Title: *Brain2Real: A Multimodal Generative Framework for Visual and Auditory Synthesis from Brain Recordings*

1. Introduction

Deciphering and externalizing human thoughts, sensed through brain recordings, has long been a goal at the intersection of healthcare, neuroscience and artificial intelligence. Electroencephalography (EEG), as an inexpensive, non-invasive and temporally precise method of measuring brain activity, holds promise for real-time brain-computer interfaces (BCIs). While recent advances in neural decoding have enabled image reconstruction from fMRI or EEG, most existing approaches are limited to unimodal outputs and rely on cumbersome signal acquisition techniques. This proposal envisions a foundational generative model that decodes EEG signals into rich, temporally coherent multimodal content including images, videos, and speech, thus creating a generalized pathway from thought to media.

2. Objectives

- To develop foundational models for EEG signals that can represent EEG data for downstream generation tasks.
- To design a unified encoder for EEG signals that captures spatial-temporal neural patterns in a task-agnostic manner.
- To build modality-specific decoders (image, video, speech) conditioned on the learned EEG representation.
- To ensure the model is scalable, robust to noise, and generalizable across subjects and recording conditions.

3. Methodology

- **EEG Signal Encoding:**
  Use Transformer-based or spatio-temporal graph neural networks to encode EEG signals into a latent representation, preserving temporal dynamics and spatial topology. Training this model with EEG recordings that are publicly available, derived from various tasks.
- **Cross-Modal Generative Architecture:**
  Design parallel decoders for image, video, and speech synthesis, inspired by diffusion models or autoregressive transformers. Latent representations from EEG will condition each decoder via cross-attention or FiLM layers.
- **Multimodal Contrastive Pretraining:**
  Introduce a multimodal contrastive loss to align EEG representations with pre-trained embeddings from CLIP (for vision) and Whisper/HuBERT (for audio), enabling better grounding and data efficiency.

- **Alignment and Synchronization:**
  For video and speech generation, temporal alignment mechanisms (e.g., CTC loss or dynamic time warping) will be explored to maintain coherence with EEG signal dynamics.

- **Data Acquisition and Augmentation:**
  Leverage public datasets (e.g., BCI2VR, DEAP, MAHNOB-HCI) and, if feasible, collect a small-scale multi-subject dataset involving EEG paired with corresponding images, video stimuli, and spoken language.

4. Expected Contributions

- A unified generative framework for producing diverse media from EEG.

- A large-scale pretrained EEG encoder transferable to downstream neural decoding tasks.
- A new benchmark and evaluation protocol for multimodal EEG-to-media generation.
- Publications in top venues of neuroscience and/or machine learning.

5. Challenges and Mitigation Strategies

- *Noisy EEG Signals:* Use self-supervised denoising and adversarial robustness techniques.

- *Data Scarcity:* Rely on pretraining with aligned synthetic data, e.g., simulated EEG or neuro-symbolic augmentations.

- *Generalization Across Subjects:* Include domain adaptation and subject-specific fine-tuning layers.

6. Impact

Such a system would significantly enhance assistive communication tools, allowing speech- or motor-impaired individuals to express themselves through images or language. It could also deepen our understanding of neural correlates of imagination, perception, and language, contributing to cognitive science and neural prosthetics.

7. Timeline

- Year 1: Developing EEG representation learning models.

- Year 2: Developing decoder models for speech and video synthesis; multitask pretraining and joint learning.

- Year 3: Model scaling, cross-subject generalization, and benchmarking models for various multimodal generation tasks.

8. References

A tailored list can include works like:

- “Mind-Reading with Recurrent Neural Networks” (Tang et al., NeurIPS)
- “Diffusion Models for Brain Decoding”
- “EEG-GAN: Generative Adversarial Networks for EEG Synthesis”
- CLIP, Whisper, and related foundational multimodal models

9. Prerequisites

Background in Electrical Engineering, ECE, Computer Science, Data Science, Information Technology.

Keen Interest in Mathematical Modeling, Computational and Machine Learning Methods

Curiosity to explore newer topics and approaches.