# LEGEND: A Language-Aligned EEG Foundation Model with Flow Matching Latent Denoising

Yiyu Gui, Mingzhi Chen, and Guibo Luo*

Guangdong Provincial Key Laboratory of Ultra High Definition Immersive Media Technology, Shenzhen Graduate School, Peking University, Shenzhen, China
luogb@pku.edu.cn

**Abstract.** Electroencephalography (EEG) foundation models are designed to learn neural representations that generalize across tasks, subjects, and recording conditions, reducing reliance on task-specific designs and improving scalability in real-world applications. However, existing approaches face two major challenges: limited semantic modeling, which restricts generalization to unseen datasets, and overreliance on masked reconstruction objectives that emphasize local details while neglecting global structure. To address these issues, we propose **LEGEND**, a **L**anguage-aligned E**EG** foundation model with flow matching lat**EN**t **D**enoising. Our framework integrates two complementary modules: Signal–Language Alignment (SLA), which reformulates EEG decoding as instruction-based generation under natural language prompts, and Flow Matching Latent Denoising (FMLD), which promotes structural abstraction through latent-space denoising. We pretrain LEGEND on 8 public EEG datasets and evaluate it on 6 datasets for full fine-tuning, as well as on 2 unseen datasets for zero-shot classification. Results show that LEGEND achieves 15 top-1 full fine-tuning results out of 18 evaluations across 7 baselines and 3 metrics while maintaining strong zero-shot performance, demonstrating its effectiveness as a general-purpose EEG foundation model. The code is available at https://github.com/5GYYYYY/LEGEND.

**Keywords:** EEG · Foundation Model · Language Alignment.

## 1 Introduction

Electroencephalography (EEG), as a non-invasive technique for recording neural activity, has been widely applied in the detection of neurological and psychiatric disorders. With the rapid growth of EEG datasets across multiple centers and tasks, increasing attention has been given to building foundation models for EEG that generalize across varying subjects, tasks, and recording configurations. These models aim to unify diverse neural decoding problems into a shared modeling framework and improve performance consistency across applications.

* Corresponding author.

However, existing methods still face two fundamental challenges that limit their broader applicability.

The first challenge lies in the limited semantic modeling ability of current EEG foundation models. As a result, many of them remain dependent on task-specific fine-tuning and struggle to generalize to unseen datasets. Several representative models reflect this limitation. BIOT [30] augments signals with spatiotemporal perturbations to construct contrastive views for pretraining. LaBraM [14] quantizes channel-wise EEG patches to learn discrete representations. CBraMod [29] employs a criss-cross Transformer with asymmetric positional encoding to strengthen spatiotemporal feature extraction. While each model demonstrates solid performance in its respective task through supervised fine-tuning, they all operate exclusively in the signal domain and lack mechanisms for semantic alignment. As a result, they are not capable of performing zero-shot inference on previously unseen datasets. SSDDB [5] recently introduces semantic-structural pretraining for brain signals, but it still focuses on discriminative signal-text alignment rather than generative instruction-based decoding.

The second challenge stems from the limitations of the masked reconstruction paradigm, widely adopted in EEG representation learning and popularized by MAE-style frameworks [13]. This approach typically reconstructs raw time-domain amplitudes or spectral magnitudes at the sampling-point level, as in Brant [31], CBraMod, and CSBrain. However, such point-wise reconstruction objectives primarily emphasize local signal details while overlooking the broader structural characteristics of EEG. As a result, such models often overfit noise and fail to learn global, invariant features critical for robust generalization.

To address these limitations, we propose LEGEND, a language-aligned EEG foundation model with flow matching latent denoising that integrates semantic alignment and structural learning. **(1) On the semantic side**, we introduce the **Signal–Language Alignment (SLA) module**, which reformulates diverse EEG tasks into a unified instruction-following framework inspired by LLaVA-style [1,17,19] multimodal instruction tuning. Leveraging autoregressive instruction-following prompts, SLA aligns EEG signals with high-level semantic concepts, enabling zero-shot inference on unseen datasets without task-specific adaptation. **(2) On the structural side**, we introduce the **Flow Matching Latent Denoising (FMLD) module**, a latent-space perturbation mechanism inspired by Flow Matching [18,20]. This approach adds structured noise to learned embeddings and trains the model to recover them via a globally consistent vector field. Unlike raw-signal reconstruction, FMLD denoises perturbed latent embeddings to promote smooth, invariant latent manifolds, emphasizing stable neural dynamics over waveform details. We pretrain LEGEND on 8 public EEG datasets and evaluate it under full fine-tuning on 6 datasets, as well as in zero-shot and one-shot settings on 2 unseen datasets. LEGEND achieves 15 top-1 rankings in 18 evaluations across 7 baselines and 3 metrics, while retaining strong zero-shot performance, supporting its use as a general-purpose EEG foundation model.

## 2 Methodology

To address the overemphasis on local details and neglect of global semantics in EEG representation learning, we propose LEGEND—a language-aligned EEG foundation model with flow matching latent denoising that integrates structural regularization and semantic supervision through two key components: Flow Matching Latent Denoising (FMLD) and Signal–Language Alignment (SLA). We next describe these two modules, which jointly define the core training objective.

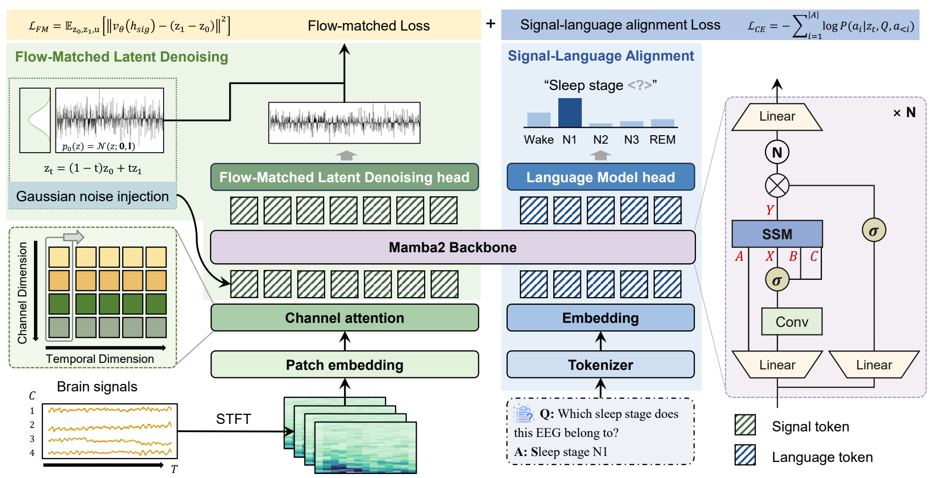


**Fig. 1.** The comprehensive architecture of LEGEND.

### 2.1 Flow Matching Latent Denoising (FMLD)

To encourage the model to capture smooth, invariant, and globally consistent brain dynamics, we introduce Flow Matching Latent Denoising (FMLD) as a generative regularization mechanism. The key idea is to learn a continuous vector field that connects a prior latent distribution with the observed latent representations produced by the signal encoder.

We begin by defining a standard Gaussian prior $p_0(z) = \mathcal{N}(z; \mathbf{0}, \mathbf{I})$ and an empirical distribution $p_1(z)$ derived from the signal encoder output $z_1$. For each training step, we sample a linear interpolation $z_t = (1-t)z_0 + tz_1$ with $t = \sigma(u)$, $u \sim \mathcal{U}(0,1)$, and $\sigma$ denoting the sigmoid function. The target vector field is set as $v_t = z_1 - z_0$.

The signal representation $h_{sig}$ from the Mamba2 backbone is passed to a denoising head $v_\theta$, implemented as a two-layer MLP with SiLU activations and hidden dimension $4D$. This network is trained to predict $v_t$, and the flow matching loss is defined as:

$$\mathcal{L}_{FM} = \mathbb{E}_{\mathrm{z}_0,\mathrm{z}_1,\mathrm{u}}\left[\left\|v_\theta\left(h_{\mathrm{sig}}\right) - \left(\mathrm{z}_1 - \mathrm{z}_0\right)\right\|^2\right] \tag{1}$$

This loss enforces the latent manifold to be continuous and smooth, thus improving the robustness and generalizability of the learned structural representations.

### 2.2 Signal–Language Alignment (SLA)

To enable zero-shot and instruction-based decoding, we introduce the Signal–Language Alignment (SLA) module, which reformulates downstream EEG tasks into a question-answering framework. Given an input signal and a natural language prompt $Q$, the model generates a sequence of tokens $A = (a_1, a_2, ..., a_{|A|})$ autoregressively, conditioned on both the signal representation and the prompt.

For instance, in a sleep stage classification task, the model receives the question “Which sleep stage does this EEG belong to?" and outputs an answer such as “Sleep stage N1". The semantic objective is defined as the cross-entropy loss over the token sequence:

$$\mathcal{L}_{CE} = -\sum_{i=1}^{|A|} \log P\left(a_i \mid \mathrm{z}_\mathrm{t}, Q, a_{<i}\right) \tag{2}$$

By grounding the signal representation in human language, SLA enables the model to align neural activity with interpretable concepts and generalize to unseen tasks without retraining.

### 2.3 Final Objective: Joint Optimization of Structure and Semantics

The overall training loss is a weighted sum of the structural and semantic objectives:

$$\mathcal{L}_{Total} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{FM} \tag{3}$$

The loss trade-off parameter $\lambda$, which controls the balance between semantic alignment and structural regularization, was set to 0.1 as determined to be optimal for training stability and performance through validation.

### 2.4 Mamba2 Backbone

To integrate the dual streams, we adopt Mamba2 [6] as our backbone architecture. Mamba2 is built on state space models (SSMs [11,10]) and offers linear computational complexity, making it particularly suitable for modeling long-range dependencies in neural signals, as also explored in recent EEG modeling studies [12].

Given an input sequence $x(t)$, Mamba2 computes its hidden state $h(t)$ via the continuous-time dynamics $h'(t) = Ah(t) + Bx(t)$. Unlike static SSMs, Mamba2 dynamically generates its parameters from the input, enabling context-aware adaptation and achieving expressivity comparable to masked attention.

In LEGEND, the input to Mamba2 is the concatenation of signal tokens (from the encoder) and task-specific query tokens (representing the semantic stream). Mamba2 performs joint reasoning over these inputs, fusing structural and semantic information in a unified latent space.

### 2.5 Channel-Adaptive Signal Encoder

To accommodate signals with varying channel configurations, we design a channel-adaptive encoder that converts raw EEG signals $x \in \mathbb{R}^{C\times T}$ into a structured time-frequency representation, where $C$ denotes the number of channels and $T$ denotes the number of samples.

Specifically, signals are first transformed using Short-Time Fourier Transform (STFT), followed by logarithmic scaling for dynamic range compression. The resulting spectrogram is partitioned via a 2D convolution layer into fixed-size patches, which are linearly projected into tokens. A multi-head self-attention module processes these tokens along the channel dimension, capturing inter-channel relationships. Finally, we perform global average pooling across channels to obtain a compact representation sequence $z_1 \in \mathbb{R}^{L\times D}$, which serves as the unified signal input for downstream processing.

## 3 Experiments and Results

### 3.1 Datasets

To construct a robust foundation model, we curated a large-scale pre-training corpus from 8 publicly available EEG datasets [2,8,25,22,32,26,4,27]. This corpus encompasses diverse clinical and cognitive neuroscience tasks, including sleep analysis and the diagnosis of various neurological and psychiatric disorders (e.g., Epilepsy, Depression, Alzheimer's). Our pretraining corpus comprises recordings from 8023 subjects, totaling over 11 million signal segments and approximately 63,000 hours of EEG data. Following our generative QA paradigm, we transformed each data sample into a question-answering pair. For instance, a sleep signal might be paired with the question, "Which sleep stage does this signal belong to? A.Wake B.N1 C.N2 D.N3 E.REM" and the corresponding answer, "B.N1".

Table 1 provides the details of downstream datasets. For both pre-training and downstream tasks, we applied a unified preprocessing pipeline. Inspired by LaBraM [14], all signals were resampled to 200 Hz, band-pass filtered between 0.1-70 Hz, and notch filtered at 50 or 60 Hz to mitigate power-line interference.

### 3.2 Downstream Experiment

We designed three distinct evaluation protocols to comprehensively benchmark LEGEND against state-of-the-art models. We compared our method against

**Table 1.** Information of downstream datasets.

| Dataset | #Classes | #Subjects | #Channels | Duration | Raw rate | Description |
|---|---|---|---|---|---|---|
| Dreams [7] | 5 | 20 | 3 | 30 s | 200 Hz | Sleep stage classification |
| ADFSU [21] | 2 | 92 | 19 | 2 s | 128 Hz | Alzheimer's detection |
| Schizo-Youth [9] | 2 | 84 | 16 | 10 s | 128 Hz | Schizophrenia diagnosis |
| SEE [24] | 2 | 30 | 1 | 4 s | 173.61 Hz | Epilepsy detection |
| MDD [23] | 2 | 61 | 19 | 5 s | 256 Hz | Depressive disorder diagnosis |
| ADHD-80 [3] | 2 | 80 | 2 | 5 s | 256 Hz | ADHD diagnosis |

three end-to-end models: EEGNet [16], EEGConformer [28], and SPaRCNet [15]; four pretraining-based methods: BIOT [30], LaBraM [14], CBraMod [29], and CSBrain [33]. Our evaluation metrics include Balanced Accuracy, F1 Score, and AUC to provide a holistic view of performance.

For downstream evaluation, we conducted experiments under three settings: full fine-tuning, one-shot classification, and zero-shot classification. **Full fine-tuning** updates all model parameters. In the **zero-shot** setting, given an EEG signal and a multiple-choice question, LEGEND autoregressively generates the most probable answer token. In the **one-shot** setting, we use one labeled example per class; for LEGEND and pre-trained baselines we perform linear probing with a frozen backbone, while end-to-end models are fully fine-tuned. Notably, since baseline models cannot perform zero-shot inference, our zero-shot and one-shot results are compared against the baselines' one-shot performance.

For robust and unbiased evaluation, we adopted a subject-wise 5-fold cross-validation strategy for five of the datasets. This ensures that the model is always tested on subjects unseen during training, reflecting a realistic clinical generalization challenge. For the SEE dataset, which has a predefined train/test split, we repeated experiments with 5 different random seeds.

### 3.3 Experimental Settings

All experiments were conducted on a Linux server with an NVIDIA A100-80GB GPU, using PyTorch 2.0.1 and CUDA 12.2. During pre-training, we used a batch size of 256, implemented with 8 gradient accumulation steps. We employed the AdamW optimizer with $\beta$=(0.9, 0.999), a weight decay of 0.1, and a peak learning rate of $2\times10^{-4}$, which was managed by a cosine annealing schedule. For downstream fine-tuning tasks, the optimizer configuration remained the same, but we used a constant learning rate of $2 \times 10^{-4}$ and a batch size of 128.

### 3.4 Results

**Full fine-tuning performance.** Table 2 reports the performance comparison between our LEGEND model and existing baselines under the full fine-tuning setting. The best results are highlighted in bold. As shown, LEGEND consistently outperforms all baselines across three metrics on sleep staging, Alzheimer's disease detection, schizophrenia diagnosis, and seizure detection. It also achieves

**Table 2. Full fine-tuning performance comparison. Bold** denotes the best result and underline denotes the second best. For LEGEND, subscripts report the absolute gap relative to the strongest baseline: ↑ denotes the improvement over the second-best model when LEGEND attains the top performance, whereas ↓ denotes the gap to the best-performing model otherwise.

| Dataset | Model | BACC | | F1 Score | | AUC | |
|---|---|---|---|---|---|---|---|
| | | mean | std. | mean | std. | mean | std. |
| Dreams | EEGNet [16] | 54.45 | 5.49 | 59.80 | 9.32 | 88.37 | 1.78 |
| | EEGConformer [28] | 31.74 | 5.22 | 31.47 | 9.98 | 72.24 | 4.34 |
| | SPaRCNet [15] | 34.80 | 8.81 | 20.69 | 6.95 | 75.49 | 3.01 |
| | BIOT [30] | 63.50 | 2.93 | 70.43 | 3.23 | 89.99 | 1.77 |
| | LaBraM [14] | 66.08 | 2.82 | 73.18 | 2.25 | 90.29 | 1.52 |
| | CBraMod [29] | 65.31 | 2.82 | 72.90 | 2.15 | 90.45 | 1.68 |
| | CSBrain [33] | 56.83 | 2.41 | 63.77 | 1.55 | 85.24 | 1.19 |
| | LEGEND | **71.92**$_{\uparrow 5.84}$ | 3.92 | **79.35**$_{\uparrow 6.17}$ | 3.40 | **94.01**$_{\uparrow 3.56}$ | 1.30 |
| ADFSU | EEGNet [16] | 88.72 | 7.93 | 96.81 | 2.19 | 96.32 | 4.19 |
| | EEGConformer [28] | 88.70 | 4.88 | 97.11 | 0.94 | 97.16 | 1.71 |
| | SPaRCNet [15] | 97.08 | 3.54 | 98.49 | 1.73 | 99.08 | 1.65 |
| | BIOT [30] | 88.78 | 5.51 | 96.50 | 1.83 | 95.63 | 4.06 |
| | LaBraM [14] | 75.70 | 3.23 | 95.64 | 1.00 | 89.57 | 2.39 |
| | CBraMod [29] | 86.15 | 6.56 | 96.88 | 1.47 | 94.23 | 5.39 |
| | CSBrain [33] | 87.55 | 4.07 | 97.16 | 1.21 | 92.08 | 3.16 |
| | LEGEND | **97.63**$_{\uparrow 0.55}$ | 1.26 | **98.52**$_{\uparrow 0.03}$ | 0.69 | **99.74**$_{\uparrow 0.66}$ | 0.30 |
| Schizo-Youth | EEGNet [16] | 70.50 | 6.03 | 74.43 | 4.64 | 76.00 | 9.63 |
| | EEGConformer [28] | 78.72 | 6.46 | 79.54 | 6.00 | 84.70 | 8.27 |
| | SPaRCNet [15] | 83.02 | 10.64 | 83.83 | 7.66 | 89.06 | 13.60 |
| | BIOT [30] | 80.65 | 9.39 | 81.66 | 7.79 | 85.57 | 8.75 |
| | LaBraM [14] | 67.80 | 7.48 | 72.97 | 5.08 | 71.63 | 11.43 |
| | CBraMod [29] | 80.70 | 9.83 | 82.95 | 7.07 | 87.96 | 8.63 |
| | CSBrain [33] | 72.99 | 8.39 | 75.82 | 4.49 | 73.90 | 9.35 |
| | LEGEND | **91.60**$_{\uparrow 8.58}$ | 7.10 | **92.15**$_{\uparrow 8.32}$ | 6.50 | **96.30**$_{\uparrow 7.24}$ | 4.32 |
| SEE | EEGNet [16] | 61.50 | 0.57 | 49.09 | 1.72 | 78.66 | 3.59 |
| | EEGConformer [28] | 70.60 | 2.16 | 69.70 | 3.88 | 78.78 | 0.64 |
| | SPaRCNet [15] | 59.94 | 1.79 | 50.69 | 15.66 | 79.94 | 2.37 |
| | BIOT [30] | 70.55 | 1.21 | 68.83 | 2.18 | 77.72 | 0.50 |
| | LaBraM [14] | 68.35 | 2.22 | 68.97 | 1.29 | 74.11 | 2.62 |
| | CBraMod [29] | 70.13 | 0.86 | **70.83** | 2.20 | 76.64 | 0.31 |
| | CSBrain [33] | 68.00 | 1.07 | 67.42 | 1.86 | 72.90 | 1.35 |
| | LEGEND | **73.34**$_{\uparrow 2.74}$ | 1.32 | 70.23$_{\downarrow 0.60}$ | 1.62 | **80.25**$_{\uparrow 0.31}$ | 1.10 |
| MDD | EEGNet [16] | 87.93 | 12.94 | 84.28 | 17.82 | 95.66 | 5.80 |
| | EEGConformer [28] | 88.09 | 8.31 | 85.61 | 11.78 | 95.05 | 4.85 |
| | SPaRCNet [15] | 85.59 | 8.05 | 83.06 | 11.02 | 92.15 | 5.93 |
| | BIOT [30] | 86.35 | 6.48 | 84.10 | 8.20 | **95.84** | 3.67 |
| | LaBraM [14] | 86.10 | 12.17 | 82.30 | 16.57 | 89.96 | 10.79 |
| | CBraMod [29] | 88.99 | 6.32 | 87.22 | 7.84 | 95.72 | 5.52 |
| | CSBrain [33] | 82.23 | 9.10 | 78.64 | 12.57 | 88.86 | 10.25 |
| | LEGEND | **90.60**$_{\uparrow 1.61}$ | 7.78 | **89.11**$_{\uparrow 1.89}$ | 9.66 | 95.42$_{\downarrow 0.42}$ | 6.71 |
| ADHD-80 | EEGNet [16] | 84.69 | 10.48 | 81.36 | 14.58 | 90.18 | 10.04 |
| | EEGConformer [28] | 90.16 | 5.64 | 89.30 | 6.34 | 94.70 | 4.48 |
| | SPaRCNet [15] | 89.99 | 8.34 | 88.23 | 10.71 | 96.75 | 3.64 |
| | BIOT [30] | 92.24 | 4.23 | 91.76 | 4.47 | 97.10 | 2.82 |
| | LaBraM [14] | 83.94 | 4.84 | 82.97 | 5.89 | 90.79 | 4.99 |
| | CBraMod [29] | 93.32 | 2.20 | 92.91 | 2.51 | **98.00** | 1.23 |
| | CSBrain [33] | 89.70 | 5.20 | 88.71 | 6.12 | 93.39 | 3.47 |
| | LEGEND | **93.65**$_{\uparrow 0.33}$ | 2.51 | **93.17**$_{\uparrow 0.26}$ | 2.97 | 96.97$_{\downarrow 1.03}$ | 2.19 |

strong results on depressive disorder and ADHD diagnosis, demonstrating the broad applicability and robust generalization capability of LEGEND.

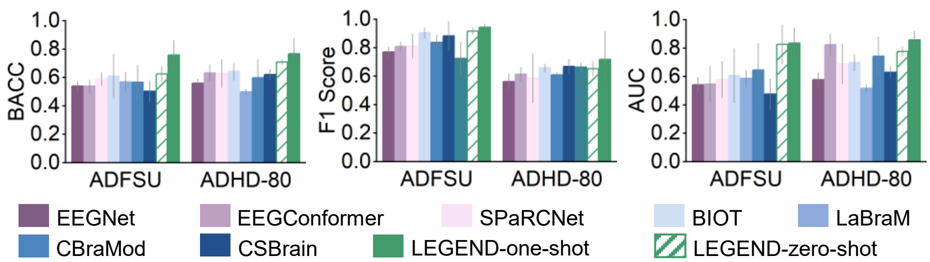


**Fig. 2.** Zero-shot and one-shot performance comparison.

**Zero-shot and one-shot performance.** Fig. 2 illustrates the one-shot and zero-shot classification performance of LEGEND on two representative neurological disorders: Alzheimer's disease (ADFSU) and ADHD (ADHD-80). The experimental setup is detailed in Section 3.2. Notably, LEGEND consistently surpasses all baselines under the one-shot setting. More impressively, its zero-shot performance exceeds that of most baselines in the one-shot scenario, highlighting the effectiveness of semantic integration in enhancing generalization. Moreover, LEGEND's one-shot performance consistently outperforms its zero-shot performance, indicating that the model can efficiently utilize limited label supervision and learn meaningful associations even from a single annotated example.

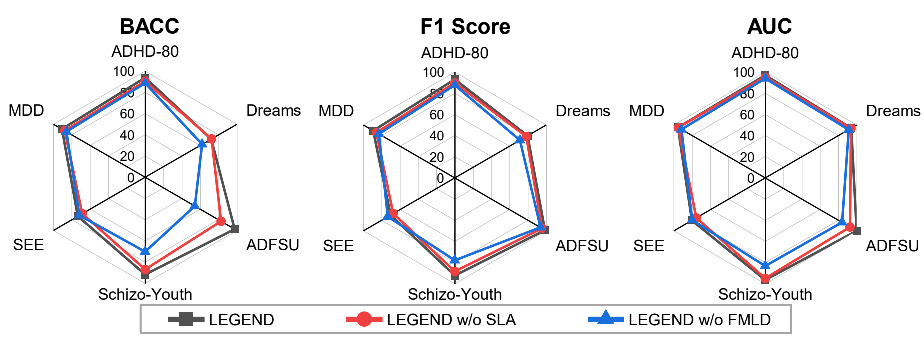


**Fig. 3.** Ablation study.

**Ablation study.** To further assess the contributions of SLA and FMLD, we conducted an ablation study by selectively removing each module, with results summarized in Fig. 3. In most datasets, removing either component led to noticeable performance degradation, although the extent of the effect varied across tasks. These results suggest that the two modules play complementary roles and that the structural regularization mechanism contributes to improving representation stability and generalization.

## 4 Conclusion

In this work, we introduced LEGEND, a language-aligned EEG foundation model with flow matching latent denoising that integrates structural representation learning with semantic alignment. Through the proposed Flow Matching Latent Denoising (FMLD) and Signal–Language Alignment (SLA) modules, our framework captures both the intrinsic dynamics of EEG signals and their task-level semantic abstractions. Extensive experiments across six benchmark datasets demonstrate that LEGEND not only achieves strong performance under full supervision but also exhibits promising zero-shot generalization to previously unseen datasets. These results highlight the potential of combining structure-aware modeling with language-guided learning to build more generalizable and semantically grounded EEG foundation models.

**Acknowledgments.** This work is supported by National Key Research and Development Program of China (2024YFE0203100), Guangdong Provincial Key Laboratory of Ultra High Definition Immersive Media Technology (Grant No.2024B1212010006), and Shenzhen Science and Technology Program (JCYJ20230807120800001).

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Alayrac, J.B., Donahue, J., Luc, P., Miech, A., et al.: Flamingo: a visual language model for few-shot learning. Advances in neural information processing systems **35**, 23716–23736 (2022)
2. Alvarez-Estevez, D., Rijsman, R.: Haaglanden medisch centrum sleep staging database (version 1.1) (Mar 2022). https://doi.org/10.13026/t79q-fr32
3. Bajestani, G.S., Abedian, S., Makhloughi, F., Raoufitabar, M., Saeedi, H.: A dataset of EEG signals from adults with adhd and healthy controls: Resting state, cognitive function, and sound listening paradigm. Mendeley Data (2023)
4. Bhagubai, M., Chatzichristos, C., Swinnen, L., Macea, J., et al.: Seizeit2 (2025). https://doi.org/10.18112/openneuro.ds005873.v1.1.0
5. Chen, M., Gui, Y., Luo, G., Yang, Y.: SSDDB: A semantic-structural dual-drive pretraining framework for brain signals. In: ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2026)
6. Dao, T., Gu, A.: Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality. Proceedings of Machine Learning Research **235**, 10041–10071 (2024)
7. Devuyst, S., Kerkhofs, M., Dutoit, T.: The DREAMS databases and assessment algorithm (may 2019). https://doi.org/10.5281/zenodo.2650142, https://zenodo.org/records/2650142

8. van Dijk, H., van Wingen, G., Denys, D., Olbrich, S., van Ruth, R., Arns, M.: The two decades brainclinics research archive for insights in neurophysiology (TD-BRAIN) database. sci data. 9, 333 (2022)
9. Gorbachevskaya, N., Borisov, S.V.: EEG of healthy adolescents and adolescents with symptoms of schizophrenia (2019)
10. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. In: First Conference on Language Modeling (2023)
11. Gu, A., Goel, K., Re, C.: Efficiently modeling long sequences with structured state spaces. In: International Conference on Learning Representations (2021)
12. Gui, Y., Chen, M., Su, Y., Luo, G., Yang, Y.: EEGMamba: Bidirectional state space models with mixture of experts for EEG multi-task classification. arXiv preprint arXiv:2407.20254 (2024)
13. He, K., Chen, X., Xie, S., Li, Y., et al.: Masked autoencoders are scalable vision learners. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15979–15988. IEEE (2022)
14. Jiang, W., Zhao, L., Lu, B.: Large brain model for learning generic representations with tremendous EEG data in BCI. In: ICLR. pp. 16405–16426 (2024)
15. Jing, J., Ge, W., Hong, S., Fernandes, M.B., et al.: Development of expert-level classification of seizures and rhythmic and periodic patterns during EEG interpretation. Neurology **100**(17), e1750–e1762 (2023)
16. Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., et al.: EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. Journal of neural engineering **15**(5), 056013 (2018)
17. Li, J., Li, D., Savarese, S., Hoi, S.: Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In: International conference on machine learning. pp. 19730–19742. PMLR (2023)
18. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. In: 11th International Conference on Learning Representations, ICLR 2023 (2023)
19. Liu, H., Li, C., Wu, Q., Lee, Y.J.: Visual instruction tuning. Advances in neural information processing systems **36**, 34892–34916 (2023)
20. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. In: The Eleventh International Conference on Learning Representations (ICLR) (2023)
21. Miltiadous, A., Tzimourta, K.D., Afrantou, T., Ioannidis, P., Grigoriadis, N., Tsalikakis, D.G., Angelidis, P., Tsipouras, M.G., Glavas, E., Giannakeas, N., Tzallas, A.T.: A dataset of EEG recordings from: Alzheimer's disease, frontotemporal dementia and healthy subjects (2024). https://doi.org/10.18112/openneuro.ds004504.v1.0.8
22. Miltiadous, A., Tzimourta, K.D., Afrantou, T., Ioannidis, P., Grigoriadis, N., Tsalikakis, D.G., Angelidis, P.A., Tsipouras, M.G., Glavas, E., Giannakeas, N., Tzallas, A.T.: A dataset of scalp EEG recordings of alzheimer's disease, frontotemporal dementia and healthy subjects from routine EEG. Data **8**, 95 (2023), https://api.semanticscholar.org/CorpusID:258952517
23. Mumtaz, W.: MDD patients and healthy controls EEG data (new) (11 2016). https://doi.org/10.6084/m9.figshare.4244171.v2
24. Panwar, S.: Single electrode eeg data of healthy and epileptic patients (2020)
25. Prado, P., Medel, V., Gonzalez-Gomez, R., Sainz-Ballesteros, A., et al.: The brainlat project, a multimodal neuroimaging dataset of neurodegeneration from underrepresented backgrounds. Scientific Data **10**(1), 889 (2023)

26. Shoeb, A.H.: Application of machine learning to epileptic seizure onset detection and treatment. Ph.D. thesis, Massachusetts Institute of Technology (2009)
27. Singh, A., Cole, R., Espinoza, A., Cavanagh, J., Narayanan, N.: Rest eyes open (2023). https://doi.org/10.18112/openneuro.ds004584.v1.0.0
28. Song, Y., Zheng, Q., Liu, B., Gao, X.: EEG conformer: Convolutional transformer for EEG decoding and visualization. IEEE Transactions on Neural Systems and Rehabilitation Engineering **31**, 710–719 (2022)
29. Wang, J., Zhao, S., et al.: CBraMod: A criss-cross brain foundation model for EEG decoding. In: ICLR. pp. 75310–75346 (2025)
30. Yang, C., Westover, M., Sun, J.: BIOT: Biosignal transformer for cross-data learning in the wild. Advances in Neural Information Processing Systems **36**, 78240–78260 (2023)
31. Zhang, D., Yuan, Z., Yang, Y., Chen, J., et al.: Brant: Foundation model for intracranial neural signal. Advances in Neural Information Processing Systems **36**, 26304–26321 (2023)
32. Zhang, G.Q., Cui, L., Mueller, R., Tao, S., et al.: The national sleep research resource: towards a sleep data commons. Journal of the American Medical Informatics Association **25**(10), 1351–1358 (2018)
33. Zhou, Y., Wu, J., et al.: CSBrain: A cross-scale spatiotemporal brain foundation model for EEG decoding. arXiv preprint arXiv:2506.23075 (2025)