Cutting Gardens

USC University of Southern California

USC Viterbi School of Engineering

LOS ANGELES GARDEN

Neuro-GPT: A Foundation Model for Advanced EEG Decoding

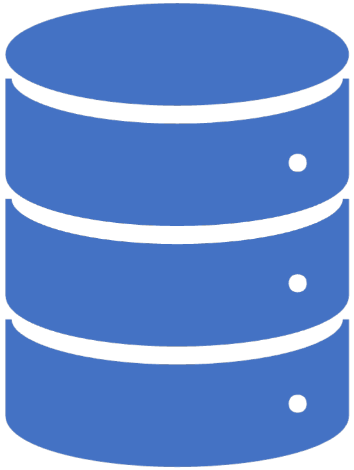
Wenhui Cui

Scarcity and Heterogeneity of Neural Data

- Challenge in Post-traumatic Epilepsy Prediction
  - fMRIs of patients with traumatic brain injury are very heterogenous
  - Only a small number of PTE subjects are available for training.
  - It is challenging for machine learning models to generalize on PTE data and extract representative features.

Tackle Data Scarcity and Heterogeneity

- Transfer knowledge from large scale datasets: transfer learning strategies.
- Learn generalizable features: self-supervised learning.
- Can we train a model on large scale data using a a self-supervised task to learn transferable features that are beneficial for our task of interest ?



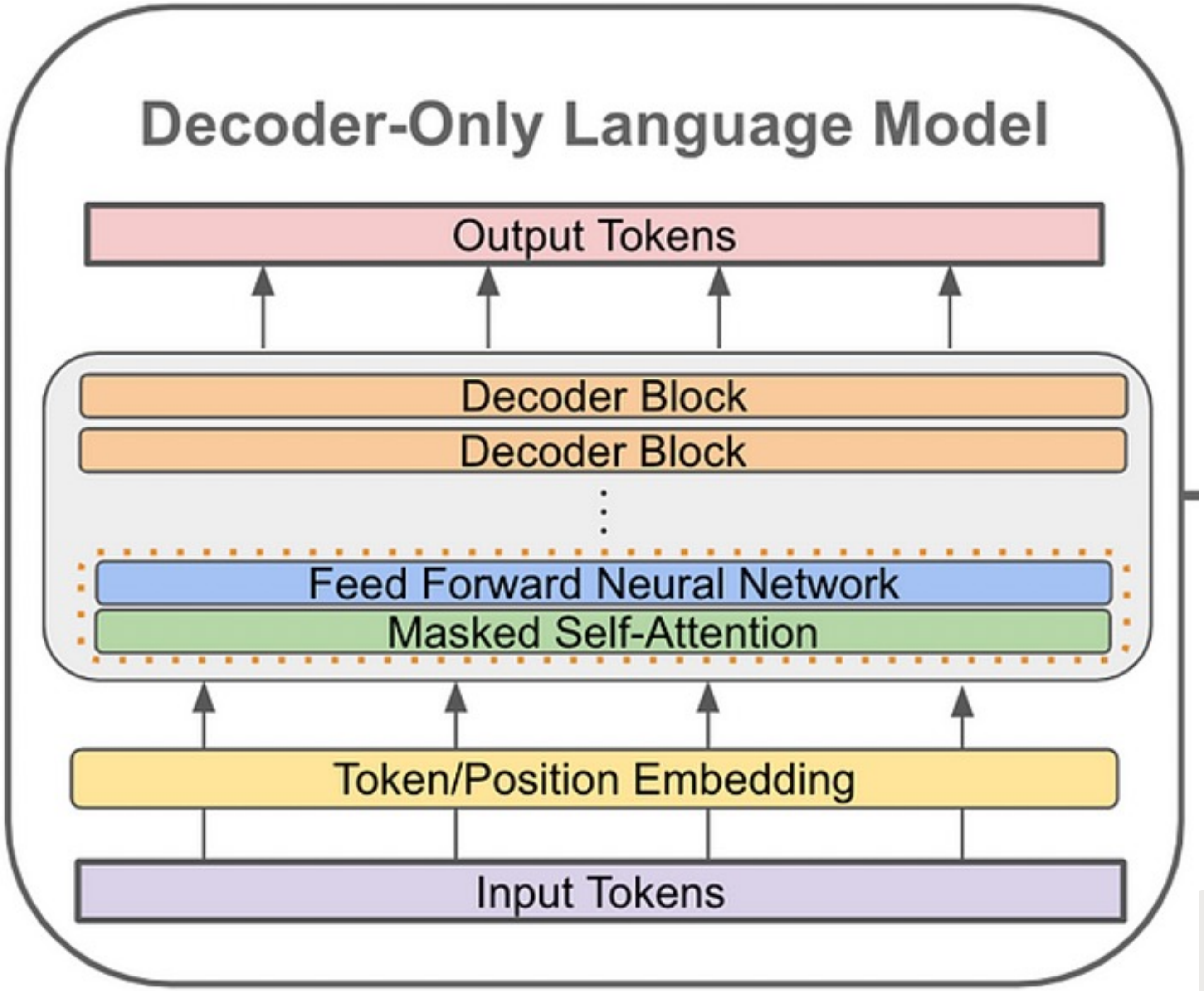
IT ALL STARTED WITH A GRAND DAY OUT

The success of ChatGPT inspired us

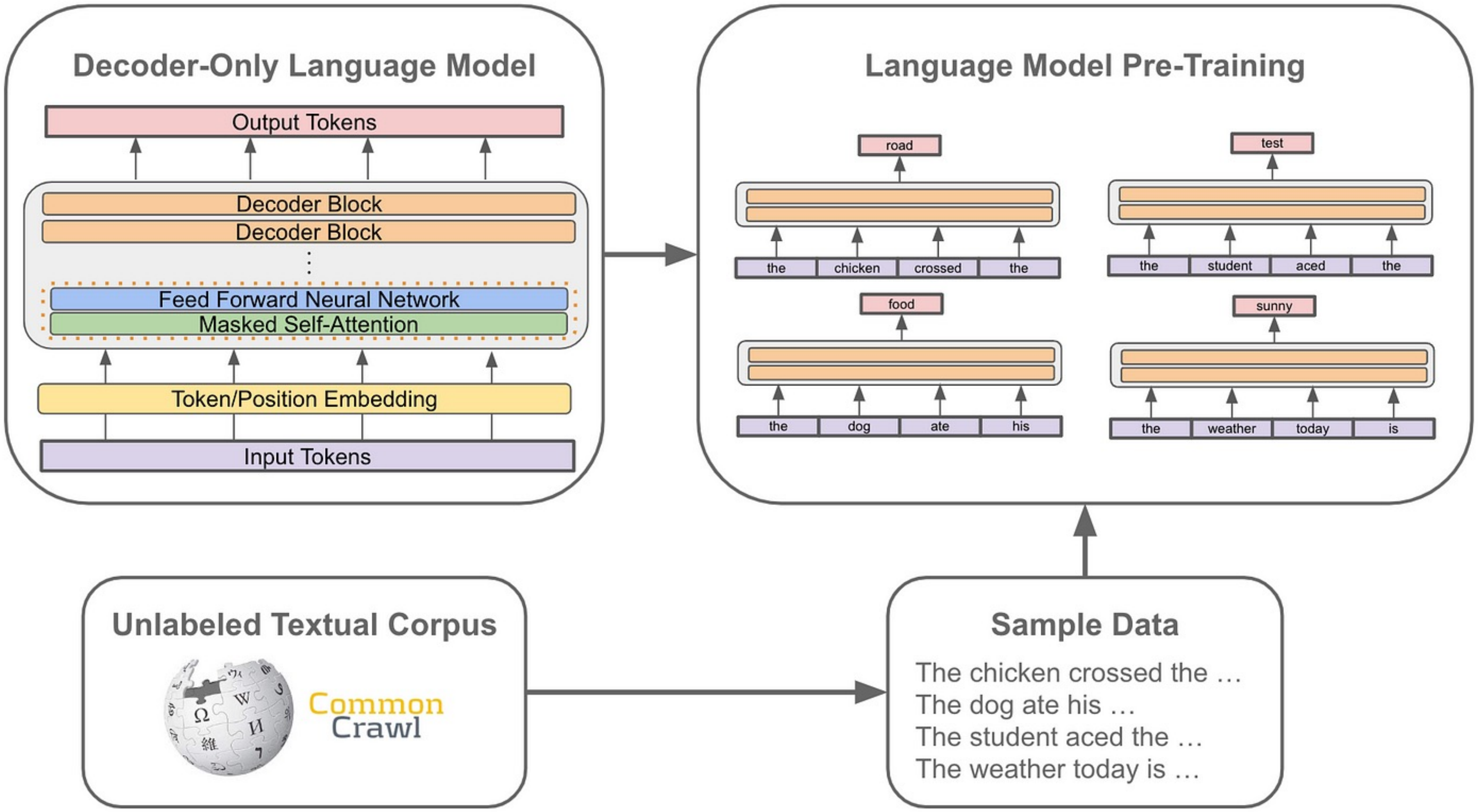
Why not develop our own Neuro-GPT for EEG decoding?

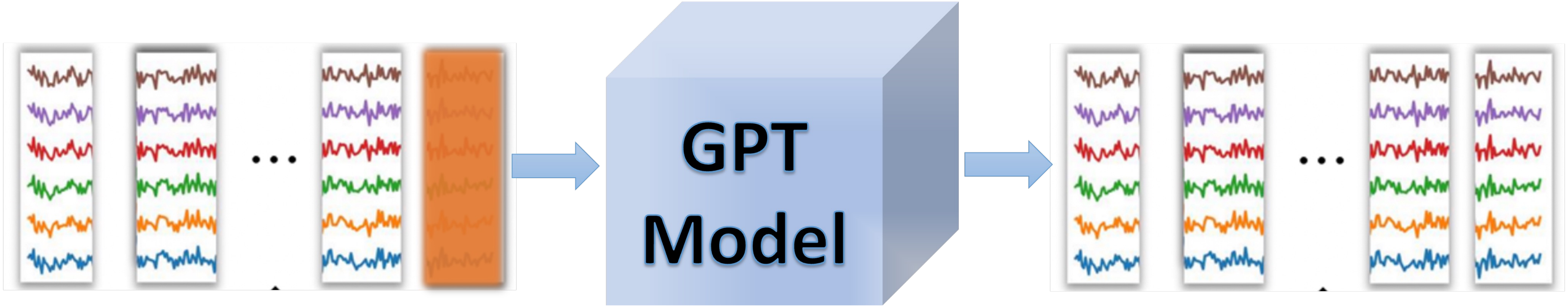
Can we apply large language models to EEG data and develop a foundation model?

GPT Model Architecture



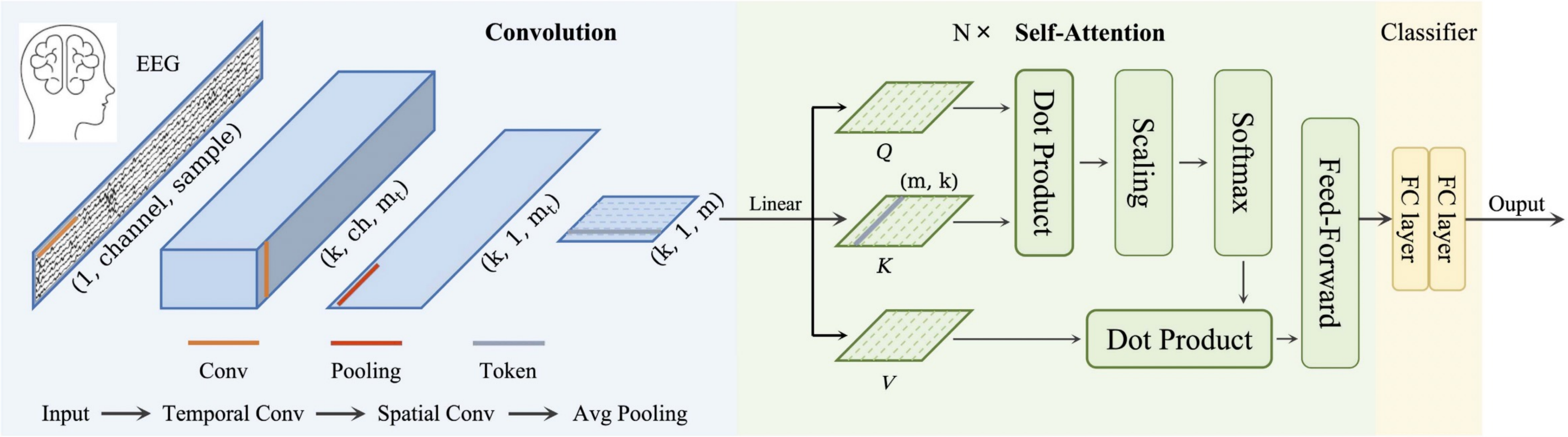
From Sentence to Time Series





From Sentence to Time Series

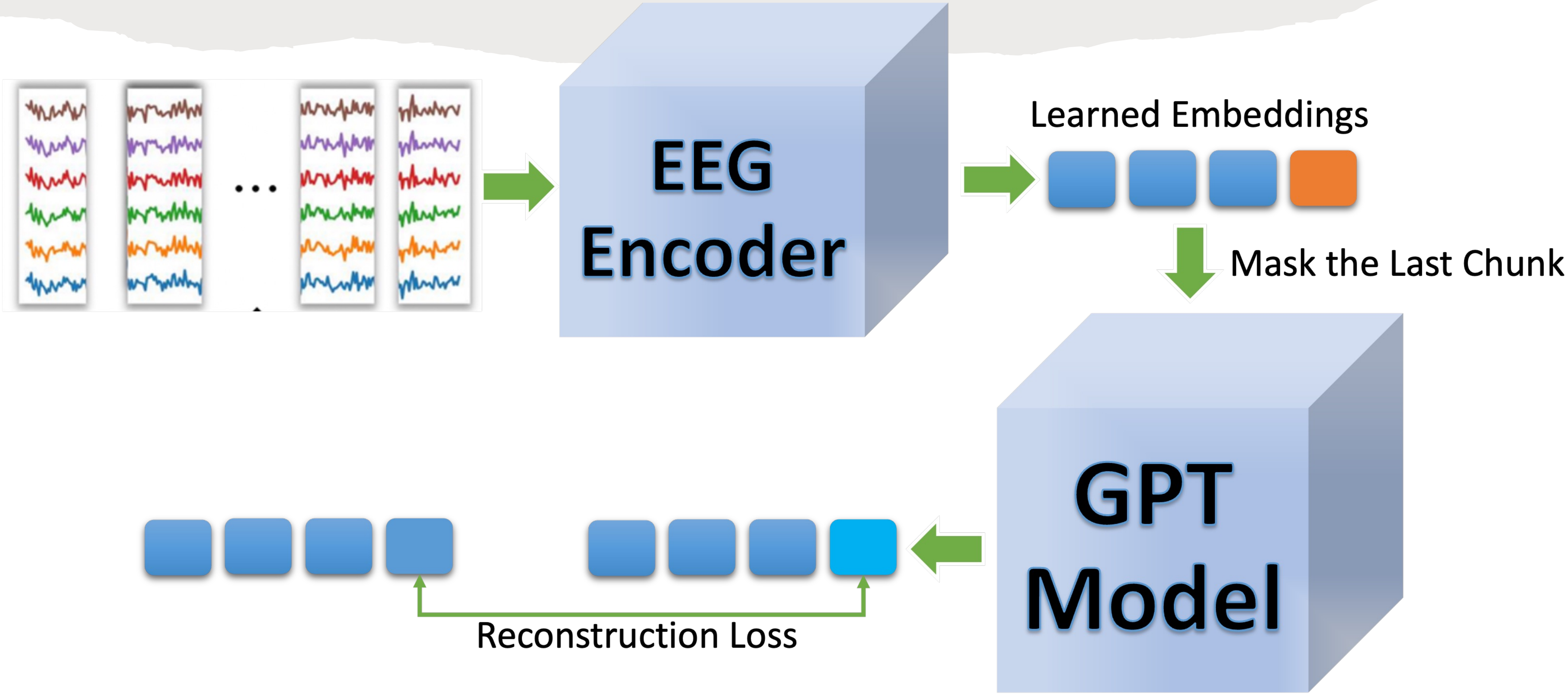
- EEG data is sequential
- Split the whole time series into chunks, each chunk can be viewed as a word or token.



EEG Encoder

- GPT Model only captures temporal patterns.
- EEG data has rich spatial-temporal information.
- EEG data is high dimensional and noisy.
- Introduce an EEG encoder to extract useful spatial-temporal features for GPT model to learn the temporal relations.

The Proposed Neuro-GPT Pipeline



Causal Reconstruction Loss

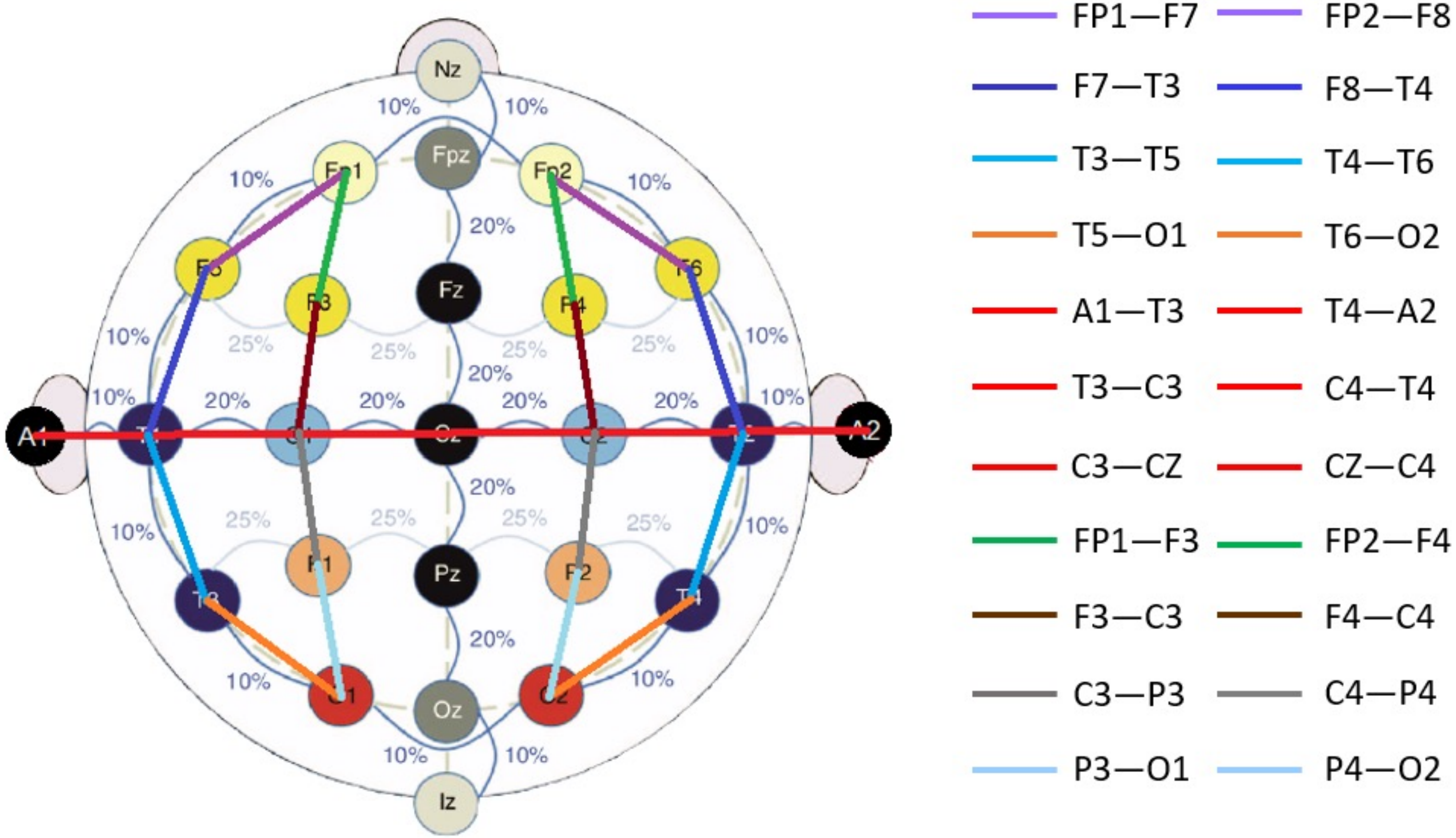
- Predict masked chunk given its preceding chunks and without seeing the future
- We mask chunk in the sequence each at a time
- Deploy the causal loss so the model learns causal temporal relations between every chunk.

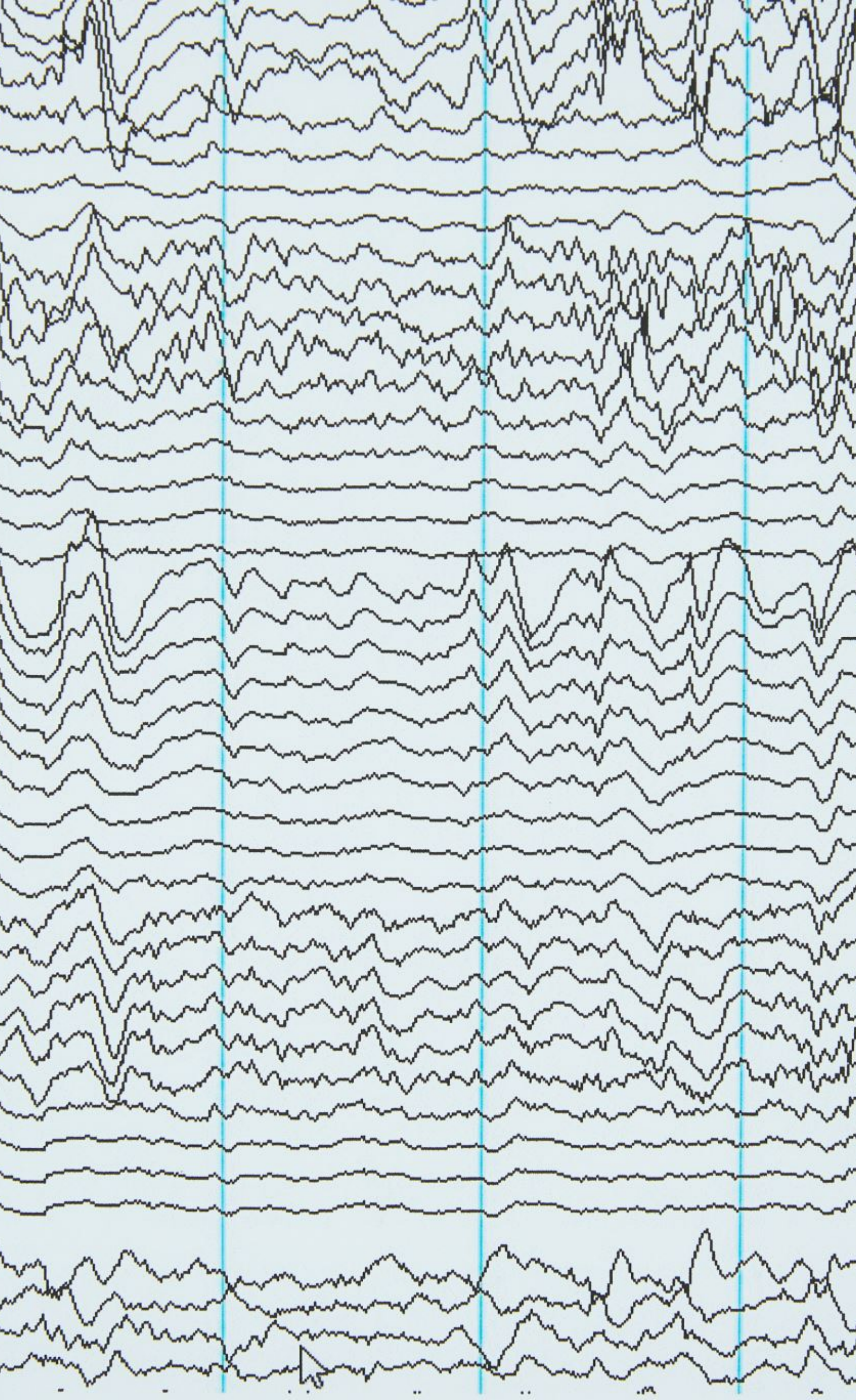
$$L = -\sum_{i=1}^{N} \log P(t_i | t_1, t_2, ..., t_{i-1})$$

- Where t represents each chunk, N is the total number of chunks.

TUH EEG Corpus

- A rich archive of 26846 clinical EEG recordings collected at Temple University.
- Heterogeneous channel configurations and sampling frequencies.
- Various duration of recordings;
- Data with epilepsy, seizure, and artifacts.
- International 10-20 system





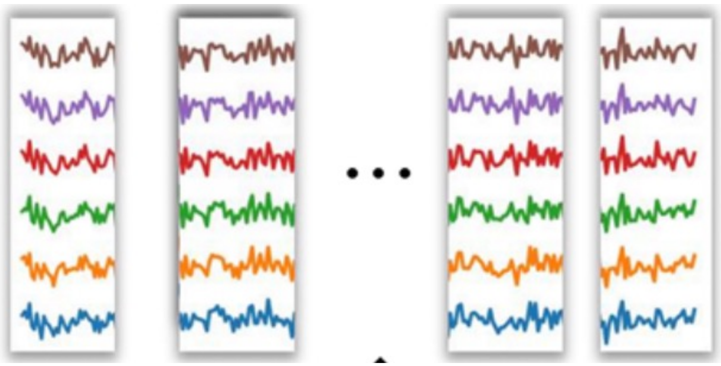
TUH EEG Data Preprocessing

- Use Brainstorm to build a preprocessing pipeline
  - Resample to 250 Hz (the majority of sampling frequency)
  - Select 22 common channels in the International 10-20 system based on the template in Brainstorm.

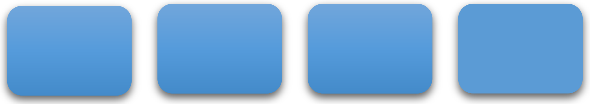
  - Bad channel removal
  - Notch filter: 60 Hz
  - Band-pass filter: 0.5 ~ 100 Hz
  - Remove DC offset and linear trend
  - Normalization over time

Input Configurations

- Split the whole recording into T seconds chunks with overlapping (2s, 0.2s overlapping)
- Randomly sample 8 contiguous chunks for each subject.



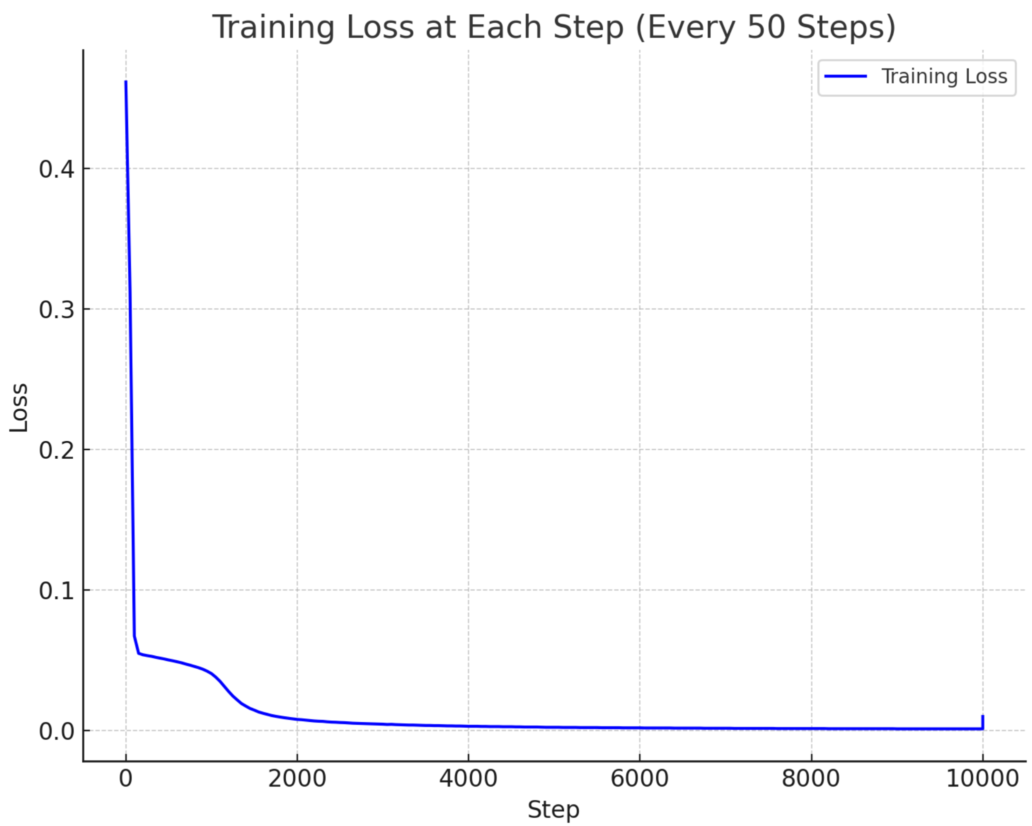
- Training batch (chunk1, chunk2, …, chunk8,
  chunk1, chunk2, …, chunk8, …)
- Each chunk has a dimension of *channel x T*
- Dimension of Input tokens to GPT model: *num_chunks x F*

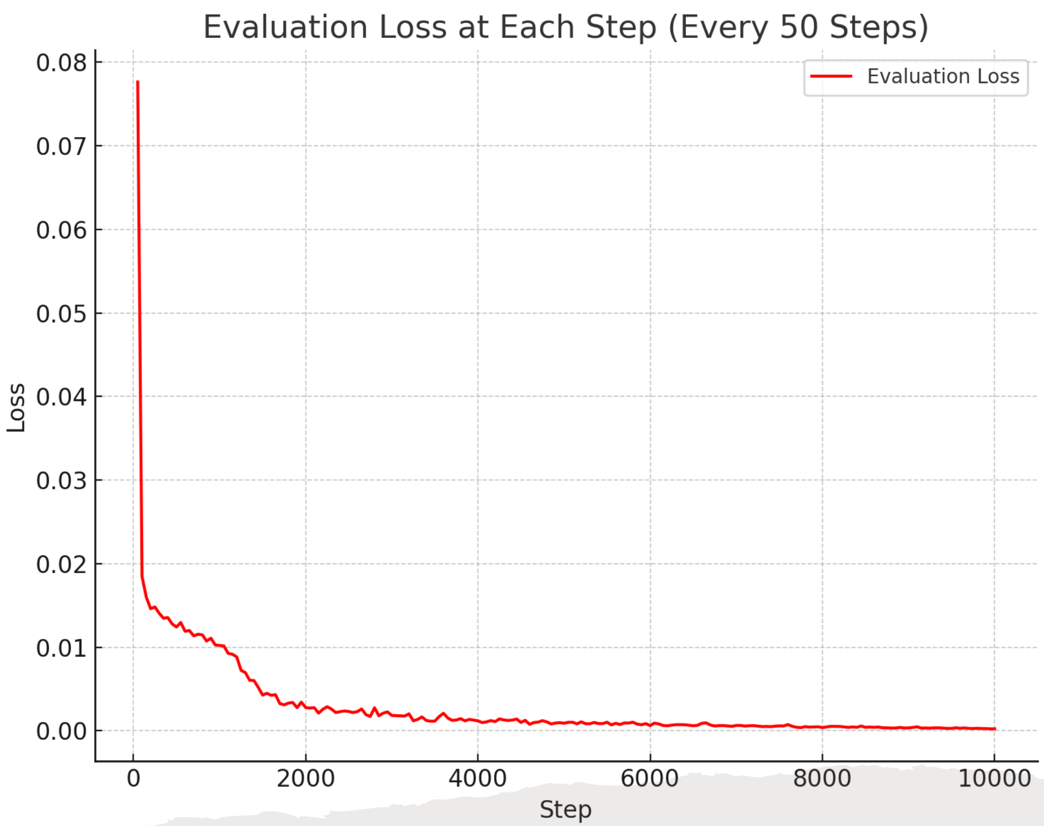


- *Where F is the dimension of flattened features produced by EEG encoder.*

Pretraining the Model

- TUH EEG Training set: 19k recordings;
- Held-out validation set: 1k recordings





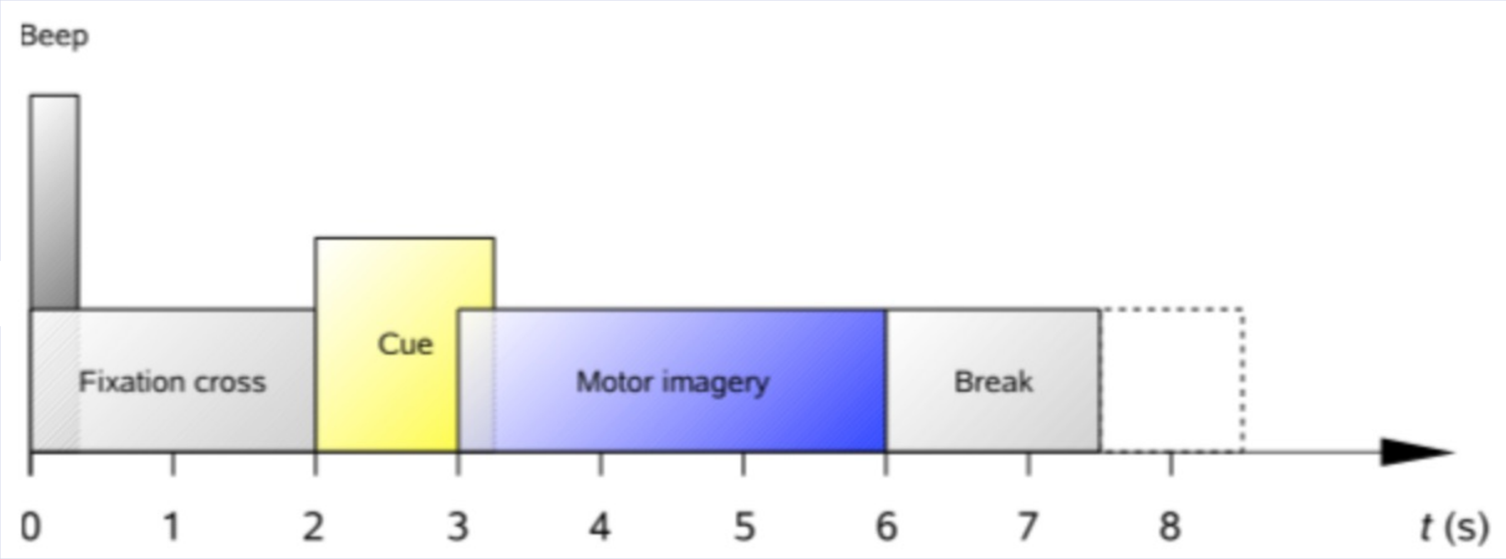
Downstream: BCI 2a Dataset

https://www.bbci.de/competition/iv/#dataset2a

- 4-class Motor Imagery EEG, cued motor imagery (left hand, right hand, feet, tongue)

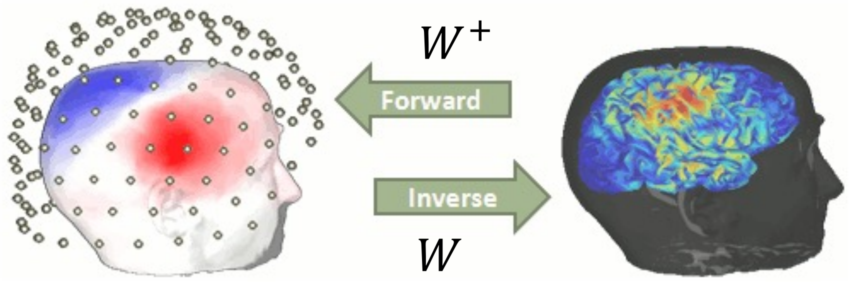
- Hz sampling

- e run consists
  al of 288 trials

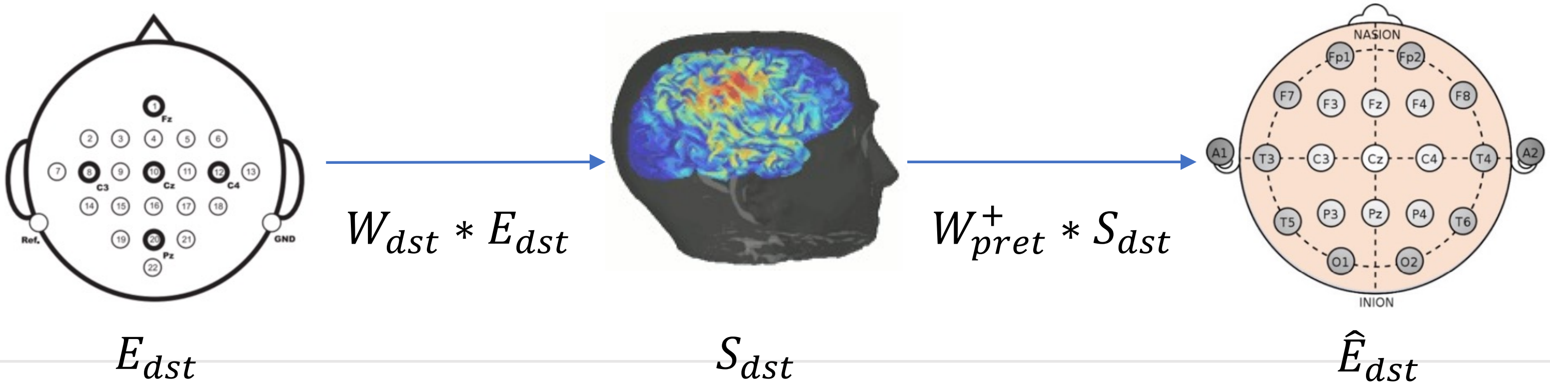


Align Downstream Data with Pretraining Data

- Different channel configurations for the pretraining and downstream data under the 10-20 system.
- Matched the channel configuration by mapping the downstream data to the channel configuration of the training data by re-sampling the signals using the source localization.
  a) Estimated linear inverse operator (inverse kernel) for the pretraining W_{pret} and downstream W_{dst} by solving the forward and inverse model.



  b) Re-sampling the downstream data



Downstream Task Finetuning

Leave-one-out cross validation, 9-fold

4-class classification.

All the trials from two sessions of the same subject are used for training and testing.

We add linear layers at the end of Neuro-GPT model to perform classification.

Three finetuning strategies:

- Finetune the whole neuro-gpt
- Finetune only the encoder
- Finetune only the GPT model.

Experiment Results

Average accuracy over 9-fold cross-validation

BENDR	Neuro-GPT	w/o Pretrain	w/ Pretrain
	Encoder+GPT	0.597 ± 0.093	0.628 ± 0.096
0.426	Encoder-only	0.606 ± 0.098	0.631 ± 0.089
	GPT-only	0.497 ± 0.061	0.507 ± 0.063

Table 1: Mean and std of 9-fold cross validation on downstream dataset

Conclusion and Future Works

- Pretraining a foundation model on large scale EEG dataset boosts downstream task performance where the data is heterogeneous and scarce.
- The foundation model encodes inherent and fundamental features of EEG that are generalizable across different datasets.
- Apply other large language models such as LLaMA, Alpaca, PaLM .....
- Evolve the EEG encoder to be channel-agnostic so it could adapt to varying numbers of channels and different montages.
- Incorporate other modalities such as fMRI to build a multi-modal Neuro-GPT.

Thanks!

Wenhui Cui[1], Woojae Jeong[1], Philipp Thölke[2], Takfarinas Medani[1], Karim Jerbi[2], Richard M. Leahy[1]

1. University of Southern California
2. University of Montreal