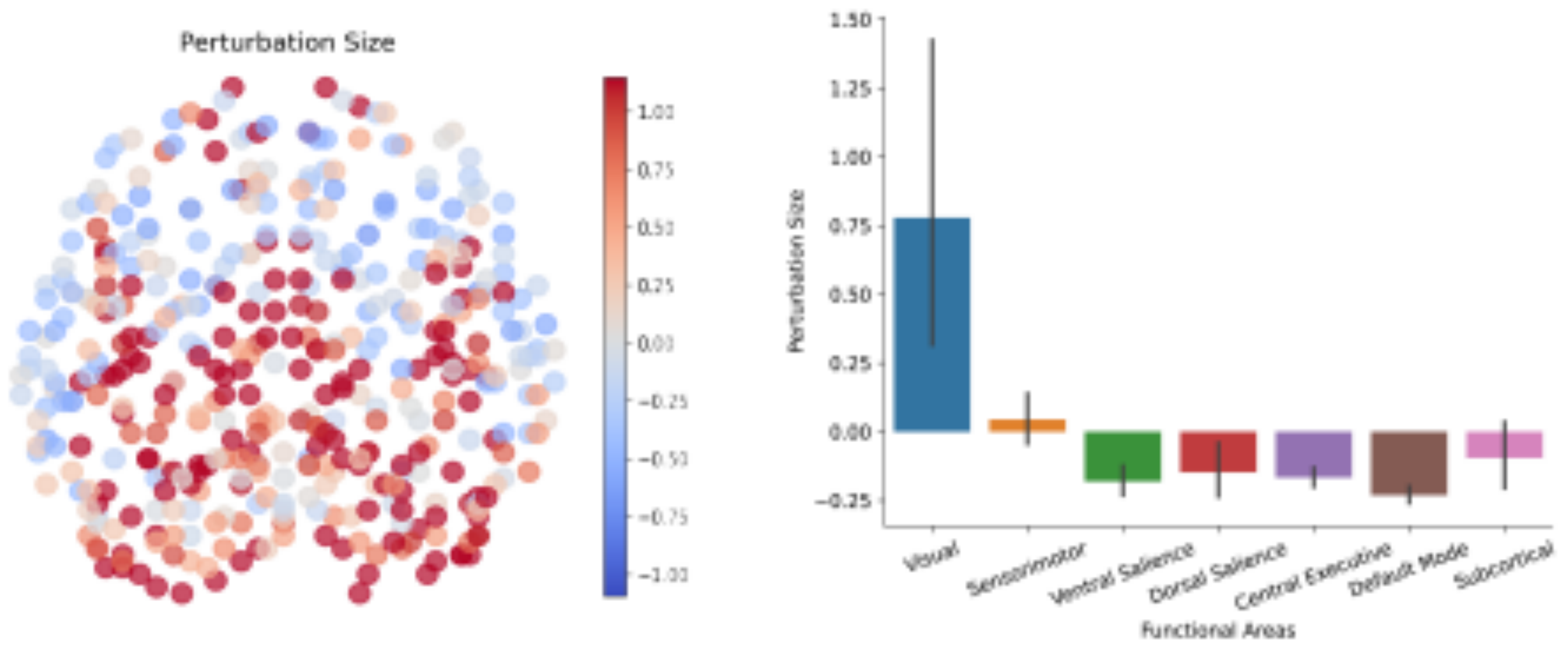


Figure 7: In silico perturbation of resting state to match task-based recordings reveals functional changes. The average magnitude of optimized perturbations to make resting state CLS tokens match target task CLS tokens. We find that the region with the largest predicted perturbation is the visual cortex, in line with expected functional changes between resting state and task-based recordings. This in silico perturbation approach demonstrates BrainLM's ability to simulate responses in a biologically meaningful manner.

- Figure 7 demonstrates in silico perturbation from resting state to task-based recordings using BrainLM.
- Optimized perturbations primarily affect the visual cortex, reflecting expected functional changes.
- - This approach showcases BrainLM's capability to simulate biologically relevant responses effectively.

Functional Network Prediction

- Evaluating BrainLM’s ability to segment fMRI parcels into 7 intrinsic functional brain networks.
- Comparing parcel classification methods including k-NN on raw data, VAE embeddings, GCN embeddings, and BrainLM’s self-attention weights.
- BrainLM's attention-based approach outperformed other methods, achieving 58.8% accuracy, demonstrating unsupervised learning of brain networks solely from pretraining.

	Accuracy (%)
BrainLM (attention weights)	**58.8**
Raw Data	39.2
Variational Autoencoder	49.4
Graph Convolutional Network	25.9

Discussion

- First foundation model for fMRI analysis
  - Self-supervised pretraining
  - Extensive brain activity data

- Demonstrates high accuracy
  - Reconstructs masked brain activity sequences
  - Generalizes well to unseen distributions

- Enables
  - biomarker discovery
  - computational modeling of brain function

- Perturbation analysis for interpretability without live brain intervention