

本研究所构建的数据集已公开发布于 GitHub 平台。

数据集 EEG_data 文件夹中包含五个主要文件夹（**data**、**assemble**、**segmentation**、**frequency**、**categorize**）以及一份受试者信息统计表（**Questionnaire_statistics.xlsx**）。

1. **data** 文件夹：包含 14 名受试者在“感兴趣”与“不感兴趣”实验条件下的脑电数据。每位受试者对应一个子文件夹，其中包含以下文件：

o original_data.txt：完整的 20 分钟原始 EEG 数据；

o intercept_data.mat：截取的中间 12 分钟数据；

o Pre-processing_data.csv：预处理后的 CSV 格式数据；

o eeglab_data.set 与 eeglab_data.fdt：可直接在 EEGLAB 中加载的 16 通道数据文件。

2. **assemble** 文件夹：对 Pre-processing_data.csv 中的数据进行通道补齐与格式规范化，便于后续处理与分析。

3. **categorize** 文件夹：基于频段能量与峰值频率特征，对受试者在“感兴趣”与“不感兴趣”状态下的数据进行分类标注，形成标签化特征集。

4. **frequency** 文件夹：对每段实验数据进行频谱分析，提取 16 通道的五个频段能量特征，包括：

o δ 波 (1 – 4 Hz)、θ 波 (4 – 8 Hz)、α 波 (8 – 13 Hz)、β 波 (13 – 30 Hz)、γ 波 (30 – 45 Hz)；
o 同时计算每个频段的峰值频率，即能量最大对应的频率点。

5. **segmentation** 文件夹：将每段实验数据按 10 秒无重叠窗口进行切割，生成标准化的时间片段，适用于时序分析与特征提取。

6. **number_10s_clips.xlsx** 文件：记录了所有数据预处理后有效分割的 10 秒片段数量。

工具包 EEG_Tool 文件夹包含以下内容：

1. **Subject Videos** 文件夹：存放实验中使用的视频材料，用于诱发受试者的学习兴趣状态。

2. **assemble.py**、**segmentation.py**、**frequency.py**、**categorize.py**：分别用于数据规整化、时间窗分割、频带能量分析与特征分类的 Python 脚本。

3. **extract_set_data.m**：用于提取 .set 文件中的 EEG 数据，适配 MATLAB 环境。

4. **Learning_interest_survey_questionnaire.pdf**：学习兴趣调查问卷文档，用于收集受者的主观兴趣评价。

5. **Informed_Consent_Form.pdf**：知情同意书，明确了研究背景目的、实验流程与配合要求、可能风险与受益、参与者权利及疑问反馈渠道以及对参与者的保护。

6. **Preprocessing_process_description.pdf**：eeglab 预处理流程详细操作说明。

7. **Questionnaire_statistics.xlsx**, **Assistance_to_downstream_analysis.xlsx**：记录了所有受者的基本信息（已做匿名处理去请）及其在实验过程中填写的问卷调查结果，用于辅助行为数据分析与个体差异研究。

8. **custom_16_electrodes.sfp**：实验的电极通道位置文件。

9. **eeglab2024.2.zip**：EEGLAB 工具箱安装包，是一款基于 MATLAB（部分支持 Octave）的开源脑电图（EEG）数据处理分析工具箱，广泛应用于神经科学研究。

该工具支持多种主流 EEG 数据格式导入，并提供从预处理到高级分析的完整流程，包括：

1. 预处理功能：伪迹去除（支持 ICA）、滤波、重参考、分段与基线校正；

2. 高级分析功能：独立成分分析（ICA）、时频分析、事件相关电位（ERP）分析、频谱分析及源定位（需配合其他工具包）；

3. 可视化功能：支持波形图、频谱图等多种形式的信号展示；

4. 操作方式：支持图形界面（GUI）与脚本化批处理，适用于不同水平的研究人员；

5.扩展性与社区支持：模块化设计，可通过插件扩展功能，拥有活跃的用户社区与详尽的文档资源。

该工具包与数据集共同构成了一个完整的分析体系，确保所有报告结果的可复现性，适用于学术研究、教学演示及临床辅助分析等多种场景。