Review of Foundation Models for EEG

November 18, 2025

1 Introduction

The development of foundation models for Electroencephalography (EEG) represents a critical evolution in neurotechnology. Leveraging self supervised learning (SSL) paradigms established in Natural Language Processing, these architectures aim to learn generalized neural representations from large scale, heterogeneous datasets. While early attempts like BENDR adapted BERT style encoders to EEG [2], current research focuses on architectural inductive biases required to process high dimensional, low SNR neural time series [1]. This review analyzes the specific structural innovations and pretraining objectives that define the current state of the art.

2 Architectural Innovations

2.1 EEGPT: Hierarchical Encoders and Dual Alignment

EEGPT [3] departs from standard flat attention mechanisms by employing a hierarchical encoder designed to decouple spatial and temporal feature extraction. The architecture utilizes a channel agnostic patch embedding layer that treats electrodes as independent sequences initially, allowing the model to handle varying channel counts (montages) without retraining. Crucially, the model integrates a "Dual Self Supervised" objective. Beyond the standard Masked Autoencoder reconstruction loss popularized in computer vision [4], EEGPT incorporates a Spatio Temporal Representation Alignment module. This module projects the encoder's latent features into a shared space where the distance between the masked view and the unmasked view is minimized via a contrastive loss. This architectural choice forces the model to encode semantic validity rather than simply memorizing stochastic noise patterns found in raw voltage traces [3].

2.2 LaBraM: VQ-VAE Tokenization and Spectrum Prediction

The Large Brain Model (LaBraM) [5] architecture is fundamentally distinguished by its discretization bottleneck. It employs a Vector Quantized Variational Autoencoder as a tokenizer. The input EEG patches are mapped to a codebook of K discrete latent vectors. The encoder backbone is a Vision Transformer variant operating on these discrete indices rather than continuous values. Uniquely, the tokenizer is not trained to reconstruct time domain signals but is optimized via "Neural Spectrum Prediction." The decoder attempts to reconstruct the Fourier spectrum of the input patch. This architectural constraint forces the VQ codebook to capture phase invariant spectral densities (alpha, beta, gamma power) rather than precise temporal waveforms, effectively embedding the FFT logic directly into the model's tokenization layer [5].

2.3 EEG Conformer: Convolutional Attention Fusion

The EEG Conformer [6] proposes a hybrid architecture that sequentially integrates local temporal filtering with global context modeling. The architecture begins with a dedicated Convolution Module comprising two distinct operations:

- **Temporal Convolution:** A 1D convolution with large kernels acting as learnable bandpass filters to extract frequency specific features along the time axis.
- **Spatial Convolution:** A depthwise separable convolution applied across the channel dimension, functionally mimicking Common Spatial Patterns [7] to act as a spatial filter.

These local features are then flattened and projected into a standard Transformer Encoder with Multi Head Self Attention. This design hardcodes the inductive bias of spectral spatial filtering (via CNNs) while leveraging the attention mechanism for long range temporal dependencies [6].

2.4 CBraMod: Factorized Criss Cross Attention

CBraMod [8] addresses the computational cost of full spatio temporal attention ($O((CT)^2)$) by factorizing the attention mechanism. The architecture introduces the Criss Cross Transformer Block, which splits the attention operation into two parallel pathways:

- **Temporal Attention Head:** Computes attention weights only along the time axis for a fixed electrode, capturing dynamic evolution.
- **Spatial Attention Head:** Computes attention weights only across channels for a fixed timestamp, capturing functional connectivity and brain topology.

This factorization reduces complexity while allowing the model to learn distinct representations for temporal dynamics and spatial synchrony, preventing the "smearing" of features often seen in flattened spatio-temporal tokens [8].

3 Signal Characteristics and Tokenization Strategies

3.1 Patching and Sampling Rate Normalization

To enable Transformer architectures, continuous EEG signals must be segmented into patch tokens. A primary challenge is the diverse sampling rates of public datasets (100 Hz to 1000 Hz) [5]. Foundation models typically enforce a unified resolution (e.g., 200 Hz) to define consistent physical time windows for patches. For instance, a patch size of 200ms translates to a fixed vector dimension of $d = 40$ at 200 Hz. Current architectures like EEGPT employ learnable position embeddings that are added to these patches to retain temporal order information after the flattening operation required for the Transformer input [3].

3.2 Frequency Domain Inductive Biases

While time domain reconstruction is intuitive, recent architectures favor frequency domain objectives. Neural signals are naturally characterized by oscillatory bands. LaBraM's architectural shift to reconstruct the Fourier spectrum forces the encoder to learn phase invariant representations. This makes the model robust to the subtle temporal variability inherent in neural recordings that would otherwise penalize a time domain reconstruction loss heavily, despite the semantic content (e.g., Alpha wave presence) being identical [10].

4 Scaling Laws and Generalization

4.1 Parameter Efficiency and Scaling

Empirical studies on EEGPT demonstrate a log linear relationship between downstream accuracy and model parameter count, scaling from 10 million to 100 million parameters [3]. However, unlike text models, EEG models face diminishing returns faster due to the lower information density of the

signal. The hierarchical designs (like EEGPT) and factorized attention (CBraMod) are architectural responses to this, attempting to increase parameter efficiency by injecting domain specific inductive biases rather than relying solely on scale.

4.2 Cross Dataset Transfer Mechanisms

The utility of these architectures lies in transfer learning. Models pretrained on the TUH corpus [9] demonstrate zero shot transfer to BCI tasks (motor imagery). This is facilitated by domain agnostic tokenization. By learning a discrete codebook (LaBraM) or aligned latent space (EEGPT), the transformer backbone models the underlying syntax of brain dynamics, such as the event related desynchronization, independently of the specific hardware sensor noise, provided the channel mapping is handled via flexible embedding layers [5].

5 Conclusion

The progression of EEG analysis towards foundation models marks a fundamental paradigm shift in neurotechnology. This review illustrates that achieving robust generalization requires architectures that go beyond generic sequence modeling to incorporate specific inductive biases aligned with neural dynamics. By embedding spectral understanding and resolving the conflicts between temporal resolution and spatial connectivity, current approaches are successfully mitigating the challenges of high inter subject variability. Moving forward, the convergence of these specialized architectural designs with scalable pretraining regimes offers a viable path toward universal neural decoders.

References

[1] M. Altaf et al., "An in-depth survey on Deep Learning-based Motor Imagery Electroencephalogram (EEG) classification," *Artificial Intelligence in Medicine*, vol. 147, 2024.

[2] D. Kostas et al., "BENDR: using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 34, 2021.

[3] G. Wang et al., "EEGPT: Pretrained Transformer for Universal and Reliable Representation of EEG Signals," *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

[4] K. He et al., "Masked Autoencoders Are Scalable Vision Learners," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[5] W. B. Jiang, L. M. Zhao, and B. L. Lu, "Large Brain Model for Learning Generic Representations with Tremendous EEG Data in BCI," *International Conference on Learning Representations (ICLR)*, 2024.

[6] Y. Song et al., "EEG-Conformer: Efficient Convolutional Transformer for EEG Decoding," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, 2023.

[7] H. Ramoser et al., "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, 2000.

[8] J. Wang et al., "CBraMod: A Criss-Cross Brain Foundation Model for EEG Decoding," *International Conference on Learning Representations (ICLR)*, 2025.

[9] I. Obeid and J. Picone, "The Temple University Hospital EEG Data Corpus," *Frontiers in Neuroscience*, vol. 10, 2016.

[10] K. Gröchenig, *Foundations of Time-Frequency Analysis*. Birkhäuser Boston, 2001.