GLAFE: WHOLE-BRAIN EEG-TO-FMRI PREDICTION VIA GLOBAL–LOCAL MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

Functional magnetic resonance imaging (fMRI) provides spatially detailed measurements of whole-brain activity, but its high cost and limited accessibility constrain its use. In contrast, electroencephalography (EEG) is inexpensive, portable, and temporally precise, yet offers limited spatial resolution. Whole-brain EEG-to-fMRI translation could thus enable accessible and detailed mapping of brain activity. Such translation requires both temporal and spatial fidelity: reconstructing the temporal dynamics of individual brain regions of interest (ROIs), while preserving the spatial structure of correlations between them as captured by their functional connectivity (FC). fMRI signals are composed of a global component, shared across all brain regions, and region-specific fluctuations which deviate from it. Motivated by this structure, we introduce **GLAFE**, a **G**lobal-**L**ocal **A**utoregressive framework for predicting whole-brain **f**MRI from **E**EG. By modeling these components separately, our approach is able to reconstruct temporal dynamics while preserving accurate FC structure. To predict the local deviations, GLAFE learns ROI-specific embeddings that query EEG features through cross-attention. It then refines these predictions through an autoregressive component based on a self-supervised fMRI foundation model, using its earlier whole-brain fMRI reconstructions. We evaluate our method across multiple datasets and subject splits, and demonstrate its generalization capabilities to unseen datasets and tasks in a zero-shot setting. Across these evaluations, GLAFE outperforms competitive baselines in both temporal dynamics and FC reconstruction metrics in almost all configurations.

1 INTRODUCTION

Functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) capture complementary aspects of brain activity. fMRI serves as a central tool for studying human brain function. Through Blood-Oxygen-Level-Dependent (BOLD) signals (where changes in blood oxygenation reflect neural activity), it provides spatially rich measurements of whole-brain activity. However, it is costly, immobile and difficult to acquire at scale, restricting its accessibility . Moreover, it has very limited temporal resolution.

In contrast, EEG is inexpensive, portable, and exhibits high temporal resolution, yet yields noisy signals with limited spatial resolution. Its scalp-level recordings provide only an indirect and spatially coarse view of neural activity. EEG reflects electromagnetic field potentials measured at the scalp, where signals from multiple neural sources are mixed by volume conduction and corrupted by physiological and environmental noise. Moreover, signals originating from activity of deeper brain structures are generally harder to detect and localize, which is particularly limiting for neurological and psychiatric conditions involving subcortical dysfunction.

A central goal is therefore to predict fMRI from EEG, i.e., to infer spatially detailed measurements of whole-brain activity from accessible EEG recordings. More specifically, we aim to predict the mean fMRI BOLD signal within brain regions of interest (ROIs). Effective EEG-to-fMRI translation requires both temporal and spatial fidelity: (i) reconstructing the *temporal* fMRI dynamics of individual brain ROIs, while (ii) faithfully preserving the *spatial* structure of correlations between their time series, as captured by their functional connectivity (FC). However, EEG-to-fMRI translation is a particularly challenging learning problem, as the two modalities fundamentally differ in their measurement mechanisms and thus in their spatial and temporal resolutions. Specifically, this requires

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

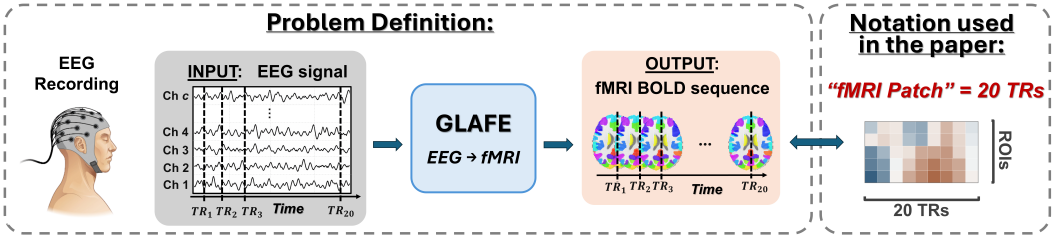

Figure 1: **Problem definition.** *We aim to predict whole-brain fMRI activity (mean BOLD signal within brain-ROIs) from EEG recordings.* **Notation in the paper:** *We define* ***"Patch-fMRI"*** *to be 20 consecutive fMRI time points (TRs), organized in a matrix where each row captures the BOLD signal of a single ROI over 20 TRs.*

inferring ROI-specific signals throughout the whole brain from spatially mixed scalp recordings, while accounting for the hemodynamic delays and temporal smoothing inherent in the BOLD signal.

Much of the earlier work on BOLD time-course prediction from EEG relied on training a separate model for each individual brain ROI (Li et al., 2024; 2025). Single-ROI prediction methods could not exploit any cross-ROI interaction. More recently, UnEBOLT (Li et al., 2026a) and UniEFS (Li et al., 2026b) have introduced a unified whole-brain prediction, along with evaluations of cross-ROI functional connectivity and the ROI dynamics (see more elaborate discussion in Sec. 2). While these prior works have significantly advanced the frontier of whole-brain fMRI prediction from EEG, they did not explicitly utilize the auto-correlations and dominant global fluctuation that are prominent in the BOLD signal, which constitute a core element in the motivations behind our approach.

We introduce **GLAFE**, a **G**lobal-**L**ocal **A**utoregressive framework for predicting whole-brain **f**MRI from **E**EG. Our method builds on top of the following fundamental observation: BOLD measurements of brain activity contain a dominant shared *global* component – the mean activation of all brain-voxels, which is commonly known as the "global signal" in the context of fMRI analysis. The global signal has been shown to correlate strongly with the dominant mode of whole-brain activity, represented by its first principal component (Carbonell et al., 2011). Subtracting this global component from the BOLD time course of individual ROIs, leaves residual signals that describe local deviations from the shared activity, revealing region- and network-specific structure that global fluctuations can obscure in the original fMRI signal (Carbonell et al., 2011; Liu et al., 2017). Across all the datasets examined in this work, these residual signals were found to exhibit lower signal-to-noise ratio and higher frequencies compared to the global signal, making them harder to reconstruct accurately. Yet, both the temporal dynamics of each ROI and especially whole-brain FC are highly dependent on the activity patterns of these local deviations. Therefore, recovering them faithfully is a key aspect of successfully predicting fMRI signals from EEG.

Motivated by this structure, GLAFE separately models the global and local components of the fMRI signal using separate EEG features, architectural pathways, and optimization regimes. To capture the local fluctuations of each ROI, it learns *ROI-specific embeddings* that query EEG features through specially designed *fMRI-EEG cross-attention* blocks. This allows each ROI to learn to attend to the relevant information for its fMRI dynamics within the EEG signals (i.e., critical electrodes, unique time points, and relevant EEG features). Apart from improving the accuracy of the fMRI signals prediction, this cross-attention mechanism between the representations of the two modalities further allows to localize which are the most critical parts within the EEG signal for the BOLD signal prediction of each brain region – i.e., which EEG electrodes are most important to each ROI, and their temporal support (see Sec. 5.3). These mappings may offer suggestive insights into how activity in different brain regions is reflected in scalp electrical potentials. Moreover, such findings may have important practical implications – e.g., if only a few EEG electrodes are available, where is optimal to place them for the prediction of a few specific ROIs of interest?

Finally, the predictions of the local fMRI deviations are refined through an autoregressive correction mechanism, which leverages a strong self-supervised fMRI foundation model (Caro et al., 2023). This foundation model was trained on a very large corpus of resting-state fMRI data, much larger than the limited *simultaneous* EEG-fMRI available dataset. We modified this foundation model to *autoregressively* predict future whole-brain fMRI from GLAFE's predictions of fMRI at earlier time points, while conditioned on the corresponding EEG data. This enables our autoregressive component to leverage the wealth of large unlabeled fMRI datasets, while exploiting the BOLD signal's temporal persistence and the cross-modality dependencies.

We evaluate GLAFE and compare its performance to prior leading methods on three different datasets: (i) a public resting-state dataset, (ii) an auditory-task dataset, and (iii) a privately collected neurofeedback dataset. We further evaluate GLAFE's prediction accuracy at different levels of granularity of the whole-brain activity parcellation into ROIs. We demonstrate GLAFE's generalization capabilities to *unseen subjects* across multiple subject-disjoint splits within all datasets, and to *unseen datasets and new tasks*, evaluated both in zero-shot settings, and in transfer learning settings. GLAFE achieves strong temporal accuracy and FC reconstruction, outperforming competitive baselines in almost all tested configurations.

Our main contributions are therefore:

- **"GLAFE" – a Whole-brain EEG-to-fMRI prediction framework,** which significantly improves FC and temporal dynamics reconstruction accuracy compared to competitive baselines.
- **Mapping most relevant EEG electrodes to each ROI's BOLD prediction,** emerges from our ROI-EEG cross-attention maps, ***suggesting optimal electrode placement*** for targeted ROI reconstruction.
- A framework for utilizing the inherent power of **Global-Local fMRI decomposition.**
- An autoregressive framework for **exploiting the inherent temporal causality of fMRI signals.**

2 Related Work

Early simultaneous EEG-fMRI work studied the mapping between EEG and BOLD signals, highlighting the important aspects of the two modalities for such mapping, such as the importance of EEG spectral composition (Rosa et al., 2010) and estimating flexible temporal transfer functions (Sato et al., 2010). Following these initial studies, research in this field increasingly focused on developing predictive models to recover regional BOLD activity from EEG alone.

Single-region EEG-to-fMRI prediction methods. The electrical fingerprint (EFP) framework used regularized models of time-frequency EEG features to predict regional BOLD, particularly in deep brain regions such as the amygdala (Meir-Hasson et al., 2014). This approach enabled EEG-only neurofeedback targeting amygdala activity (Keynan et al., 2018), with subsequent EFPs developed for the ventral striatum and right inferior frontal gyrus (Singer et al., 2023; Or-Borichev et al., 2023). Later single-region work introduced more general deep-learning methods applicable to any target ROI (Li et al., 2024). Prominently, NeuroBOLT combined spatiotemporal and multiscale spectral features, demonstrating strong predictive performance of BOLD signal on several ROIs (Li et al., 2025). However, training separate models for individual ROIs limits scalability and does not utilize cross-region dependencies throughout the brain.

Whole-brain EEG-to-fMRI prediction methods. An important advancement in this line of work is the introduction of whole-brain methods for predicting all ROIs' fMRI signals simultaneously. UnEBOLT (Li et al., 2026a) combined shared EEG representations with learnable ROI embeddings, introducing the evaluation of both the BOLD signal of each ROI and whole-brain FC. More recently, UniEFS (Li et al., 2026b) incorporated spatial priors from unlabled fMRI pretraining and conditions on demographic and physiological metadata. These methods have made significant progress in whole-brain fMRI predictions from EEG, and are currently the leading methods in this field. However, their main cross-subject evaluations use only a single fixed data-split across scans and subjects, which does not provide enough statistical validation given the very small size of *simultaneous* EEG–fMRI datasets. Therefore, we evaluate and compare GLAFE to previous methods (for which code is available) on *multiple* dataset splits, in order to perform a more exhaustive evaluation.

3 Method

3.1 Problem Formulation and Global–Local Decomposition

Whole-brain fMRI BOLD time series at the ROI level are predicted from simultaneous EEG recordings, as illustrated in Figure 1. Each ROI's fMRI time series is obtained by averaging its voxel signals at each time point, as discussed in Sec. 4.1). Let $X \in \mathbb{R}^{C\times T_e}$ denote an EEG recording with C electrodes and T_e temporal samples, and let $Y \in \mathbb{R}^{N\times T_f}$ contain the corresponding fMRI time series for N ROIs and T_f time points. The input EEG signals and the target fMRI time-series are aligned using the recorded fMRI acquisition times. For each fMRI time point t, we extract the immediately preceding EEG window $X_t \in \mathbb{R}^{C\times w}$, where w is the window length in temporal samples. This

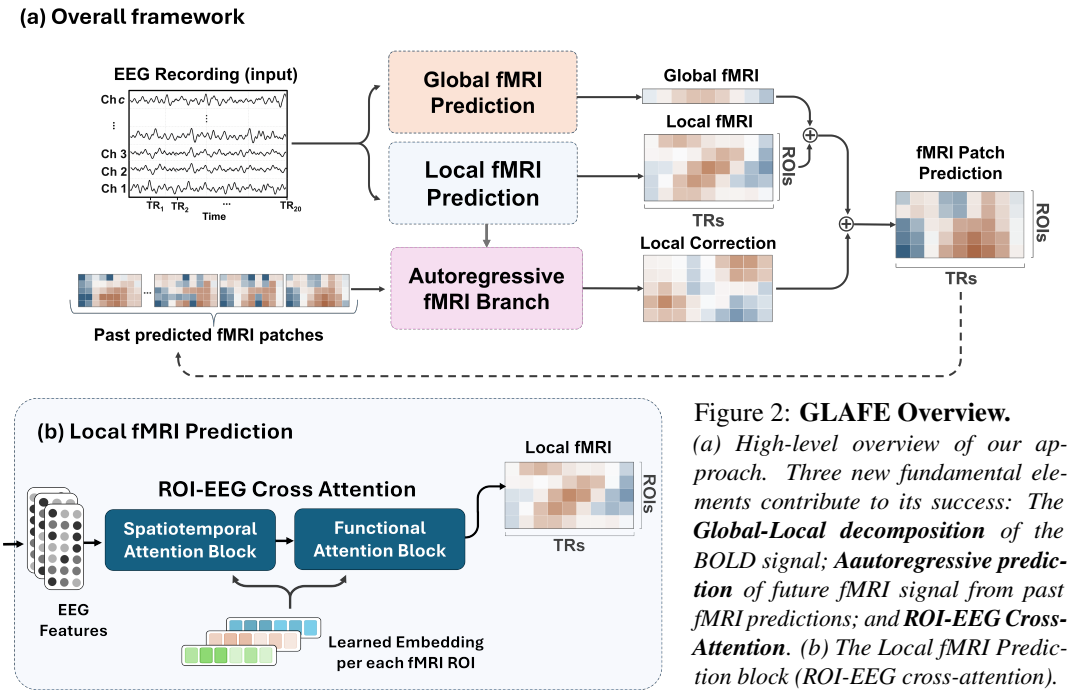

Figure 2: **GLAFE Overview.** *(a) High-level overview of our approach. Three new fundamental elements contribute to its success: The **Global-Local decomposition** of the BOLD signal; **Aautoregressive prediction** of future fMRI signal from past fMRI predictions; and **ROI-EEG Cross-Attention**. (b) The Local fMRI Prediction block (ROI-EEG cross-attention).*

length is a tunable parameter guided by the typical timescale of the delayed hemodynamic response. We selected a duration of 16 seconds based on reconstruction performance, consistent with prior work (Li et al., 2025; 2026a).Each EEG window is paired with a target vector $Y_t \in \mathbb{R}^N$, containing the fMRI signal values of all ROIs at time point t. We organize prediction into non-overlapping patches of 20 consecutive fMRI time points (fMRI patches, see visualization in Fig. 1) , each paired with its corresponding sequence of 20 overlapping EEG windows. This patch structure supports the autoregressive generation and the computation of functional connectivity over the contiguous patch, enabling FC-based losses (see Sections 3.3 and 3.4 for details).

Global and local fMRI signals. We define the global fMRI signal at each time point as the mean fMRI signal across all ROIs, and the local fMRI signal of each ROI as its deviation from this mean:

$$Y_{i,t} = G_t + D_{i,t}, \qquad G_t = \frac{1}{N}\sum_{i=1}^{N} Y_{i,t}. \tag{1}$$

The global component describes a shared temporal trajectory, whereas the local deviations describe how each brain-region fluctuates around it. Predicting the global signal alone would assign the same time course to every ROI. Recovering the deviations is therefore necessary to distinguish regional dynamics and their connectivity patterns. We evaluate all temporal and FC-based metrics on the reconstructed full signal, which contains both components.

Model overview. We reconstruct the whole-brain fMRI signals for each ROI i at each time point t by combining three additive predictions: $\widehat{Y}_{i,t} = \widehat{G}_t + \widehat{D}^{\mathrm{EEG}}_{i,t} + \Delta D^{\mathrm{AR}}_{i,t}$, where $\widehat{G}_t$ represents the predicted global signal shared across all ROIs, $\widehat{D}^{\mathrm{EEG}}_{i,t}$ denotes the local deviations of each ROI from it, and $\Delta D^{\mathrm{AR}}_{i,t}$ denotes the autoregressive correction to these local deviations. The local and correction terms are predicted from the input EEG window and the past fMRI context, respectively. After prediction, we subtract the mean across ROIs at each time point from the predicted local deviations and corrections, ensuring that both predicted signals have zero spatial mean. Consequently, the spatial mean of the final prediction is exactly $\widehat{G}_t$.

3.2 Prediction of Global and Local FMRI Signals from EEG

EEG Feature Extraction. We use REVE (Ouahidi et al., 2025), a self-supervised EEG foundation model, to encode each window X_t into features. The signal of individual electrodes is divided into 1-second overlapping temporal segments, producing a feature vector for each electrode-time

position within the window. REVE's pretrained backbone remains frozen, while two separate Low-Rank Adaptation (LoRA) modules (Hu et al., 2021) tune its representations for the global and local prediction pathways independently. We extract the features from three intermediate layers of REVE to capture different levels of representation. For each EEG window X_t, the resulting feature tensors for each pathway, $F_t^{\mathrm{G}}, F_t^{\mathrm{L}} \in \mathbb{R}^{L \times A \times d_F}$ represent features from L REVE layers, with a d_F-dimensional feature vector at each of the A electrode-time positions.

Global fMRI Signal Prediction. We average F_t^{G} across layers and electrode-time positions to obtain a single representation of the EEG windows. The pooled representation is passed through an MLP head to predict the global fMRI signal, $\widehat{G}_t$.

Local fMRI Signal Prediction via ROI-EEG Cross Attention. To extract the relevant information from the EEG for each region's local signal, GLAFE learns an embedding $e_i \in \mathbb{R}^d$ for each ROI. This embedding queries the local pathway's EEG features through two successive cross-attention blocks. The *spatiotemporal (ST) block* learns which electrode–time information is relevant to each ROI in the current EEG window. It follows standard multi-head cross-attention, using ROI embeddings as queries and local EEG features F_t^{L} as both keys and values, while learned electrode and temporal embeddings are added to the keys. Softmax attention aggregates information across electrode-time positions separately for each REVE feature layer. An output projection and residual MLP then produce ROI-specific features $F_t^{\mathrm{ST}} \in \mathbb{R}^{N \times L \times d_F}$.

The *functional block* modifies this multi-head cross attention mechanism to learn general (input-independent) preferences of each ROI for different types of EEG features. After a linear projection, the ROI embeddings query learned layer–feature embeddings which serve as keys, producing attention coefficients over EEG feature types. Omitting softmax allows signed attention coefficients to weight the value vectors. These vectors are produced from expanding the scalar ST block output for each layer–feature combination using a low-rank linear transformation. Aggregating the weighted value vectors across the layer–feature positions and concatenating the heads produces $F_t^{\mathrm{F}} \in \mathbb{R}^{N \times d}$, where d is the model's hidden dimension. Finally, a shared linear head maps each row $h_{t,i}$ of F to a scalar, and subtracting the mean across ROIs yields the local EEG-based prediction $\widehat{D}_{i,t}^{\mathrm{EEG}}$. A detailed view of the architecture is provided in Fig. 5 in the Appendix.

3.3 Autoregressive Correction of Local fMRI Signals

To address the scarcity of paired EEG–fMRI data, we use BrainLM (Caro et al., 2023), an fMRI foundation model pretrained through self-supervision on a much larger unpaired fMRI corpus. Originally trained to reconstruct randomly masked fMRI patches, BrainLM is adapted here for autoregressive refinement. Its encoder receives a window of up to 10 consecutive fMRI patches, with the final target patch masked and preceding patches visible as context. Fine-tuning with this causal mask adapts BrainLM's representations for future fMRI prediction, exploiting temporal correlations within ROIs and dependencies across regions. To condition these representations on EEG, we combine each ROI's target-patch decoder representation with its local EEG features $h_{t,i}$ averaged over the corresponding windows. An MLP jointly predicts local-signal corrections for all target-patch time points from this combined representation. These corrections compensate for EEG prediction errors, with the corrected local signal supervised against the ground truth (Sec. 3.4). At inference, GLAFE predicts the first patch from EEG alone and uses its preceding fMRI predictions as autoregressive context for subsequent patches.

3.4 Two-Stage Training

Our training pipeline consists of two main stages: (i) EEG-based prediction: which establishes the global and local predictions directly from EEG, and (ii) autoregressive local refinement, which refines the local dynamics and FC while preserving the learned global signal. *Stage 1* trains both EEG pathways with autoregressive correction disabled, supervising both the full prediction and its local component. At the end of this stage, a checkpoint is selected based on the validation temporal R of both the global and local signal prediction, which becomes the starting point of Stage 2. In *Stage 2* the global is frozen to preserve the learned temporal dynamics, and both the local pathway and the autoregression mechanism are trained. FC-based and spatial and correlation losses supervise the local prediction and its autoregressive correction separately. For the final model selection, we maximize full-signal validation FC correlation among Stage-2 checkpoints that retain near-maximal temporal

correlation. Cross-dataset fine-tuning uses only the Stage-1 training procedure, with autoregressive correction disabled, and selects the final checkpoint using the same full-signal criteria as in end-to-end training. Detailed objectives, optimization settings, and selection rules are provided in Appendix B.

4 Experimental Setup

4.1 Datasets

We evaluate GLAFE on three different datasets of simultaneously recorded EEG and fMRI: (i) a public resting-state dataset (Li et al., 2025), (ii) an auditory-task dataset (Li et al., 2025), and (iii) a privately collected neurofeedback task dataset. These are described below, and the full dataset-specific cleaning and preprocessing details are provided in Appendix A.

NeuroBOLT Resting-State Dataset. We use NeuroBOLT's simultaneous resting-state EEG-fMRI dataset (Li et al., 2025): 29 scans from 22 subjects after artifact-based exclusion, acquired with eyes closed using a 3T scanner (TR = 2.1 s) and 32-channel EEG. Each of three splits contains 18:5:6 train:validation:test scans. One of these is the split used in NeuroBOLT and UnEBOLT (Li et al., 2025; 2026a), and two were randomly generated.

NeuroBOLT Auditory-Task Dataset. The auditory-task dataset comprises 16 simultaneous EEG-fMRI scans from 10 subjects, acquired at a different site using a 3T Siemens Prisma scanner (TR = 2.1 s) and 32-channel EEG with a slightly different electrode configuration (Li et al., 2025). Participants responded to auditory tones with right-handed button presses. All three splits were randomly generated, each with 9:3:4 train:validation:test scans. At 128 ROIs, only 13 scans were available (6:3:4); only training counts decrease, with validation and test counts unchanged.

Privately Collected Neurofeedback Data. We collected 109 simultaneous EEG-fMRI scans from 40 healthy subjects during neurofeedback targeting regulation of hippocampal activity, with up to three sessions per subject. MRI was acquired on 3T scanners (TR = 2.0 s) with T2*-weighted gradient-echo EPI, alongside 30-channel EEG at 5 kHz. All three splits were randomly generated, each with 76:16:16 train:validation:test scans.

Across datasets, we apply the same EEG and fMRI filtering and normalization procedures (Appendix A). Using recorded fMRI acquisition times, we align the modalities and segment them into paired EEG windows and whole-brain fMRI frames (Sec. 3.1). We evaluate each dataset using DiFuMo (Dadi et al., 2020) parcellations of 64, 128, and 256 ROIs. For each parcellation, averaging all voxel signals within each ROI at every time point yields N ROI time series per scan. In each split, all scans from a subject remain in one partition to prevent leakage between training and evaluation.

4.2 Baselines and Evaluation Metrics

We compare our model against 4 baselines. Two of them are EEG-to-fMRI translation models-NeuroBOLT (Li et al., 2025) and UneBOLT (Li et al., 2026a). The former predicts BOLD signal of each ROI independently, whereas the latter jointly predicts whole-brain activity and represents a state-of-the-art approach to EEG-to-fMRI translation. We further compare against two EEG foundation models: REVE (Ouahidi et al., 2025) and BIOT (Yang et al., 2023), which we adapted to predict fMRI BOLD time-series from EEG (following each model's protocol for finetuning). All baselines were evaluated using the same dataset splits, ROI granularities, and evaluation metrics as our method.

We note that UniEFS (Li et al., 2026b), a recently published method for whole-brain EEG-to-fMRI prediction, reports in their paper stronger results compared to previous baselines. Unfortunately, we could not compare to them, as their code was not made available at the time of the ICLR submission. Moreover, a comparison to the results reported in their paper could not be done either, for two reasons: (i) Their full training data (which is a significantly larger super-set of the publicly available NeuroBOLT Resting-State Dataset) was not made available (at least not by the time of submission), and (ii) UniEFS is trained also with population *metadata* (age, sex, etc.), providing additional training information compared to training on EEG data alone.

We evaluate predictions using both temporal and structural metrics. For evaluating the temporal fidelity of the predicted fMRI scans, we compare the predicted and ground-truth BOLD time series for each ROI and report the Pearson correlation (R) between them, averaged over ROIs and scans. To evaluate whether the predicted scans preserve the cross-region correlation structure, we also compute

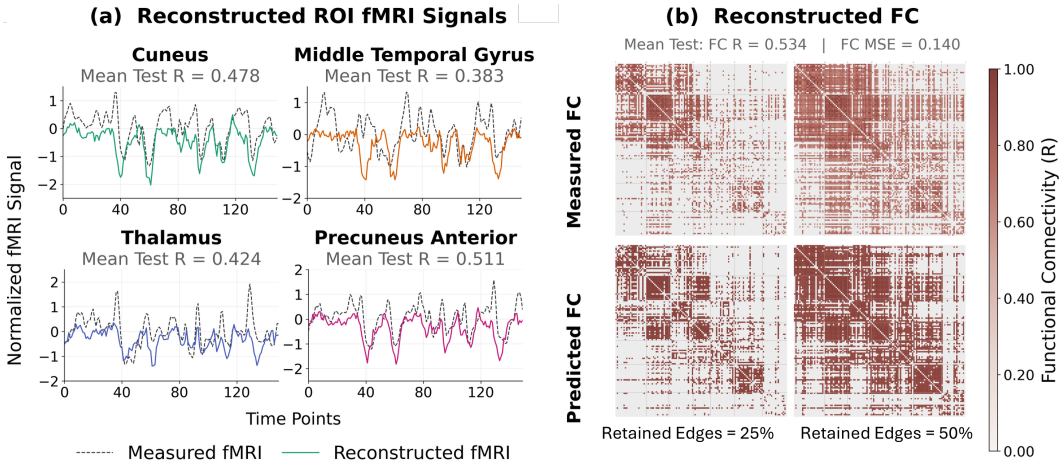

Figure 3: **Examples of reconstructed ROI time-series & FC matrices.** *(a) Sample segments from a reconstructed fMRI scan of a held-out subject, showing 4 sample ROIs against their ground truth signals. The reported mean R per ROI is averaged across all scans in the test partition. (b) Examples of reconstructed whole-brain FC matrices (128 ROIs) from 2 sampled scans in the test partition, against their measured ground truth. For visual clarity, we omit entries with the lowest cross-ROI correlation in both predicted and ground truth matrices, showing the 25% and 50% highest entries (Retained Edges) in the left and right matrices, respectively.*

Method	Temporal R ↑			FC R ↑			FC MSE ↓		
	64 ROIs	128 ROIs	256 ROIs	64 ROIs	128 ROIs	256 ROIs	64 ROIs	128 ROIs	256 ROIs
BIOT (Yang et al., 2023)	0.297 ±0.066	0.285 ±0.068	0.264 ±0.078	0.302 ±0.169	0.348 ±0.162	0.495 ±0.123	0.207 ±0.095	0.245 ±0.105	0.295 ±0.095
REVE (Ouahidi et al., 2025)	0.320 ±0.059	0.306 ±0.060	0.273 ±0.080	0.500 ±0.093	0.502 ±0.117	0.554 ±0.095	0.118 ±0.069	0.142 ±0.066	**0.162 ±0.066**
NeuroBOLT (Li et al., 2025)	0.326 ±0.063	0.304 ±0.068	0.287 ±0.078	0.409 ±0.143	0.413 ±0.160	0.537 ±0.128	0.165 ±0.082	0.204 ±0.092	0.207 ±0.068
UneBOLT (Li et al., 2026a)	0.330 ±0.105	0.304 ±0.110	0.278 ±0.099	0.455 ±0.136	0.463 ±0.149	0.551 ±0.123	0.133 ±0.077	0.157 ±0.083	0.193 ±0.073
GLAFE (Ours)	**0.352 ±0.142**	**0.341 ±0.136**	**0.303 ±0.130**	**0.521 ±0.116**	**0.520 ±0.117**	**0.562 ±0.102**	**0.111 ±0.064**	**0.141 ±0.071**	0.169 ±0.064

Table 1: Generalization across Subjects. *EEG-to-fMRI performance on the Resting-State Dataset for unseen subjects evaluated for 3 atlas sizes. Values are mean ± std over the test scans in each split, averaged across three splits.* ***Bold*** *and underline mark the best and 2nd-best results, ranked by mean.*

the functional connectivity (FC) matrices. For each scan, the FC matrix is defined as the Pearson correlation between all pairs of ROI time series. We compute FC matrices for both the predicted and ground-truth scans, and compare them using MSE and Pearson correlation over the upper triangle of the matrices, excluding the unity diagonal.

5 Results and Analysis

5.1 Generalization across subjects

Our first experiment evaluates GLAFE's generalization capabilities across subjects within a single dataset, reconstructing whole-brain fMRI scans of held-out, unseen subjects. To this end, we use the Resting-State EEG-fMRI dataset introduced by NeuroBOLT (Li et al., 2025). The dataset contains only 29 scans from 22 subjects, with six scans in the original test split. As a result, the choice of held-out subjects can substantially influence the performance of the trained models. Therefore, to obtain more statistically reliable results, we extend the single-split evaluations used in prior works (Li et al., 2025; 2026a;b) to a three-split evaluation. One of our three splits is the original split used in these prior works and two were randomly generated. We compare GLAFE to several state-of-the-art baseline methods (detailed in Sec. 4.2), evaluating both ROI-level fMRI time-series prediction and whole-brain FC reconstruction. Table 1 reports the performance of all methods averaged over the same three subjects-splits for atlas sizes of 64, 128, and 256 ROIs. GLAFE achieves *both the highest temporal* Pearson correlation *and the highest FC* Pearson correlation *at every atlas size*, and the lowest FC-MSE on two of them, outperforming all baselines in almost all comparisons. This advantage is also evident in the performance on each of the individual subject-splits (see Tables 6–8 in Appendix).

Ablations study. We performed ablation analysis on the subject-split used in prior work (Li et al., 2025; 2026a) for the 64 ROIs brain atlas, evaluating the effects of removing two key elements of

GLAFE: (i) the Global-Local decomposition (this was obtained by using the local pathway and autoregression correction on the full signal), and (ii) the autoregression mechanism. These resulted in deterioration in the quality of fMRI reconstruction: (i) Removing the Global-Local decomposition decreased temporal R by 0.056 and FC R by 0.042. (ii) Disabling the autoregression decreased temporal R by 0.025 and FC R by 0.009.

5.2 Generalization across datasets and tasks

We further evaluate GLAFE's generalization across datasets and to new cognitive tasks. For this purpose we use the Auditory-Task Dataset and the Neurofeedback Dataset (Sec. 4.1). We first assess generalization to unseen subjects (as described in Sec. 5.1) by training and testing on subject-splits within each dataset (referred to in Table 2 as "From Scratch" training-mode). We then evaluate generalization from the Resting-State Dataset to the two other datasets in both finetuning and zero-shot settings, using a checkpoint trained on a single resting-state subject-split. Similarly to Sec. 5.1, we evaluate GLAFE on three data splits over the new datasets. We compare GLAFE with UnEBOLT (Li et al., 2026a) (the only state-of-the-art whole-brain EEG-to-fMRI method for which there is available code). Table 2 reports performance at 64 ROIs, averaged across the same three subject-splits per target dataset. For evaluations on other brain atlas granularities (128, 256), see Table 5 in the Appendix.

GLAFE outperforms UnEBOLT on the Auditory-Task Dataset, *across all evaluation metrics, training modes and atlas sizes*. On the Neurofeedback Dataset this advantage holds *across all metrics and training modes at 64 ROIs* and persists in most comparisons at the two finer atlas sizes. Overall, GLAFE achieves the best performance in *49 out of the 54 comparisons* across training modes, evaluation metrics, and atlas sizes on the two datasets. It is interesting to note that finetuning improves GLAFE's temporal and FC correlations over training "from scratch" on both datasets, suggesting that resting-state training teaches useful representations across tasks. On the auditory task, even zero-shot prediction achieves similar or higher correlations than training "from scratch", potentially benefiting from the larger resting-state training dataset. These transfer benefits are more consistent for correlations than for FC-MSE, which improves with fine-tuning only on the auditory task, but not with zero-shot.

5.3 ROI-EEG Cross-Attention Patterns

GLAFE uses a cross-attention mechanism between representations of the two modalities (EEG and fMRI) in order to predict the local BOLD signal of each ROI (see Sec. 3.2). Apart from improving the accuracy of fMRI prediction, this cross-attention mechanism also allows to *localize which EEG electrodes are most critical to the recovery of each ROI.* This is explained next. The cross-modal attention scores are obtained from the query-key inner products between the ROI embedding and EEG features. These indicate the most important electrodes and critical time points (1-second segments) within each 16-second EEG window. To visualize and analyze the ROI-specific attention maps, we summarize them in two ways: (i) averaging over temporal segments – this produces the *spatial* attention scores across the EEG electrodes, and (ii) averaging over electrodes produces the *temporal* attention scores. We then average these maps across all scans within a dataset to obtain the mean attention maps of electrodes and times per brain-ROI.

Method	Training Mode	Auditory-Task Dataset			Neurofeedback Dataset		
		Temp R ↑	FC R ↑	FC MSE ↓	Temp R ↑	FC R ↑	FC MSE ↓
UneBOLT (Li et al., 2026a)	From scratch	0.262 ±0.085	0.431 ±0.168	0.179 ±0.070	0.248 ±0.106	0.425 ±0.084	0.147 ±0.086
GLAFE (Ours)	From scratch	**0.293 ±0.109**	**0.516 ±0.119**	**0.113 ±0.037**	**0.257 ±0.113**	**0.487 ±0.088**	**0.102 ±0.056**
UneBOLT (Li et al., 2026a)	Finetuning	0.272 ±0.109	0.448 ±0.178	0.174 ±0.070	0.249 ±0.105	0.405 ±0.088	0.172 ±0.097
GLAFE (Ours)	Finetuning	**0.318 ±0.110**	**0.591 ±0.108**	**0.098 ±0.046**	**0.258 ±0.125**	**0.518 ±0.093**	**0.167 ±0.089**
UneBOLT (Li et al., 2026a)	Zero-Shot	0.252 ±0.113	0.416 ±0.139	0.200 ±0.094	0.162 ±0.109	0.152 ±0.107	0.156 ±0.086
GLAFE (Ours)	Zero-Shot	**0.308 ±0.108**	**0.512 ±0.052**	**0.156 ±0.069**	**0.195 ±0.117**	**0.223 ±0.081**	**0.140 ±0.083**

Table 2: Generalization across Datasets & Tasks. *EEG-to-fMRI performance on the Auditory-Task and the Neurofeedback datasets with 64 ROIs. Means and across-scan standard deviations are averaged over three splits. "Finetuning" and "Zero-Shot" evaluations use the resting-state checkpoints.* ***Bold*** *and underline mark the best and second-best model within each dataset, metric, and training mode, respectively.*

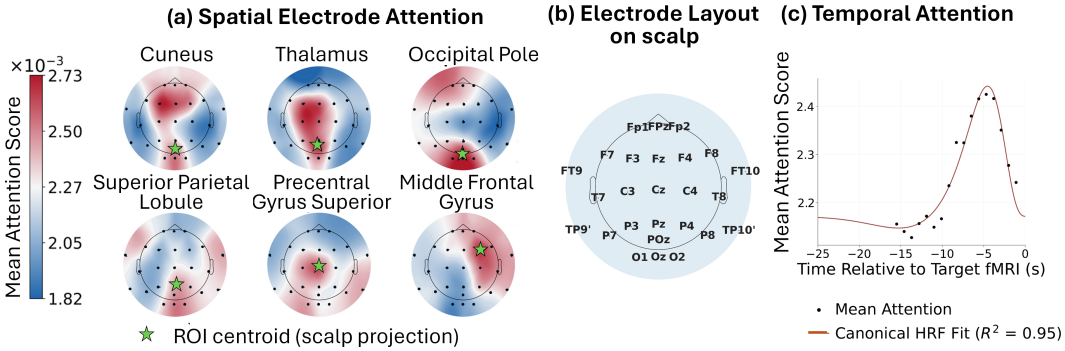

Figure 4: **ROI-EEG Cross-Attention Maps.** *Visualizing the mean Spatio-Temporal Attention maps which emerge from the Resting-State dataset. (a) Spatial electrode attention maps superimposed on the EEG electrode layout, for 6 sample brain-ROIs. The green star marks the projection of each ROI-center on its closest scalp point. (b) Flattered layout of the EEG electrodes on the scalp. (c) The temporal attention map (averaged over all electrodes) shows a distinct temporal pattern, which precedes the corresponding ROI BOLD signal by a few seconds, and resembles almost perfectly (R^2=0.95) the fMRI Hemodynamic Response Function (HRF).*

Fig. 4(a) shows the mean electrode attention maps for several brain-ROIs which emerged from the Resting-State dataset. The projected location of the ROI center to the closest point on the scalp is marked by a green star on each attention map. We observe that for many ROIs, the most predictive electrodes (those residing in high attention score areas) are located near the ROI's closest scalp point (the green star) as intuitively expected, but not in all ROIs. More broadly, these attention maps identify the most important electrode positions for predicting each ROI's fMRI activity. This has potential implications for targeted EEG acquisition: if one has only a few electrodes to record with, or if one is interested in recovering only a few brain-ROIs, GLAFE's attention map could suggest the optimal placement of the EEG electrodes for best prediction of each ROI's BOLD signal. The temporal attention map in Fig. 4(c) show a distinct temporal pattern in the EEG signal, which precedes the corresponding ROI BOLD signal by a few seconds. This recovered temporal pattern is highly consistent with the well known hemodynamic response function (HRF; R^2=0.95), providing evidence to support the accuracy of GLAFE's attention maps. The canonical HRF describes a delayed response between the fMRI signal and the source neural activity. While EEG records electrical activity without any delay, BOLD reflects a slower blood-flow response. Accurate EEG-to-fMRI prediction must therefore account for this distinct temporal delay pattern. This temporal delay is automatically detected by GLAFE's cross-modal attention mechanism, without having to explicitly specify the time delay between the modalities (which might also vary between ROIs).

Discussion. We introduced GLAFE, a whole-brain EEG-to-fMRI framework, which is based on three fundamental new elements: (i) Exploiting the power of Global-Local decomposition of the fMRI BOLD signal; (ii) Autoregressive prediction of future fMRI signal from past fMRI predictions, which exploits the inherent temporal causality in fMRI; and (iii) An ROI-EEG Cross-Attention mechanism. GLAFE outperforms previous competitive methods both in prediction accuracy of the temporal dynamics of individual brain-ROIs, and in reconstruction fidelity of whole-brain functional connectivity (FC) between ROIs. Moreover, evaluations on transfer-learning to new datasets and cognitive tasks demonstrate the generalization capabilities of GLAFE. This is demonstrated both in finetuning settings and in zero-shot settings. Beyond improving prediction accuracy, our ROI-EEG cross-attention reveals attention-maps in the EEG electrode space and ROI-specific HRF-like time delays between the two modalities. Theses findings may have important practical implications on optimal electrode placement for recovery of specific ROIs of interest, or when one has only a few EEG electrodes to record with.

Interesting future extensions include exploring the predictive power of GLAFE's reconstructed fMRI on downstream tasks, such as predicting age, sex, or psychiatric measures. Comparing it to the predictive potential of real measured fMRI, or to that of the input EEG recordings, would assess whether GLAFE's fMRI reconstructions preserve *clinically-relevant information* or even add predictive benefits over EEG alone. Finally, the important potential implication of optimal electrode placement guided by the learned EEG-fMRI attention maps, requires direct validation through electrode-ablation experiments and recordings with reduced electrode sets. Theses are part of our future work.

AI USE STATEMENT

We used generative AI tools to assist with code implementation and to improve the clarity, fluency, and grammar of the manuscript. The authors reviewed all AI-assisted code and text and take full responsibility for the final manuscript.

REPRODUCIBILITY STATEMENT

We provide the details necessary to reproduce GLAFE in the main paper, with complete architecture, training, optimization and model-selection details provided in Appendix B. Dataset preprocessing procedures, experimental splits, and evaluation protocols are described in Section 4 and Appendix A. Upon publication, we will release our full source code and trained model checkpoints. We will also release the model predictions used for evaluation, enabling direct comparison under additional evaluation metrics.

REFERENCES

Felix Carbonell, Pierre Bellec, and Amir Shmuel. Global and System-Specific Resting-State fMRI Fluctuations Are Uncorrelated: Principal Component Analysis Reveals Anti-Correlated Networks. *Brain Connectivity*, 1(6):496–510, December 2011. doi: 10.1089/brain.2011.0065. URL https://doi.org/10.1089/brain.2011.0065.

Josue Ortega Caro, Antonio H. De O. Fonseca, Christopher Averill, Syed A. Rizvi, Matteo Rosati, James L. Cross, Prateek Mittal, Emanuele Zappala, Daniel Levine, Rahul M. Dhodapkar, Insu Han, Amin Karbasi, Chadi G. Abdallah, and David Van Dijk. BrainLM: A foundation model for brain activity recordings, September 2023. URL http://biorxiv.org/lookup/doi/10.1101/2023.09.12.557460.

Kamalaker Dadi, Gaël Varoquaux, Antonia Machlouzarides-Shalit, Krzysztof J. Gorgolewski, Demian Wassermann, Bertrand Thirion, and Arthur Mensch. Fine-grain atlases of functional modes for fMRI analysis. *NeuroImage*, 221:117126, 2020. doi: 10.1016/j.neuroimage.2020.117126. URL https://doi.org/10.1016/j.neuroimage.2020.117126.

Vladimir Fonov, Alan C. Evans, Kelly Botteron, C. Robert Almli, Robert C. McKinstry, D. Louis Collins, and Brain Development Cooperative Group. Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54(1):313–327, 2011. doi: 10.1016/j.neuroimage.2010.07.033. URL https://doi.org/10.1016/j.neuroimage.2010.07.033.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-Rank Adaptation of Large Language Models, October 2021. URL http://arxiv.org/abs/2106.09685. arXiv:2106.09685 [cs].

Jackob N. Keynan, Avihay Cohen, Gilan Jackont, Nili Green, Noam Goldway, Alexander Davidov, Yehudit Meir-Hasson, Gal Raz, Nathan Intrator, Eyal Fruchter, Keren Ginat, Eugene Laska, Marc Cavazza, and Talma Hendler. Electrical fingerprint of the amygdala guides neurofeedback training for stress resilience. *Nature Human Behaviour*, 3(1):63–73, December 2018. ISSN 2397-3374. doi: 10.1038/s41562-018-0484-3. URL https://www.nature.com/articles/s41562-018-0484-3.

Yamin Li, Ange Lou, Ziyuan Xu, Shiyu Wang, and Catie Chang. Leveraging sinusoidal representation networks to predict fMRI signals from EEG, January 2024. URL http://arxiv.org/abs/2311.04234. arXiv:2311.04234 [eess].

Yamin Li, Ange Lou, Ziyuan Xu, Shengchao Zhang, Shiyu Wang, Dario J. Englot, Soheil Kolouri, Daniel Moyer, Roza G. Bayrak, and Catie Chang. NeuroBOLT: Resting-state EEG-to-fMRI Synthesis with Multi-dimensional Feature Mapping, September 2025. URL http://arxiv.org/abs/2410.05341. arXiv:2410.05341 [eess].

Yamin Li, Ange Lou, Chang Li, Shiyu Wang, Haatef Pourmotabbed, Ziyuan Xu, Shengchao Zhang, Dario J Englot, Soheil Kolouri, Daniel Moyer, Roza G Bayrak, and Catie Chang. UnEBOLT: A Unified Model for EEG-to-BOLD Translation and Functional Connectivity Reconstruction. 2026a.

Yamin Li, Shiyu Wang, Chang Li, Ange Lou, Haatef Pourmotabbed, Sarah Goodale, Dario J. Englot, Daniel Moyer, Roza G. Bayrak, and Catie Chang. Mind the State: Towards Unified, Context-Aware EEG-to-fMRI Synthesis. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306 of *Proceedings of Machine Learning Research*. PMLR, 2026b. URL https://soupeeli.github.io/UniEFS/.

Thomas T. Liu, Alican Nalci, and Maryam Falahpour. The global signal in fMRI: Nuisance or Information? *NeuroImage*, 150:213–229, 2017. doi: 10.1016/j.neuroimage.2017.02.036. URL https://doi.org/10.1016/j.neuroimage.2017.02.036.

Yehudit Meir-Hasson, Sivan Kinreich, Ilana Podlipsky, Talma Hendler, and Nathan Intrator. An EEG Finger-Print of fMRI deep regional activation. *NeuroImage*, 102:128–141, November 2014. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.11.004. URL https://linkinghub.elsevier.com/retrieve/pii/S1053811913010963.

Ayelet Or-Borichev, Guy Gurevitch, Ilana Klovatch, Ayam Greental, Yulia Lerner, Dino J. Levy, and Talma Hendler. Neural and functional validation of fMRI-informed EEG model of right inferior frontal gyrus activity. *NeuroImage*, 266:119822, February 2023. ISSN 10538119. doi: 10.1016/j.neuroimage.2022.119822. URL https://linkinghub.elsevier.com/retrieve/pii/S1053811922009430.

Yassine El Ouahidi, Jonathan Lys, Philipp Thölke, Nicolas Farrugia, Bastien Pasdeloup, Vincent Gripon, Karim Jerbi, and Giulia Lioi. REVE: A Foundation Model for EEG – Adapting to Any Setup with Large-Scale Pretraining on 25,000 Subjects, October 2025. URL http://arxiv.org/abs/2510.21585. arXiv:2510.21585 [cs].

M.J. Rosa, J. Kilner, F. Blankenburg, O. Josephs, and W. Penny. Estimating the transfer function from neuronal activity to BOLD using simultaneous EEG-fMRI. *NeuroImage*, 49(2):1496–1509, January 2010. ISSN 10538119. doi: 10.1016/j.neuroimage.2009.09.011. URL https://linkinghub.elsevier.com/retrieve/pii/S1053811909009902.

João R. Sato, Carlo Rondinoni, Marcio Sturzbecher, Draulio B. De Araujo, and Edson Amaro. From EEG to BOLD: Brain mapping and estimating transfer functions in simultaneous EEG-fMRI acquisitions. *NeuroImage*, 50(4):1416–1426, May 2010. ISSN 10538119. doi: 10.1016/j.neuroimage.2010.01.075. URL https://linkinghub.elsevier.com/retrieve/pii/S1053811910000972.

Neomi Singer, Gilad Poker, Netta Dunsky-Moran, Shlomi Nemni, Shira Reznik Balter, Maayan Doron, Travis Baker, Alain Dagher, Robert J Zatorre, and Talma Hendler. Development and validation of an fMRI-informed EEG model of reward-related ventral striatum activation. *NeuroImage*, 276:120183, August 2023. ISSN 10538119. doi: 10.1016/j.neuroimage.2023.120183. URL https://linkinghub.elsevier.com/retrieve/pii/S1053811923003348.

Chaoqi Yang, M. Brandon Westover, and Jimeng Sun. BIOT: Cross-data Biosignal Learning in the Wild, May 2023. URL http://arxiv.org/abs/2305.10351. arXiv:2305.10351 [eess].

A DATASETS PREPROCESSING

NeuroBOLT Resting-State Dataset. The released resting-state data provide ROI time series parcellated into 64 ROIs using DiFuMo, following the NeuroBOLT imaging pipeline (Li et al., 2025). This pipeline includes multi-echo fMRI denoising and nuisance regression. ROI time series were low-pass filtered at 0.15 Hz, mean-centered, and scaled by the per-ROI 95th-percentile absolute amplitude. Before release, EEG was corrected for scanner-gradient and ballistocardiographic (BCG) artifacts. We then applied the EEG preprocessing pipeline of Ouahidi et al. (2025): resampling to 200 Hz, band-pass filtering from 0.5–99.5 Hz, notch filtering at 50 Hz, restriction to EEG channels, session-wise z-scoring, and clipping at ± 15 standard deviations.

NeuroBOLT Auditory-Task Dataset. The auditory-task data followed the same cleaning and preprocessing pipeline as the resting-state data (Li et al., 2025). fMRI preprocessing included multi-echo denoising and nuisance regression; ROI time series were low-pass filtered at 0.15 Hz, mean-centered, and scaled by the per-ROI 95th-percentile absolute amplitude. EEG underwent scanner-gradient and BCG artifact correction. Subsequent EEG preprocessing followed Ouahidi et al. (2025): resampling to 200 Hz, band-pass filtering from 0.5–99.5 Hz, notch filtering at 50 Hz, restriction to EEG channels, session-wise z-scoring, and clipping at ± 15 standard deviations.

Privately Collected Neurofeedback Data. fMRI was processed with fMRIPrep, mapped to MNI152NLin2009cAsym standard space (Fonov et al., 2011) at 2 mm resolution, and denoised with six motion regressors and four polynomial drift terms. Motion regression and ROI standardization followed the corresponding steps in Singer et al. (2023). Standardized ROI signals were low-pass filtered at 0.15 Hz, mean-centered, and scaled by the per-ROI 95th-percentile absolute amplitude, as for the NeuroBOLT datasets. EEG scanner-gradient and cardioballistic artifacts were removed following the corresponding MR-artifact steps of Singer et al. (2023). The first 60 seconds of each scan were cropped because of a technical issue with EEG scanner-gradient artifact removal. Post-artifact EEG preprocessing matched that used for the NeuroBOLT datasets: resampling to 200 Hz, band-pass filtering from 0.5–99.5 Hz, notch filtering at 50 Hz, restriction to EEG channels, session-wise z-scoring, and clipping at ± 15 standard deviations.

B IMPLEMENTATION DETAILS

B.1 ARCHITECTURE SETTINGS AND DIMENSIONS

Table 3 summarizes the architecture using the notation of the Methods section. The tensors F_t^{G} and F_t^{L} contain the retained REVE features after separate linear projections for the global and local pathways. Recording lengths T_e and T_f vary by scan; C depends on the EEG montage, and N on the atlas granularity.

Table 3: Architecture settings and dimensions. Arrows denote input, hidden, and output widths. MLPs use GELU between their linear layers; tensor shapes refer to one EEG window unless stated otherwise.

Parameter	Setting
Inputs and EEG features	
ROIs, N; electrodes, C	$N \in \{64, 128, 256\}$; $C = 26$ for resting state, montage-dependent otherwise
EEG sampling and window length	200 Hz; 16 s, giving $w = 3200$ temporal samples
EEG segment length / overlap	200 / 20 temporal samples (1 / 0.1 s); 17 segments per electrode per window
Electrode–time positions, A	$17C$; 442 for the resting-state montage
fMRI patch length	20 time points, paired with 20 EEG windows
EEG foundation model; retained layers	REVE-large; $L = 3$, using layers 7, 15, and 22 (one-based)
EEG feature projection, d_F	$1216 \to 200$ independently per pathway; $d_F = 200$
Feature tensors F_t^{G}, F_t^{L}	$3 \times A \times 200$
LoRA adapters	Separate global/local adapters; rank 4, LoRA $\alpha = 8$, dropout 0.25
ROI embedding / hidden dimension, d	256
Spatiotemporal (ST) block	
Attention heads; per-head dimensions	4 heads; query/key width 64, value width 50
Electrode and temporal embeddings	Width $d_F = 200$; added to keys only
Output projection; residual MLP	$200 \to 200$; $200 \to 800 \to 200$, separately per retained layer
ROI-specific features F_t^{ST}	$N \times 3 \times 200$
Functional block	
Attention heads; per-head dimensions	4 heads; query/key width 32, value width 64
Layer–feature positions	$Ld_F = 600$ learned keys per head
Low-rank value mapping	Rank 16 per layer and head
Output F_t^{F}; ROI vector $h_{t,i}$	$N \times 256$; 256 features per ROI after head concatenation
Global and local EEG prediction heads	
Global MLP	$200 \to 256 \to 1$, after averaging over layers and electrode–time positions
Shared local linear head	$256 \to 1$, with a learned scalar gain; followed by mean subtraction across ROIs
Autoregressive correction	
fMRI foundation model / decoder features	Pretrained 13M BrainLM; 512 features per ROI at the target patch
Context window	Up to 9 preceding patches plus the masked target; 10 patch positions in total
EEG conditioning and concatenation	Mean of the 20 local EEG vectors $h_{t,i}$ per ROI; $512 + 256 = 768$ combined features
Correction MLP	Input LayerNorm; $768 \to 128 \to 20$, shared across ROIs
Correction initialization	Final-layer weights and biases initialized to zero
Dropout outside LoRA	0 in the ST and functional blocks and the global and correction heads

The pretrained REVE backbone remains frozen, with LoRA applied to its query/key/value, attention-output, and feed-forward projections. The functional block combines each ROI's ST features using signed coefficients and low-rank value maps shared across ROIs. BrainLM supplies decoder features to the correction MLP; its original prediction head is not used.

B.2 Optimization and training objectives

We train Stages 1 and 2 for ten and five epochs, respectively, using AdamW with a learning rate of 10^{-4} and batches of two fMRI patches. Losses are computed independently within each 20-time-point patch and averaged over valid samples; patches in a batch are not concatenated for temporal or FC losses. Each stage completes its prescribed duration before checkpoint selection.

Stage 1: EEG-based prediction. Both EEG pathways are trained with autoregressive correction disabled, including their LoRA adapters, projections, readouts, and shared ROI embeddings. MSE and FC-correlation losses supervise the full EEG prediction $\widehat{Y}^{\mathrm{EEG}} = \widehat{G} + \widehat{D}^{\mathrm{EEG}}$ against Y, and its local component $\widehat{D}^{\mathrm{EEG}}$ against D. This explicit local supervision encourages reconstruction of regional deviations as well as the shared signal.

Stage 2: local refinement. We restore the selected Stage-1 checkpoint and freeze the global pathway and shared EEG parameters, including ROI embeddings. The local EEG pathway continues training, while BrainLM's encoder, decoder, and correction head are fine-tuned. MSE, FC-MSE, spatial-correlation, and FC-correlation losses supervise both EEG-based and corrected local predictions against the true local deviations; temporal-correlation loss additionally supervises the EEG-based prediction. Both the EEG conditioning representation and the EEG prediction used in the corrected-prediction loss are detached without changing their values. Consequently, corrected-prediction losses update only the autoregressive pathway, while the local EEG pathway learns from its own losses.

Combined objective and loss weights. For stage $s \in \{1, 2\}$ and supervised prediction r, the loss is

$$\begin{aligned}\mathcal{L}_{s,r} = {} & \alpha_{s,r}\ell_{\mathrm{MSE}}(\widehat{Z}_r, Z_r) + \beta_{s,r}\ell_{\text{FC-MSE}}(\widehat{Z}_r, Z_r) + \gamma_{s,r}\ell_{\text{FC-corr}}(\widehat{Z}_r, Z_r) \\ & + \delta_{s,r}\ell_{\mathrm{temp}}(\widehat{Z}_r, Z_r) + \eta_{s,r}\ell_{\mathrm{spatial}}(\widehat{Z}_r, Z_r).\end{aligned} \tag{2}$$

Here, r indexes the full and local EEG predictions in Stage 1, and the EEG-based and corrected local predictions in Stage 2; Z_r is the corresponding target. The stage objective sums these contributions using the weights in Table 4, with zero coefficients omitting inactive losses.

Table 4: Loss weights. Stage-1 weights apply separately to full and local EEG predictions. Stage-2 losses supervise local deviations; corrected-prediction losses update only the autoregressive pathway.

	Stage 1	Stage 2	
Loss	Full and local EEG	Local EEG	Corrected local
MSE	1	0.5	1.0
FC-MSE	0	0.025	0.1
Spatial correlation	0	0.025	0.025
FC correlation	0.05	0.025	0.025
Temporal correlation	0	0.025	0

Loss definitions. MSE measures squared prediction errors averaged over valid time points and ROIs. Temporal- and spatial-correlation losses are one minus the mean Pearson correlation over time within each ROI and across ROIs at each time point, respectively. The latter measures the instantaneous regional activity pattern. Constant temporal traces and zero-variance spatial vectors are excluded from the respective correlation averages.

FC losses use the $N(N-1)/2$ strict-upper-triangle entries of the predicted and target patchwise ROI correlation matrices. FC-MSE matches individual connectivity values through their mean squared difference; FC-correlation loss is one minus the Pearson correlation between the two edge vectors,

matching their overall pattern. Stage-2 FC losses act on local deviations, whereas reported FC metrics use the full reconstructed ROI signals.

B.3 AUTOREGRESSIVE CONTEXT DURING TRAINING

Only preceding fMRI patches are visible to BrainLM; the target remains masked and future patches are not exposed. During Stage 2, each entire context window uses fixed reconstructions of the training scan from the selected Stage-1 model with probability 0.3, and measured training fMRI otherwise. The choice is made per window, not per constituent patch, and the reconstructions remain fixed throughout Stage 2. This exposes the correction pathway to both measured histories and prediction errors. Correction losses require preceding context; EEG losses also supervise context-free patches, including the first patch of each scan.

B.4 VALIDATION AND CHECKPOINT SELECTION

Full-scan validation. Selection uses complete validation scans generated as at inference: EEG alone in Stage 1, and earlier model-generated fMRI as autoregressive context in Stage 2. Measured validation fMRI serves only as the evaluation target. Full- and local-signal temporal correlations are computed over each scan, averaged over ROIs, then over scans; global temporal correlation is averaged over scans. Full-signal FC correlation compares the strict upper triangles of predicted and measured full-scan FC matrices and is averaged over scans.

Stage-1 handoff. After ten epochs, we consider checkpoints from epochs three onward, retain those achieving at least 90% of the best eligible validation global temporal correlation, and select the one with the highest local temporal correlation. This checkpoint initializes Stage 2 and supplies its fixed reconstructed training histories.

Final model selection. We consider only the five Stage-2 checkpoints, retain those achieving at least 90% of that stage's best validation full-signal temporal correlation, and select the one with the highest full-signal FC correlation. In both selection steps, all eligible epochs are considered, regardless of when temporal correlation peaks. Ties are resolved by higher global temporal correlation for the handoff or higher full-signal temporal correlation for final selection, then by the earlier epoch.

Stage-1-only fine-tuning. Cross-dataset fine-tuning starts from a resting-state checkpoint and adapts the EEG pathways using the Stage-1 losses and trainable parameters, with autoregressive correction disabled. Final selection retains strong validation full-signal temporal correlation and maximizes full-signal FC correlation among the retained checkpoints.

C ARCHITECTURE VISUALIZATION

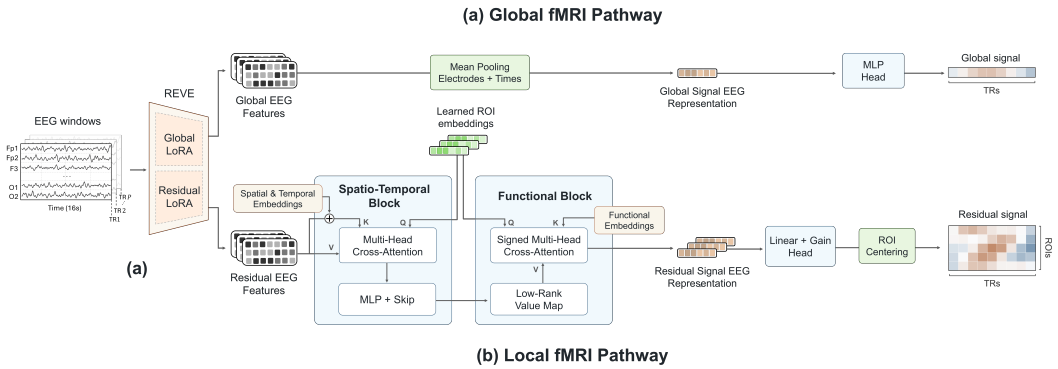

Figure 5: **Architecture of Global and Local Pathways.** (a) Global fMRI pathway. (b) Local fMRI pathway.

D ADDITIONAL RESULTS

128 ROIs

Method	Training Mode	Auditory task			Neurofeedback		
		Temp R ↑	FC R ↑	FC MSE ↓	Temp R ↑	FC R ↑	FC MSE ↓
UneBOLT (Li et al., 2026a)	From scratch	0.255±0.080	0.432±0.086	0.235±0.076	0.230±0.102	0.417±0.092	0.190±0.102
GLAFE (Ours)	From scratch	**0.263±0.111**	**0.515±0.097**	**0.130±0.046**	**0.233±0.116**	**0.451±0.080**	**0.144±0.073**
UneBOLT (Li et al., 2026a)	Fine-tuning	0.278±0.093	0.465±0.103	0.209±0.071	0.229±0.105	0.365±0.110	**0.222±0.117**
GLAFE (Ours)	Fine-tuning	**0.292±0.105**	**0.557±0.086**	**0.130±0.044**	**0.234±0.125**	**0.488±0.094**	0.227±0.105
UneBOLT (Li et al., 2026a)	Zero-shot	0.274±0.104	0.448±0.114	0.193±0.083	0.159±0.098	0.215±0.102	**0.174±0.084**
GLAFE (Ours)	Zero-shot	**0.299±0.105**	**0.494±0.085**	**0.192±0.070**	**0.179±0.113**	**0.226±0.071**	0.205±0.107

256 ROIs

Method	Training Mode	Auditory task			Neurofeedback		
		Temp R ↑	FC R ↑	FC MSE ↓	Temp R ↑	FC R ↑	FC MSE ↓
UneBOLT (Li et al., 2026a)	From scratch	0.239±0.077	0.526±0.074	0.251±0.064	0.194±0.089	0.484±0.120	0.235±0.095
GLAFE (Ours)	From scratch	**0.255±0.094**	**0.531±0.098**	**0.158±0.031**	**0.204±0.103**	**0.518±0.145**	**0.196±0.067**
UneBOLT (Li et al., 2026a)	Fine-tuning	0.251±0.105	0.496±0.115	0.244±0.061	**0.201±0.089**	0.442±0.132	0.280±0.117
GLAFE (Ours)	Fine-tuning	**0.278±0.091**	**0.576±0.073**	**0.170±0.050**	0.200±0.113	**0.535±0.155**	**0.273±0.102**
UneBOLT (Li et al., 2026a)	Zero-shot	0.244±0.105	0.442±0.129	0.244±0.062	0.127±0.084	**0.333±0.123**	**0.248±0.087**
GLAFE (Ours)	Zero-shot	**0.273±0.086**	**0.511±0.066**	**0.233±0.060**	**0.150±0.098**	0.299±0.108	0.322±0.120

Table 5: EEG-to-fMRI performance on the auditory-task and neurofeedback datasets with 128 and 256 ROIs. Means and across-scan standard deviations are averaged over three splits. Fine-tuning and zero-shot evaluation use resting-state checkpoints. Bold and underline mark the best and second-best model within each dataset, metric, and training mode.

Method	Split	Temporal R ↑	FC R ↑	FC MSE ↓
BIOT (Yang et al., 2023)	NeuroBOLT	0.390±0.080	0.415±0.130	0.140±0.084
	42	0.245±0.049	0.268±0.179	0.193±0.119
	45	0.255±0.070	0.222±0.199	0.289±0.083
	Average	0.297±0.066	0.302±0.169	0.207±0.095
REVE (Ouahidi et al., 2025)	NeuroBOLT	0.421±0.059	0.487±0.087	0.084±0.057
	42	0.284±0.062	0.445±0.115	0.135±0.107
	45	0.254±0.055	**0.569±0.078**	**0.134±0.043**
	Average	0.320±0.059	0.500±0.093	0.118±0.069
NeuroBOLT (Li et al., 2025)	NeuroBOLT	0.414±0.069	0.433±0.102	0.121±0.059
	42	0.310±0.060	0.392±0.155	0.159±0.127
	45	0.253±0.060	0.403±0.171	0.215±0.060
	Average	0.326±0.063	0.409±0.143	0.165±0.082
UneBOLT (Li et al., 2026a)	NeuroBOLT	0.411±0.037	**0.546±0.094**	0.090±0.056
	42	0.318±0.084	0.448±0.108	0.136±0.116
	45	0.261±0.195	0.370±0.206	0.174±0.060
	Average	0.330±0.105	0.455±0.136	0.133±0.077
GLAFE (Ours)	NeuroBOLT	**0.449±0.065**	0.519±0.107	**0.075±0.058**
	42	**0.325±0.132**	**0.487±0.131**	**0.121±0.084**
	45	**0.282±0.230**	0.557±0.111	0.137±0.051
	Average	**0.352±0.142**	**0.521±0.116**	**0.111±0.064**

Table 6: Resting-state EEG-to-fMRI performance on the NeuroBOLT dataset with 64 ROIs. Split rows show mean ± source-reported within-split std. Average rows report the arithmetic mean of the three split means and of their reported stds, as in the main results table. Bold and underline mark the best and second-best means across methods within each split or Average row, using unrounded values. NeuroBOLT split denotes the subject spit used in Li et al. (2025; 2026a) and splits 42 and 45 denote the random seed used for their generation.

Method	Split	Temporal R ↑	FC R ↑	FC MSE ↓
BIOT (Yang et al., 2023)	NeuroBOLT	0.376±0.079	0.430±0.083	0.180±0.092
	42	0.236±0.052	0.382±0.205	0.224±0.128
	45	0.243±0.074	0.232±0.198	0.330±0.095
	Average	0.285±0.068	0.348±0.162	0.245±0.105
REVE (Ouahidi et al., 2025)	NeuroBOLT	0.402±0.064	0.501±0.094	0.095±0.055
	42	0.275±0.065	0.469±0.126	0.153±0.094
	45	0.242±0.051	**0.536±0.131**	**0.177±0.048**
	Average	0.306±0.060	0.502±0.117	0.142±0.066
NeuroBOLT (Li et al., 2025)	NeuroBOLT	0.389±0.080	0.447±0.082	0.168±0.059
	42	0.292±0.062	0.432±0.171	0.194±0.137
	45	0.230±0.061	0.361±0.227	0.249±0.079
	Average	0.304±0.068	0.413±0.160	0.204±0.092
UneBOLT (Li et al., 2026a)	NeuroBOLT	0.376±0.053	0.525±0.086	**0.092±0.057**
	42	0.305±0.095	0.471±0.166	0.175±0.124
	45	0.231±0.180	0.394±0.194	0.205±0.068
	Average	0.304±0.110	0.463±0.149	0.157±0.083
GLAFE (Ours)	NeuroBOLT	**0.434±0.070**	**0.552±0.097**	0.094±0.066
	42	**0.314±0.121**	**0.506±0.112**	**0.143±0.086**
	45	**0.274±0.216**	0.502±0.143	0.187±0.061
	Average	**0.341±0.136**	**0.520±0.117**	**0.141±0.071**

Table 7: Resting-state EEG-to-fMRI performance on the NeuroBOLT dataset with 128 ROIs. Split rows show mean ± source-reported within-split std. Average rows report the arithmetic mean of the three split means and of their reported stds, as in the main results table. Bold and underline mark the best and second-best means across methods within each split or Average row, using unrounded values. NeuroBOLT split denotes the subject spit used in Li et al. (2025; 2026a) and splits 42 and 45 denote the random seed used for their generation.

Method	Split	Temporal R ↑	FC R ↑	FC MSE ↓
BIOT (Yang et al., 2023)	NeuroBOLT	0.355±0.095	0.572±0.056	0.233±0.081
	42	0.216±0.060	0.474±0.144	0.271±0.106
	45	0.221±0.079	0.439±0.168	0.380±0.097
	Average	0.264±0.078	0.495±0.123	0.295±0.095
REVE (Ouahidi et al., 2025)	NeuroBOLT	0.343±0.091	0.586±0.089	0.133±0.055
	42	0.261±0.081	0.530±0.082	**0.178±0.075**
	45	0.216±0.068	**0.545±0.114**	**0.176±0.067**
	Average	0.273±0.080	0.554±0.095	**0.162±0.066**
NeuroBOLT (Li et al., 2025)	NeuroBOLT	0.372±0.087	0.597±0.056	0.165±0.047
	42	0.269±0.073	0.544±0.130	0.196±0.094
	45	0.220±0.075	0.471±0.197	0.261±0.063
	Average	0.287±0.078	0.537±0.128	0.207±0.068
UneBOLT (Li et al., 2026a)	NeuroBOLT	0.342±0.036	**0.612±0.071**	0.136±0.055
	42	**0.276±0.091**	0.532±0.133	0.194±0.097
	45	0.217±0.171	0.509±0.165	0.247±0.066
	Average	0.278±0.099	0.551±0.123	0.193±0.073
GLAFE (Ours)	NeuroBOLT	**0.391±0.073**	0.594±0.075	**0.126±0.055**
	42	0.274±0.114	**0.548±0.102**	0.190±0.078
	45	**0.244±0.205**	0.544±0.129	0.193±0.058
	Average	**0.303±0.130**	**0.562±0.102**	0.169±0.064

Table 8: Resting-state EEG-to-fMRI performance on the NeuroBOLT dataset with 256 ROIs. Split rows show mean ± source-reported within-split std. Average rows report the arithmetic mean of the three split means and of their reported stds, as in the main results table. Bold and underline mark the best and second-best means across methods within each split or Average row, using unrounded values. NeuroBOLT split denotes the subject spit used in Li et al. (2025; 2026a) and splits 42 and 45 denote the random seed used for their generation.