



NeuroLM: A Universal Multi-task Foundation Model for Bridging the Gap between Language and EEG Signals

☼ 상태	완료
≡ Journal	ICLR
≡ Year	2024.09
≡ Summary	EEG 데이터를 자연어처럼 변환하여 LLM과 결합하고, 멀티태스크 학습을 통해 다양한 EEG 태스크를 하나의 모델에서 수행할 수 있도록 만든 최초의 EEG 기반 foundation model 개발
≡ Limitations	단일 작업 최적화 모델에 비해 뒤처짐, 하이퍼파라미터 민감도 높음, EEG-text 정렬이 공간적 정렬만 이루어져 한계가 있음
🔗 Link	https://arxiv.org/abs/2409.00101
≡ category	EEG Fine-tuning LLM Pre-training

<https://github.com/935963004/NeuroLM>

Introduction

EEG(뇌전도) 데이터는 뇌-컴퓨터 인터페이스(BCI) 및 의료 분야에서 중요한 역할을 하고 있지만, 낮은 신호 대 잡음비(SNR), 데이터의 비정상성(Nonstationarity), 다양한 데이터 수집 방식 등으로 인해 일반화된 표현을 학습하는 것이 어려움. 본 연구에서는 MLLMs를 활용해 EEG 신호를 기존 LLM에 통합하여 여러 EEG 작업을 통합하는 새로운 방향을 제시하려고 함. 하지만, LLM을 활용해 EEG 패턴을 이해하는 데에는 몇 가지 어려움이 있음:

1. EEG-text 임베딩 정렬 : EEG 세그먼트에서 의미 정보를 추출하기가 어렵기 때문에 EEG-text 쌍 존재하지 않음
2. LLM을 통한 효과적인 representation learning : EEG 신호 추출을 위해 주로 마스킹된 EEG 모델링을 사용함 → LLM에서 일반 정보를 학습하는 방법은 아직 부족함
3. 다양한 EEG task에 대한 통합 : 여러 EEG 작업을 통합하는 것은 작업의 다양성과 특수성 때문에 복잡함

따라서, 본 연구에서는 LaBraM을 기반으로 NeuroLM을 제안함. 25,000시간의 EEG 데이터로 사전 학습하여 여러 EEG 태스크를 하나의 모델로 해결할 수 있도록 설계함. 세 가지 단계로 학습 진행됨: 1) 연속 EEG 신호를 neural codebook에서 불연속 코드로 인코딩하기 위해 벡터 양자화된 시간-주파수 예측을 사용하여 텍스트 정렬 neural tokenizer가 학습됨, 2) neural tokenizer의 VQ 인코더를 고정하여 compact embedding을 추출하고 이를 LLM의 입력으로 사용함. LLM이 인과적 EEG 표현을 학습할 수 있도록 하기 위해, 본 연구는 multi-channel autoregressive pre-training을 제안함. 3) multi-task instruction tuning을 사용하여 비정상 탐지, 이벤트 분류, 감정 인식, 수면 단계 분류, 인지 작업 부하 예측, 뇌파의 속도 분류를 포함한 여섯 가지 작업에 대한 능력을 강화함.

즉, NeuroLM의 주요 내용은 다음과 같음:

1. Text-aligned neural tokenizer embeddings : 정렬을 통해 LLM 강점 극대화함
2. Large-scale multi-channel autoregressive pre-training : EEG 채널 간 인과적 표현 학습하여 일반화 성능 높임
3. Joint multi-task tuning and inference : 다운스트림 작업 최적화

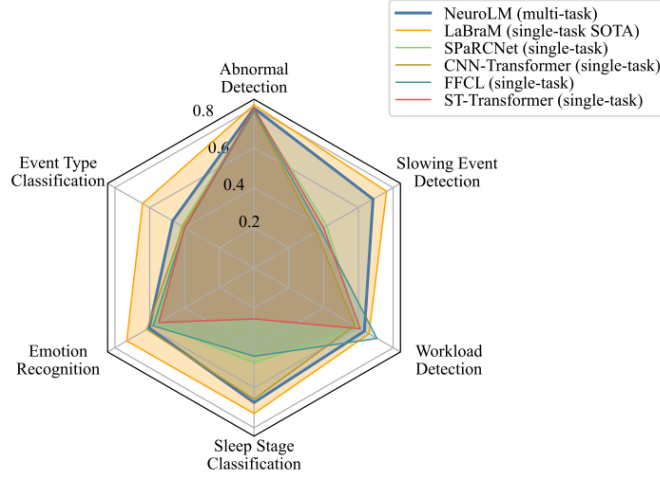


Figure 1: Comparison on six tasks.

Method

Text-Aligned Neural Tokenizer Training

EEG 데이터를 직접 다루는 것이 어렵기 때문에, EEG 신호를 LLM이 이해할 수 있도록 텍스트와 정렬된 토큰으로 변환하는 과정이 필요함.

▼ formulation

- 다채널 EEG 신호 : $X \in \mathbb{R}^{C \times T}$ (C 는 채널 수, T 는 총 타임스탬프)
- EEG 샘플 : $x \in \mathbb{R}^{C \times L}$ (L 은 윈도우 크기)
- 총 샘플 수 : $\lfloor \frac{T}{L} \rfloor$
- EEG 샘플을 non-overlap 패치로 나눔 : $x = \{x_{ij} \in \mathbb{R}^P | i = 1, \dots, C, j = 1, \dots, N\}$
 - (P 는 패치 크기, $N = \lfloor \frac{L}{P} \rfloor$)

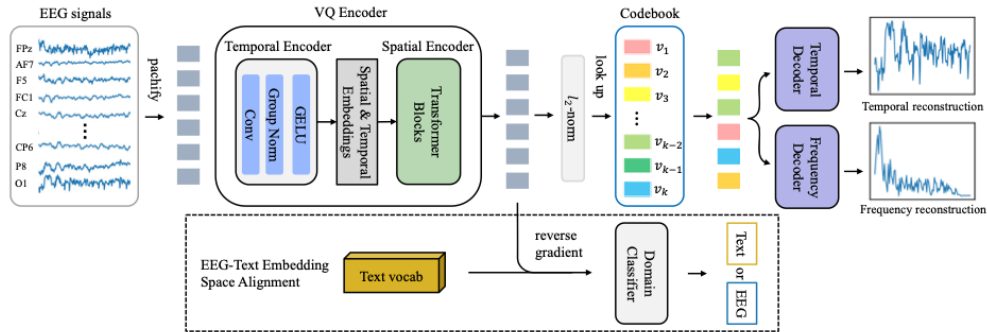


Figure 2: The architecture design of text-aligned neural tokenizer training. The neural tokenizer is trained by reconstructing both temporal and frequency domain of input EEG signals to discretize them into discrete neural tokens. To align EEG and text embedding space, we utilize a domain classifier through adversarial training.

• Neural Tokenizer 구성

- VQ(벡터 양자화) 인코더 : EEG 신호를 이산적인 토큰(discrete tokens)으로 변환 → EEG 신호를 정해진 코드북에 매핑하여 토큰화
 - $z_i = \arg \min_j \|\ell_2(h_i) - \ell_2(v_i)\|_2$
 - h_i 는 패치 표현, v_i 는 코드북 임베딩

- 즉 EEG 샘플은 $z = [z_1, \dots, z_N]$ 로 토큰화됨
- Codebook : 미리 학습된 EEG 토큰 사전
 - $V \in \mathbb{R}^{K \times D}$
- Temporal & Frequency Decoder : EEG 신호의 시간적, 주파수적 정보 확인
- Domain Classifier : EEG와 텍스트 데이터가 같은 임베딩 공간에서 학습되도록 유도. Adversarial Training을 사용하여, LLM이 EEG 데이터를 언어처럼 이해하도록 만들

• Temporal-frequency prediction

- EEG 패치마다 DFT(Discrete Fourier Transform) 적용 → 오일러 공식으로 변환
 - $\tilde{x}_{i,j}^m = \sum_{n=1}^M x[n] \cos\left(\frac{2\pi}{M}mn\right) - jx[n] \sin\left(\frac{2\pi}{M}mn\right)$
- 코드북 임베딩으로 양자화한 후, 정규화된 neural 임베딩을 2개의 디코더에 넣음
 - 주파수 및 시간 특성을 학습
- codebook learning의 최적화 타겟은 다음과 같음:

$$\mathcal{L}_1 = \sum_{\mathbf{x} \in \mathcal{D}} \sum_i \underbrace{\|o_i^t - x_i\|_2^2 + \|o_i^f - f_i\|_2^2}_{\text{reconstruction loss}} + \underbrace{\|\text{sg}(\ell_2(h_i)) - \ell_2(v_{z_i})\|_2^2}_{\text{codebook loss}} + \underbrace{\|\ell_2(h_i) - \text{sg}(\ell_2(v_{z_i}))\|_2^2}_{\text{commitment loss}},$$

- sg는 순방향 계산 중 부분 도함수가 0인 stop-gradient 연산자

• EEG-text embedding space alignment

- 보통 VLM은 대규모 image-text 페어로 학습한 pre-trained CLIP 이미지 인코더를 사용함
- but, EEG-text 정렬은 고려해야 할 사항이 많음
 - 뇌파 세그먼트는 한 사람의 감정과 정신 상태를 포함할 뿐만 아니라, 신체 움직임과 의학적 정상성을 나타낼 수도 있기 때문에 복잡한 정보가 포함되어 있음
 - eeg-text 페어를 구성하는 데 사용할 수 있는 라벨링된 뇌파 데이터는 제한적임
- 따라서, 임베딩 방식이 아닌 텍스트 공간 방식으로 뇌파 정렬 제안
- domain classifier C 제안 : 임베딩이 EEG에서 왔는지, text에서 왔는지 예측하는 역할
 - VQ 인코더 후 gradient reverse layer를 추가 → VQ 인코더로부터 온 임베딩이 text 임베딩과 같은 공간에 존재하게 됨
 - text-aligned neural tokenizer training의 목적 함수 : $\min \mathcal{L}_1 + \lambda \sum_i d_i \log C(h_i)$
 - d_i 는 EEG 또는 text domain의 라벨, λ 는 0~1로 스케일링하는 변수

• VQ encoder architecture

- LaBraM과 거의 동일함
- temporal encoder : 여러 개의 1-D convolution 블록으로 구성 → 각 EEG 패치의 시간적 특징을 추출하는 것이 목표
- 그 후, 표준 10-20 국제 시스템에 따라 학습 가능한 시간/공간 임베딩을 추가하여 시간과 채널 정보 모두 주입
- spatial encoder : vanilla Transformer 블록으로 구성 → 패치 간 상호작용 학습



EEG 데이터를 벡터 양자화하여 이산적인 토큰으로 변환하고, LLM이 학습할 수 있도록 EEG-텍스트 공간을 정렬함으로써, EEG 신호를 자연어처럼 다룰 수 있도록 변환

Multi-Channel Autoregressive Pre-training

- NeuroLM은 multi-channel autoregression, 즉 attention mask를 통해 훈련됨
- 눈에 보이는 EEG 토큰을 기반으로 다음 토큰을 예측하여, 모델이 EEG 인과 관계의 특수 패턴을 학습

- multi-channel autoregressive pre-training은 멀티태스크 instruction 튜닝의 성능에 기여함

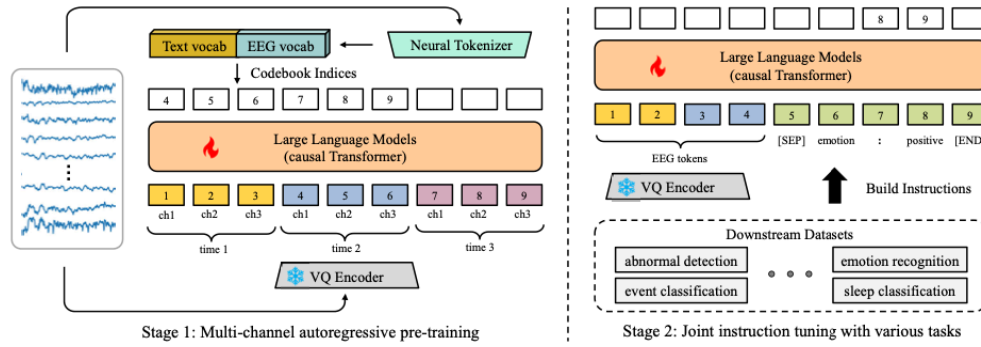


Figure 3: Schematic of NeuroLM training. **Left:** We first pre-train NeuroLM via multi-channel autoregression with EEG tokens output by the frozen VQ encoder. **Right:** The multi-task instruction tuning enables NeuroLM to perform various BCI tasks within a single model.

- 각 EEG 토큰이 현재 및 이전 시간 단계의 모든 채널의 토큰을 관찰할 수 있는 **stair-stepping mask** 정의함
 - 현재 및 이전 시점의 채널을 모두 활용하여 예측 수행 → LLM 구조를 활용하여 EEG 신호의 시계열적 패턴을 효과적으로 학습
 - dark cell은 해당 요소가 attention에 참여해야 함을 나타냄

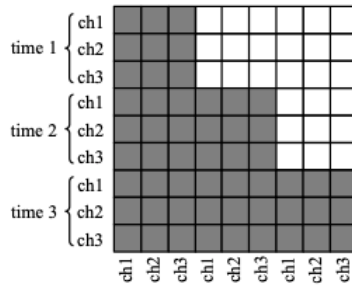


Figure 4: The stair-stepping mask. Each row indicates attention masks for an EEG token.



기존 자연어 모델이 단어를 하나씩 예측하듯이, EEG 채널별로 다음 신호를 예측하는 방식으로 학습하여, EEG의 시계열적 관계를 효과적으로 반영

Multi-Task Instruction Tuning

- 각각의 downstream에 대해 instruction 데이터 구축
- [SEP] 토큰으로 EEG와 text instructions를 합침



하나의 모델에서 여러 EEG 태스크를 해결할 수 있도록, instruction 튜닝을 적용하여 LLM이 다양한 태스크에 적응 가능

Experiments

Downstream Datasets

- NeuroLM은 6개의 다양한 EEG 태스크에서 평가됨

- **TUAB**: 비정상 탐지 (정상/비정상)
- **TUEV**: 이벤트 유형 분류 (6개 클래스)
- **SEED**: 감정 인식 (긍정/중립/부정)
- **HMC**: 수면 단계 분류 (5개 클래스)
- **Workload**: 작업 부하 탐지 (낮음/높음)
- **TUSL**: EEG 슬로잉 이벤트 탐지

Table 1: Information of datasets used for downstream evaluation.

Dataset	#Channel	Sampling Rate	Duration	#Sample	Task
TUAB	23	256 Hz	10 seconds	409,455	Binary classification
TUEV	23	256 Hz	5 seconds	112,491	6-class classification
SEED	62	1000 Hz	4 seconds	38,475	3-class classification
HMC	4	256 Hz	30 seconds	137,243	5-class classification
Workload	19	500 Hz	4 seconds	2,088	Binary classification
TUSL	23	256 Hz	10 seconds	245	3-class classification

Experimental Setup

- model configurations
 - GPT-2 base
 - B(254M), L(500M), XL(1696M)
- environment settings
 - 8개 NVIDIA A100-80G GPUs
 - Python 3.11.8 / PyTorch 2.2.2 + CUDA 12.1

Experimental Results

Table 2: Results on TUAB and TUEV.

Methods	Multi-task	TUAB			TUEV		
		Balanced Acc.	AUC-PR	AUROC	Balanced Acc.	Cohen’s Kappa	Weighted F1
SPaRCNet	✗	0.7896±0.0018	0.8414±0.0018	0.8676±0.0012	0.4161±0.0262	0.4233±0.0181	0.7024±0.0104
ContraWR	✗	0.7746±0.0041	0.8421±0.0104	0.8456±0.0074	0.4384±0.0349	0.3912±0.0237	0.6893±0.0136
CNN-Transformer	✗	0.7777±0.0022	0.8433±0.0039	0.8461±0.0013	0.4087±0.0161	0.3815±0.0134	0.6854±0.0293
FFCL	✗	0.7848±0.0038	0.8448±0.0065	0.8569±0.0051	0.3979±0.0104	0.3732±0.0188	0.6783±0.0120
ST-Transformer	✗	0.7966±0.0023	0.8521±0.0026	0.8707±0.0019	0.3984±0.0228	0.3765±0.0306	0.6823±0.0190
BIOT	✗	0.7959±0.0057	0.8792±0.0023	0.8815±0.0043	0.5281±0.0225	0.5273±0.0249	0.7492±0.0082
LaBraM-Base	✗	0.8140±0.0019	0.8965±0.0016	0.9022±0.0009	0.6409±0.0065	0.6637±0.0093	0.8312±0.0052
NeuroLM-B	✓	0.7826±0.0065	0.6975±0.0081	0.7816±0.0079	0.4560±0.0048	0.4285±0.0048	0.7153±0.0028
NeuroLM-L	✓	0.7876±0.0034	0.7099±0.0034	0.7876±0.0034	0.4132±0.1235	0.4414±0.0996	0.7387 ±0.0400
NeuroLM-XL	✓	0.7969 ±0.0091	0.7219 ±0.0082	0.7884 ±0.0194	0.4679 ±0.0356	0.4570 ±0.0498	0.7359±0.0219

Table 3: Results on SEED and HMC.

Methods	Multi-task	SEED			HMC		
		Balanced Acc.	Cohen’s Kappa	Weighted F1	Balanced Acc.	Cohen’s Kappa	Weighted F1
SPaRCNet	✗	0.5596±0.0244	0.3464±0.0372	0.5585±0.0297	0.4756±0.1109	0.3147±0.1315	0.4108±0.1310
ContraWR	✗	0.6106±0.0078	0.4220±0.0129	0.6137±0.0085	0.4242±0.0541	0.2340±0.0554	0.2987±0.0288
CNN-Transformer	✗	0.6161±0.0384	0.4262±0.0601	0.6150±0.0463	0.6573±0.0141	0.5961±0.0105	0.6896±0.0065
FFCL	✗	0.5808±0.0322	0.3732±0.0462	0.5743±0.0402	0.4427±0.0702	0.2542±0.0654	0.2902±0.0485
ST-Transformer	✗	0.5479±0.0091	0.3261±0.0169	0.5505±0.0091	0.2559±0.0141	0.0503±0.0183	0.1428±0.0122
BIOT	✗	0.7097±0.0024	0.5682±0.0051	0.7134±0.0027	0.6862±0.0041	0.6295±0.0113	0.7091±0.0147
LaBraM-Base	✗	0.7318±0.0019	0.5994±0.0031	0.7354±0.0021	0.7286±0.0101	0.6812±0.0073	0.7554±0.0024
NeuroLM-B	✓	0.5554±0.0075	0.3393±0.0117	0.5599±0.0068	0.6737 ±0.0050	0.6188 ±0.0057	0.7126 ±0.0034
NeuroLM-L	✓	0.6006±0.0047	0.4067±0.0063	0.6048±0.0050	0.6658±0.0550	0.5929±0.0715	0.6896±0.0504
NeuroLM-XL	✓	0.6034 ±0.0010	0.4082 ±0.0036	0.6063 ±0.0030	0.5761±0.1084	0.4795±0.1466	0.5883±0.1286

Table 4: Results on Workload and TUSL.

Methods	Multi-task	Workload			TUSL		
		Balanced Acc.	AUC-PR	AUROC	Balanced Acc.	Cohen's Kappa	Weighted F1
SPaRCNet	✗	0.5977±0.0071	0.6638±0.0314	0.6717±0.0172	0.4185±0.0452	0.1399±0.0799	0.3500±0.0968
ContraWR	✗	0.6966±0.0332	0.7668±0.0408	0.7685±0.0317	0.5857±0.0662	0.3567±0.0968	0.5458±0.0798
CNN-Transformer	✗	0.5793±0.0230	0.5306±0.0459	0.5663±0.0349	0.3575±0.0151	0.0306±0.0179	0.2235±0.0251
FFCL	✗	0.7069±0.0197	0.7823±0.0099	0.7857±0.0234	0.3819±0.0688	0.0628±0.0888	0.2120±0.0786
ST-Transformer	✗	0.6103±0.0056	0.5716±0.0071	0.6375±0.0078	0.4000±0.0329	0.0860±0.0449	0.3793±0.0459
BIOT	✗	0.6655±0.0665	0.7189±0.0722	0.7342±0.0536	0.5758±0.0303	0.2012±0.0212	0.2394±0.0040
LaBraM-Base	✗	0.6609±0.0204	0.7174±0.0234	0.7272±0.0165	0.7625±0.0131	0.6407±0.0304	0.7614±0.0210
NeuroLM-B	✓	0.6172±0.0113	0.5824±0.0080	0.6253±0.0160	0.6734±0.0436	0.5107±0.0617	0.6743±0.0394
NeuroLM-L	✓	0.6311±0.0250	0.5869±0.0155	0.6247±0.0339	0.5314±0.0530	0.2961±0.0810	0.5243±0.0680
NeuroLM-XL	✓	0.6345±0.0442	0.5889±0.0423	0.6130±0.0764	0.6845±0.0304	0.5194±0.0461	0.6839±0.0297

Ablation on Robustness

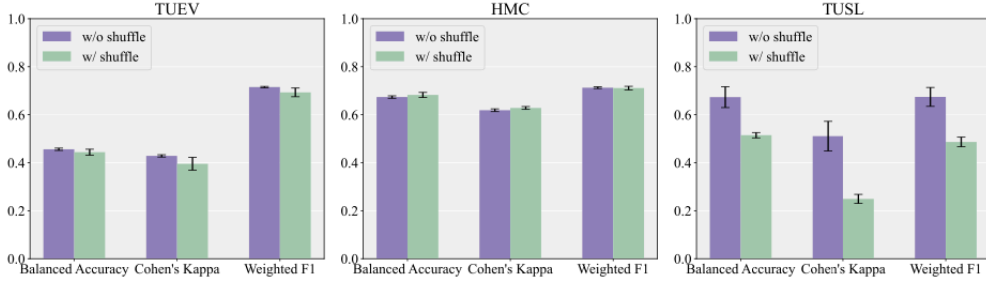


Figure 5: Ablation study on whether shuffling the options of instructions.

Ablation on Instruction Data Size

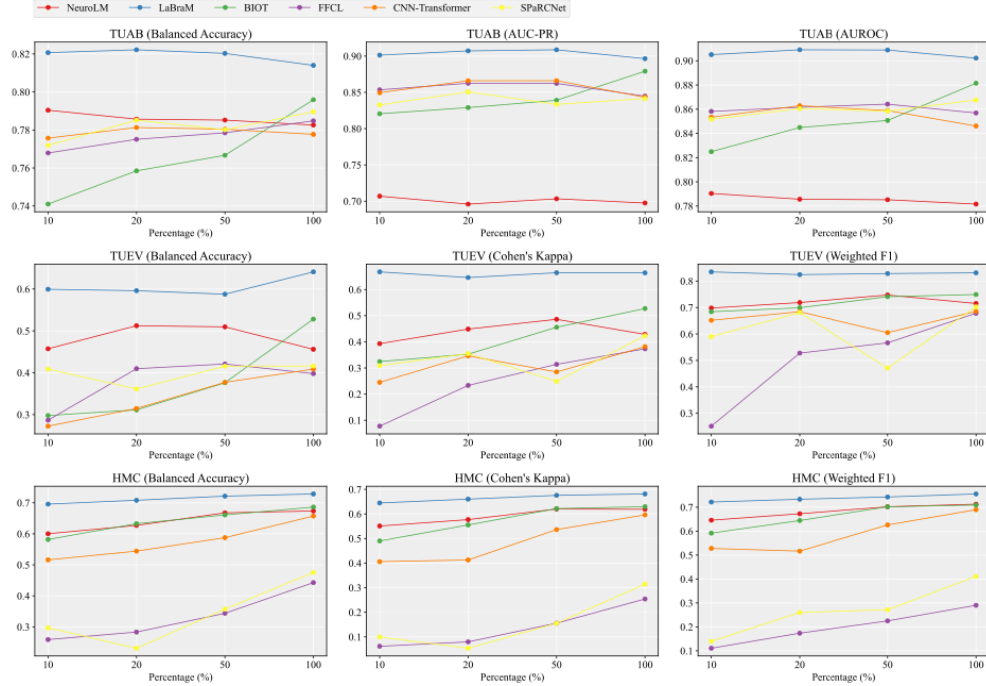


Figure 6: Comparison of different methods under different proportions of instruction data.

Visualization Curves of Multi-Channel Autoregression

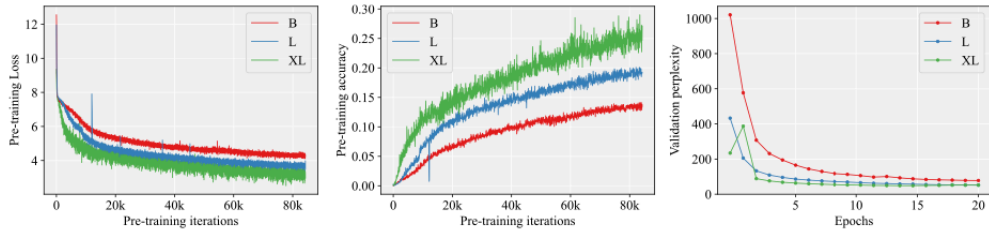


Figure 7: The training and validation visualization of multi-channel autoregressive pre-training.

Ablation on Multi-Channel Autoregressive Pre-training

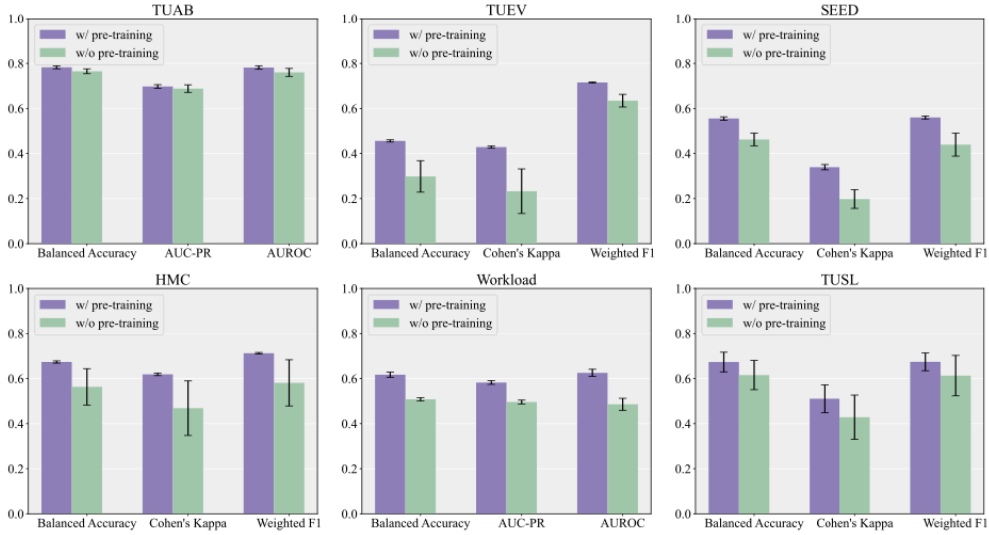


Figure 8: Ablation study on multi-channel autoregressive pre-training.

Conclusion

- NeuroLM은 최초의 EEG 멀티태스크 기초 모델(Foundation Model)
- EEG 데이터를 자연어처럼 다룰 수 있도록 변환하는 신경 토크나이저를 개발
- 멀티태스크 학습을 통해 단일 모델에서 여러 태스크 수행 가능
- 기존 모델 대비 보다 효율적으로 다양한 EEG 태스크를 해결할 수 있음

Related Work

Instruction Design

Table 5: Information of instruction design for downstream datasets.

Dataset	Instruction Description
TUAB	[SEP] Question: Is this EEG segment abnormal? Answer: {Yes, No} [END]
TUEV	[SEP] Question: Which event type does this EEG segment belong to? Options: (A) spike and slow wave. (B) generalized periodic epileptiform discharge. (C) periodic lateralized epileptiform discharge. (D) eye movement. (E) artifact. (F) background. Answer: {(A), (B), (C), (D), (E), (F)} [END]
SEED	[SEP] Question: Which emotion type does this EEG segment belong to? Answer: {Positive, Neutral, Negative} [END]
HMC	[SEP] Question: Which sleep type does this EEG segment belong to? Options: (A) Wake. (B) NREM-1. (C) NREM-2. (D) NREM-3. (E) REM. Answer: {(A), (B), (C), (D), (E)} [END]
Workload	[SEP] Question: Is this EEG segment of high workload? Answer: {Yes, No} [END]
TUSL	[SEP] Question: Which type does this EEG segment belong to? Options: (A) background. (B) seizure. (C) slowing. Answer: {(A), (B), (C)} [END]

Pre-training Dataset Description

Table 6: Information of datasets used for pre-training.

Dataset	#Channel	Rate (Hz)	Time (h)	Description
TUEG (Obeid & Picone, 2016)	17-23	250-1024	~24,000	A corpus of 26,846 clinical EEG recordings collected at Temple University Hospital.
SEED Series (Zheng et al., 2018; Liu et al., 2021; 2022)	62	1000	170.54	These datasets including SEED-IV (15 subjects), SEED-V (20 subjects), SEED-GER (8 subjects), and SEED-FRA (8 subjects) in response to emotional videos.
BCI Competition IV-1 (Blankertz et al., 2007)	59	1000	8.21	A motor imagery dataset containing 2 classes of left hand, right hand, foot (+ idle state) for 7 subjects.
Emobrain (Savran et al., 2006)	64	1024	4.94	A multimodal emotion dataset including 16 subjects. The emotions were elicited through a selected subset of the IAPS dataset.
Grasp and Lift (Luciw et al., 2014)	32	500	11.72	A dataset containing 12 subjects performing grasp-and-lift (GAL) trials.
Inria BCI (Margaux et al., 2012)	56	600	29.98	A P300-based spelling dataset including 26 subjects.
Motor Movement/Imagery (Schalk et al., 2004)	64	160	47.3	A motor imagery dataset consisting of 109 volunteers performing 2 baseline tasks, motor movement, and motor imagery.
Raw EEG Data (Trujillo, 2020)	64	256	34.35	EEG was recorded during reported Information-Integration categorization and reported multidimensional Rule-Based categorization tasks.
Resting State (Trujillo et al., 2017)	64	256	3.04	A dataset comprising 22 subjects for a resting task of 8 mins with 4 mins of eyes closed and 4 mins of eyes open.
Siena Scalp EEG Database (Detti et al., 2020)	31	512	30.47	A database consisting of 14 patients.
SPIS Resting State (Torkamani-Azar et al., 2020)	64	2048	0.83	A dataset including 10 subjects, 2.5 minutes recording in eyes-closed and eyes-open prior to a 105-minute session of Sustained Attention to Response Task with fixed-sequence and varying ISIs.
Target Versus Non-Target (Korczowski et al., 2019)	32	512	16	A dataset including 50 subjects playing Brain Invaders, a visual P300 Brain-Computer Interface using oddball paradigm with adaptive Riemannian Geometry (no-calibration).
Self-collected EEG corpus (Jiang et al., 2023; 2021; Luo et al., 2022; Li et al., 2021; Tao & Lu, 2020)	62	1000	342.23	A mixed self-collected EEG datasets of more than 140 subjects under various conditions.

Detailed Experimental Settings

Hyperparameter Settings

Table 7: Hyperparameters for neural tokenizer.

Hyperparameters		Values
Temporal Encoder	Input channels	{1,16,16}
	Output channels	{16,16,16}
	Kernel size	{15,3,3}
	Stride	{8,1,1}
	Padding	{7,1,1}
Transformer encoder layers		12
Transformer decoder layers		3
Hidden size		768
MLP size		3072
Attention head number		12
Codebook size		8192×128
EEG Batch size		512
Text Batch size		128
Peak learning rate		5e-5
Minimal learning rate		1e-5
Learning rate scheduler		Cosine
Optimizer		AdamW
Adam β		(0.9,0.999)
Weight decay		1e-4
Total epochs		50
Warmup epochs		5
Data overlap		None
Gradient clipping		None

Table 8: Hyperparameters for autoregressive pre-training.

Hyperparameters	NeuroLM-B	NeuroLM-L	NeuroLM-XL
Model size	254M	500M	1696M
Transformer encoder layers	12	24	48
Hidden size	768	1024	1600
MLP size	3072	4096	6400
Attention head number	12	16	25
EEG batch size	480 (B), 512 (L, XL)		
Text batch size	32 (B), 64 (L, XL)		
Peak learning rate	6e-4		
Minimal learning rate	6e-5		
Learning rate scheduler	Cosine		
Optimizer	AdamW		
Adam β	(0.9,0.95)		
Weight decay	0.1		
Total epochs	20		
Warmup epochs	2		
Data overlap	None		
Gradient clipping	1		

Table 9: Hyperparameters for instruction tuning.

Hyperparameters	Values
Instruction batch size	512
Text batch size	128
Peak learning rate	5e-4 (B), 5e-5 (L), 2e-5 (XL)
Minimal learning rate	5e-5 (B), 5e-6 (L), 2e-6 (XL)
Learning rate scheduler	Cosine
Optimizer	AdamW
Adam β	(0.9,0.95)
Weight decay	0.1
Total epochs	5 (B, L), 3 (XL)
Warmup ratio	0.1
Gradient clipping	1

Metrics

Baselines

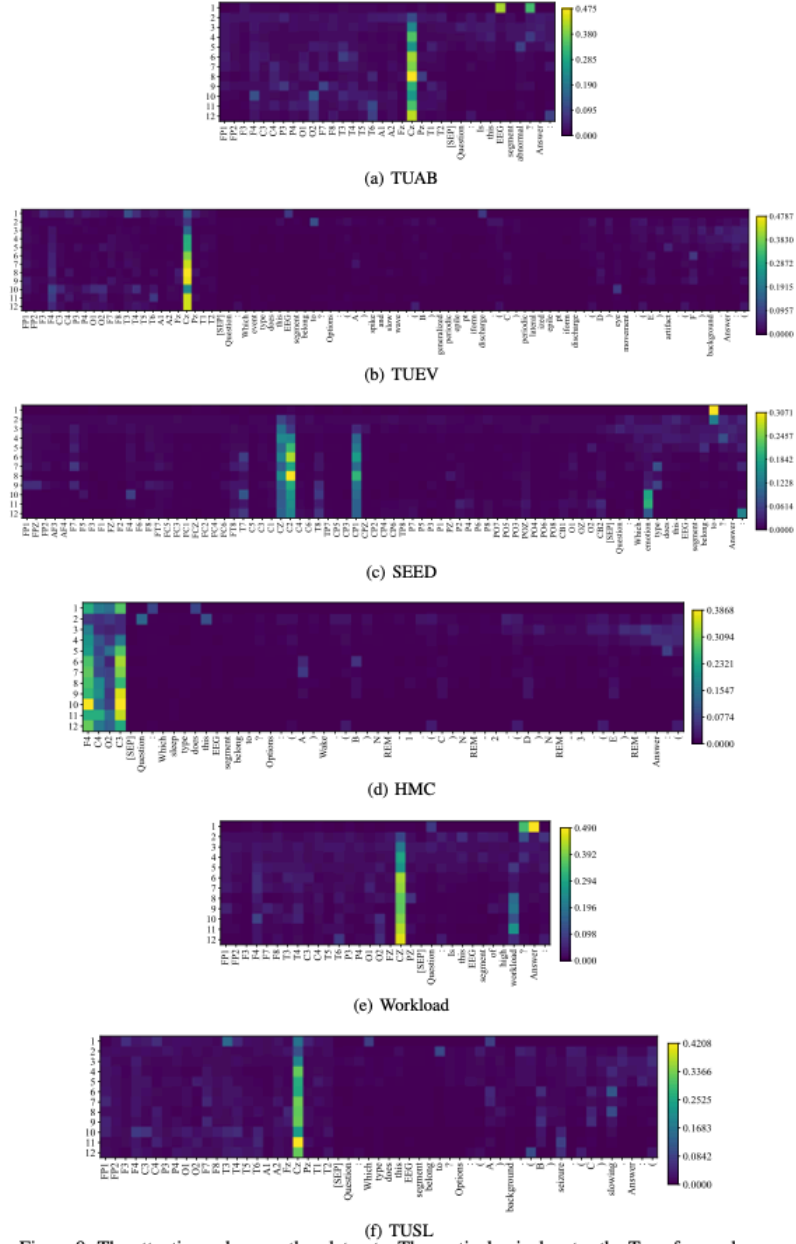


Figure 9: The attention value on other datasets. The vertical axis denotes the Transformer layers.

Attention Visualization

Analysis of Neural Tokenizer

Ablation on Temporal-Frequency Prediction

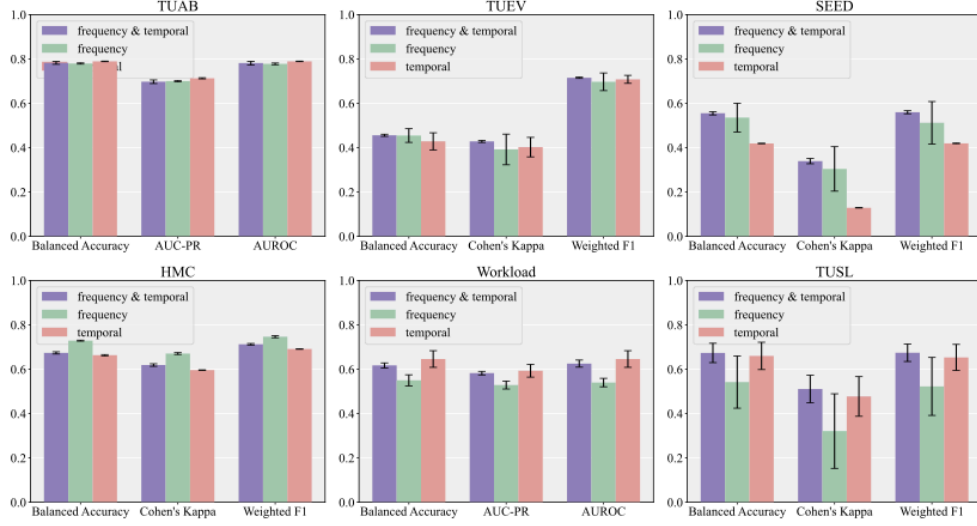


Figure 10: Ablation study on reconstructing temporal or frequency domain in neural tokenizer.

Visualization of EEG and Text Embeddings

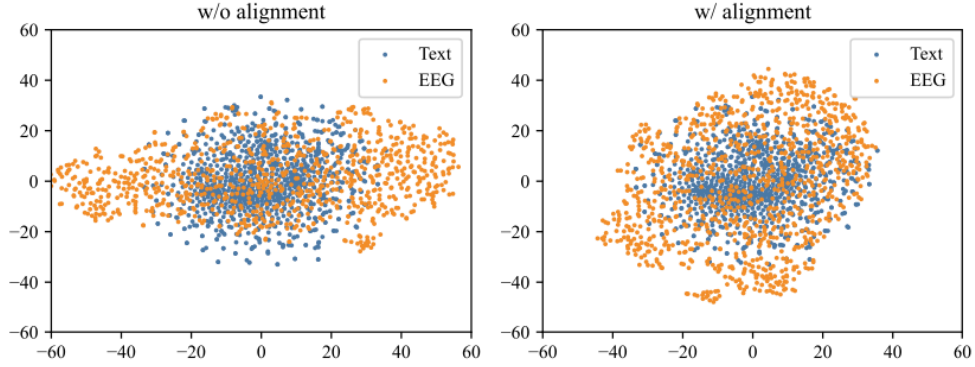


Figure 11: Representation visualization of EEG and text by t-SNE. **Left:** training neural tokenizer without alignment. **Right:** training neural tokenizer with alignment.

Ablation on Different Pre-training Epochs

Table 10: Results on TUAB and TUEV.

Pre-trained Epochs	TUAB			TUEV		
	Balanced Acc.	AUC-PR	AUROC	Balanced Acc.	Cohen's Kappa	Weighted F1
5	0.7763±0.0037	0.6891±0.0040	0.7762±0.0038	0.4609±0.0696	0.3914±0.0670	0.6975±0.0249
10	0.7738±0.0040	0.6995 ±0.0038	0.7647±0.0143	0.4693 ±0.0175	0.4625 ±0.0091	0.7353 ±0.038
15	0.7780±0.0050	0.6951±0.0054	0.7735±0.0112	0.4557±0.0277	0.4216±0.0215	0.7131±0.0101
20	0.7826 ±0.0065	0.6975±0.0081	0.7816 ±0.0079	0.4560±0.0048	0.4285±0.0048	0.7153±0.0028

Table 11: Results on SEED and HMC.

Pre-trained Epochs	SEED			HMC		
	Balanced Acc.	Cohen's Kappa	Weighted F1	Balanced Acc.	Cohen's Kappa	Weighted F1
5	0.5641 ±0.0103	0.3505 ±0.0172	0.5679 ±0.0123	0.6956 ±0.0136	0.6269 ±0.0053	0.7118±0.0048
10	0.5553±0.0089	0.3376±0.0143	0.5592±0.0091	0.6763±0.0054	0.6161±0.0069	0.7065±0.0057
15	0.5543±0.0156	0.3365±0.0244	0.5572±0.0164	0.6543±0.0151	0.5991±0.0153	0.6674±0.0265
20	0.5554±0.0075	0.3393±0.0117	0.5599±0.0068	0.6737±0.0050	0.6188±0.0057	0.7126 ±0.0034

Table 12: Results on Workload and TUSL.

Pre-trained Epochs	Workload			TUSL		
	Balanced Acc.	AUC-PR	AUROC	Balanced Acc.	Cohen's Kappa	Weighted F1
5	0.5816±0.0235	0.5483±0.0160	0.5815±0.0236	0.5342±0.0235	0.2950±0.0345	0.5241±0.0260
10	0.6540 ±0.0192	0.6123 ±0.0165	0.6501 ±0.0178	0.5920 ±0.0560	0.3884±0.0876	0.5984 ±0.0574
15	0.5701±0.0282	0.5426±0.0177	0.5682±0.0255	0.5910±0.0629	<u>0.3915</u> ±0.0999	0.5868±0.0536
20	<u>0.6172</u> ±0.0113	<u>0.5824</u> ±0.0080	<u>0.6253</u> ±0.0160	0.6734 ±0.0436	0.5107 ±0.0617	0.6743 ±0.0394

Discussion

- limitations
 - 특정 단일 작업 기준선을 초과할 수 있지만, 각 다운스트림 데이터셋에서 엔드 투 엔드로 훈련된 최신 방법들에 비해서는 여전히 부족함
 - 하이퍼파라미터 설정에 다소 민감하여, 신중한 조정 없이는 만족스러운 결과를 도출하지 못할 수 있음
 - 고품질 EEG-text 쌍이 제한적으로 제공되기 때문에, 이 논문에서는 EEG와 언어 사이의 대략적인 정렬, 즉 공간적인 정렬만을 사용하여 LLM이 EEG 토큰에서 유용한 정보를 추출하는 데 어려움을 겪을 수 있음
- outlook
 - 보다 고급 LLM을 기본 모델로 활용 : 이 논문은 비교적 작은 LLM인 GPT-2를 사용하면서도 다중 작업 패러다임에서 유망한 결과를 달성했지만, LLaMA 3와 같은 최신의 고급 오픈소스 LLM을 활용하는 것이 NeuroLM의 다중 작업 학습 능력을 크게 향상시킬 수 있음
 - 전문가 혼합 접근 방식을 채택 : EEG와 언어 간의 모달리티 격차를 고려할 때, 모달리티 특화 전문가를 사용하는 것이 LLM으로 멀티모달 학습을 개선할 수 있음
 - EEG 샘플을 미리 정의된 문장으로 설명하고, VQ 훈련 단계에서 대조 학습 손실을 추가하여 EEG와 텍스트 설명을 정렬하는 것과 같은 보다 세밀한 EEG 및 텍스트 정렬 방법을 개발하면 성능을 더욱 향상시킬 수 있음