

Cover Letter: Discussion Leader Presentation

Research Methods

April 7, 2026

Paper: Wei-Bang Jiang, Yansen Wang, Bao-Liang Lu, and Dongsheng Li. “NeuroLM: A Universal Multi-task Foundation Model for Bridging the Gap between Language and EEG Signals.” Published as a conference paper at ICLR 2025.

1. General Topic, Specific Activity, and Research Questions

General topic. Multi-task foundation models that bridge EEG signals and natural language for brain-computer interfaces.

Specific activity. The authors study human neurological activities across six EEG tasks: abnormal detection, event classification, emotion recognition, sleep staging, cognitive workload, and slowing event detection. Rather than requiring separate fine-tuning for each task, the authors propose treating EEG signals as a “foreign language” and integrating them into an LLM to handle all tasks within a single model via instruction tuning.

Research questions. (1) “Aligning EEG and text embeddings presents a great challenge.” (2) “How to learn generic information within the LLM paradigm remains an unsolved issue.” (3) “Developing a model that can seamlessly handle various tasks without compromising performance on any individual task is a major challenge.”

2. Problem, Claims, Evidence, and Statistical Analysis

Problem. Due to the variability of the EEG signal and tasks, prior EEG foundation models require fully fine-tuning on each downstream task separately. This wastes computational and storage resources and confines each model to a single task.

Claims. A single NeuroLM can handle six diverse BCI tasks simultaneously if aligning EEG embeddings to the text space via adversarial training, pre-training with multi-channel autoregression on 25,000 hours of EEG data, and applying multi-task instruction tuning, (Its largest variant (NeuroLM-XL) reaches 1.7B parameters, a record for EEG signal processing.)

Evidence. The authors evaluated NeuroLM on six downstream datasets: TUAB (abnormal detection), TUEV (event classification), SEED (emotion recognition), HMC (sleep staging), Workload (cognitive workload), and TUSL (slowing event classification). Ablation studies cover the necessity for EEG-text alignment (t-SNE visualization), multi-channel autoregressive pre-training, robustness of option switching, instruction data scaling, temporal vs. frequency reconstruction targets and pre-training epoch selection. Although NeuroLM generally underperforms SOTA, the authors acknowledge that comparison with single task baselines is “not entirely fair” since those baselines are individually fine-tuned.

Statistical analysis. Binary and Multi class classification results are reported as mean \pm standard deviation over three random seeds. Metrics include balanced accuracy, AUROC, AUC-PR, Cohen’s Kappa, and weighted F1.