

EEGAgent: A Unified Framework for Automated EEG Analysis Using Large Language Models

Sha Zhao^{1,2,*}, Mingyi Peng^{1,2,*}, Haiteng Jiang^{3,4,1}, Tao Li^{3,4,1}, Shijian Li^{1,2}

¹State Key Laboratory of Brain-machine Intelligence, Zhejiang University

²College of Computer Science and Technology, Zhejiang University

³Department of Neurobiology, Affiliated Mental Health Center & Hangzhou Seventh People's Hospital, Zhejiang University School of Medicine

⁴MOE Frontier Science Center for Brain Science and Brain-machine Integration, Zhejiang University
{szhao, mingyipeng, h.jiang, litaozjusc, shijianli}@zju.edu.cn

Abstract

Scalable and generalizable analysis of brain activity is essential for advancing both clinical diagnostics and cognitive research. Electroencephalography (EEG), a non-invasive modality with high temporal resolution, has been widely used for brain states analysis. However, most existing EEG models are usually tailored for individual specific tasks, limiting their utility in realistic scenarios where EEG analysis often involves multi-task and continuous reasoning. In this work, we introduce EEGAgent, a general-purpose framework that leverages large language models (LLMs) to schedule and plan multiple tools to automatically complete EEG-related tasks. EEGAgent is capable of performing the key functions: EEG basic information perception, spatiotemporal EEG exploration, EEG event detection, interaction with users, and EEG report generation. To realize these capabilities, we design a toolbox composed of different tools for EEG preprocessing, feature extraction, event detection, etc. These capabilities were evaluated on public datasets, and our EEGAgent can support flexible and interpretable EEG analysis, highlighting its potential for real-world clinical applications.

Code — <https://github.com/rebootingLine/EEGAgent>

Introduction

Brain diseases are increasingly recognized as a major global health challenge, often leading to profound functional impairments and progressive cognitive decline (Steinmetz et al. 2024). To support early diagnosis and effective intervention, the ability to accurately and efficiently assess brain function is of paramount importance. Among existing neurophysiological techniques, electroencephalography (EEG) stands out as a widely accessible, non-invasive tool that offers high temporal resolution and low operational cost. By capturing electrical signals generated by neuronal activity via scalp electrodes, EEG provides rich, multi-channel time-series data that encode dynamic information about both physiological processes and pathological changes. Despite its value, EEG analysis presents substantial challenges. The signals are highly non-stationary, exhibit considerable inter-subject variability, and contain complex waveform patterns that are

difficult to interpret (Shen and Lin 2019). Manual analysis by domain experts remains the standard in clinical practice, but this approach is time-consuming, labor-intensive, and inherently subjective. **These constraints severely limit the scalability and efficiency of EEG applications, especially in settings requiring continuous monitoring or multi-faceted analysis.**

Recent advances in machine learning (ML) and deep learning (DL) have enabled more automated and scalable approaches to EEG analysis (Zhou et al. 2025b). Modern data-driven models can learn meaningful representations directly from raw EEG signals and achieve high accuracy across tasks such as seizure detection, sleep staging, and neurological disorder diagnosis (Li et al. 2025). These approaches have led to notable improvements in analysis speed, accuracy, and reproducibility. **However, most existing EEG models remain task-specific and are optimized for isolated objectives.** This leads to the so-called task isolation problem, wherein each model addresses only a narrow slice of EEG analysis without considering the broader context or the interdependence of concurrent tasks. In clinical scenarios, EEG analysis often involves multi-task, continuous reasoning—for example, detecting epileptiform discharges while also assessing cognitive states or sleep transitions (Zhao et al. 2024). **Most current task-specific EEG models fail to meet this requirement, necessitating an urgent development of a unified and flexible framework to address this critical need.** Fortunately, recent breakthroughs in large language models (LLMs) offer a promising pathway toward overcoming these limitations (Zhang et al. 2024). With their powerful capabilities in task planning, contextual reasoning, and tool integration, LLM-powered agents can manage complex, multi-step workflows. They dynamically interpret user intent and interface with external modules, making them ideal candidates for scheduling multi-task EEG analysis in a coordinated and intelligent manner.

Inspired by this potential, we propose **EEGAgent**, a general-purpose, LLM-enhanced intelligent framework for EEG analysis. **EEGAgent is designed to bridge the gap between single-task models and multiple EEG analysis demands in the real world.** It achieves this by integrating perception, task understanding, and dynamic model scheduling into a unified system. EEGAgent brings together tra-

ditional feature engineering, multiple deep learning backbones, and an LLM-based policy planning engine to enable end-to-end coordination and execution of diverse EEG tasks. Through this integration, EEGAgent performs modular and adaptive scheduling, dynamically selecting appropriate pipelines based on the characteristics of the input data and the goals of the analysis. Furthermore, the system is equipped with domain-specific knowledge bases and the ability to interface with external information sources, enhancing its contextual awareness and interpretability. With this architecture, EEGAgent supports the entire workflow, from signal preprocessing and feature extraction to classification, decision reasoning, and interactive reporting. By providing an intelligent, multi-task-capable EEG analysis platform, our framework aims to advance the field toward greater automation, adaptability, and clinical readiness, helping unlock the full potential of EEG as a diagnostic and monitoring modality. Our key contributions are as follows:

- We propose the EEGAgent framework, which enables unified scheduling and automated execution of multi-task EEG analysis by integrating traditional and deep learning-based methods, thereby improving overall system efficiency and adaptability. To our best knowledge, we are the first to design an agent for EEG analysis.
- EEGAgent incorporates key EEG-specific capabilities such as context awareness, flexible spatiotemporal analysis, accurate event localization, and automated interactive reporting, facilitating comprehensive and adaptive EEG analysis and interpretation.
- We evaluate the capability of EEGAgent on public datasets, demonstrating its clinical potential through the support for interactive workflows and automated report generation, which can enhance diagnostic efficiency and assist clinical decision-making.

Related Works

EEG Analysis

Electroencephalography (EEG) is widely used in sleep staging, emotion recognition, seizure detection, and cognitive workload assessment, all of which require models capable of extracting informative patterns from noisy brain signals. Existing methods fall into two major categories: traditional machine learning with handcrafted features and deep learning models that learn directly from raw data.

Traditional ML approaches remain useful where efficiency or interpretability is important. Naive Bayes has been applied to seizure and drowsiness detection (Ashok and Purushothaman 2016), Random Forests to mental workload and emotion recognition (Messaoud and Chavez 2021), and Linear Discriminant Analysis (LDA) to motor imagery-based BCIs (Fu et al. 2020; Santos, San Martin, and Fraga 2023). Other classical classifiers such as SVM, KNN, and Logistic Regression are frequently used as baselines or in hybrid systems (Saeidi et al. 2021). Deep learning methods have shown strong performance by automatically extracting spatial-temporal features from EEG signals (Zhou et al. 2025a; Wang et al. 2024). CNNs excel in sleep staging and motor imagery (Efe and Ozsen 2023; Khademi,

Ebrahimi, and Kordy 2022), RNNs (e.g., LSTM, GRU) perform well in sequential tasks such as seizure and emotion detection (Mekruksavanich and Jitpattanakul 2023; Zhao et al. 2025; Wang et al. 2025b), and Transformer-based models further improve the modeling of long-range dependencies (Sun, Xie, and Zhou 2021; Wei et al. 2023). However, most existing EEG models are designed for isolated tasks under controlled conditions, which limits their practical applicability. In real-world scenarios, tasks such as artifact removal, event detection, and signal classification often need to be performed together, underscoring the need for a unified, context-aware framework that can flexibly support multi-task EEG analysis.

LLM-Enhanced Agent Framework

Recent advances in large language models (LLMs) have enabled the creation of intelligent agents that can perform complex reasoning and manage multi-step tasks (Wang et al. 2025a). Originally designed for next-token prediction, LLMs now demonstrate impressive capabilities in language understanding, planning, and decision-making (Brown et al. 2020; Radford et al. 2019). When augmented with modules for planning, memory, perception, and tool use, they can coordinate workflows, invoke external tools, and synthesize information to support high-level decision-making (Yao et al. 2023b; Shinn et al. 2023). Techniques such as Chain of Thought (CoT) (Wei et al. 2022) and Tree of Thought (ToT) (Yao et al. 2023a) enhance LLM reasoning via stepwise decomposition and hypothesis exploration. Planning modules help break down goals into subtasks, while memory mechanisms enable both contextual tracking and long-term knowledge retention (Huang et al. 2022; Wang et al. 2023; Wu et al. 2022). Retrieval-Augmented Generation (RAG) further improves reasoning by allowing access to external knowledge bases such as clinical guidelines (Lewis et al. 2020).

However, directly applying LLM agents to EEG analysis poses unique challenges. EEG signals are highly variable and non-stationary, and they require domain-specific interpretation. Tailoring LLMs to accommodate these complexities is essential to ensure their utility in EEG-related applications. While most existing EEG models remain task-specific, recent trends in LLM-driven agent frameworks offer a new direction. These agents leverage planning, memory, and tool integration to coordinate analytic workflows and generate structured outputs. Extending such capabilities to EEG analysis could enable more flexible, multi-task solutions that align with clinical workflows.

Method

The EEGAgent is an intelligent system designed to autonomously plan and execute EEG analysis tasks. As illustrated in Fig. 1, this framework leverages a series of tools that automate the execution of these tasks, thereby forming several distinctive capabilities tailored to diverse EEG analysis scenarios. Specifically, the agent demonstrates the following core competencies:

- **Perception:** The EEGAgent is aware of the EEG data context, including subject metadata, available channels,

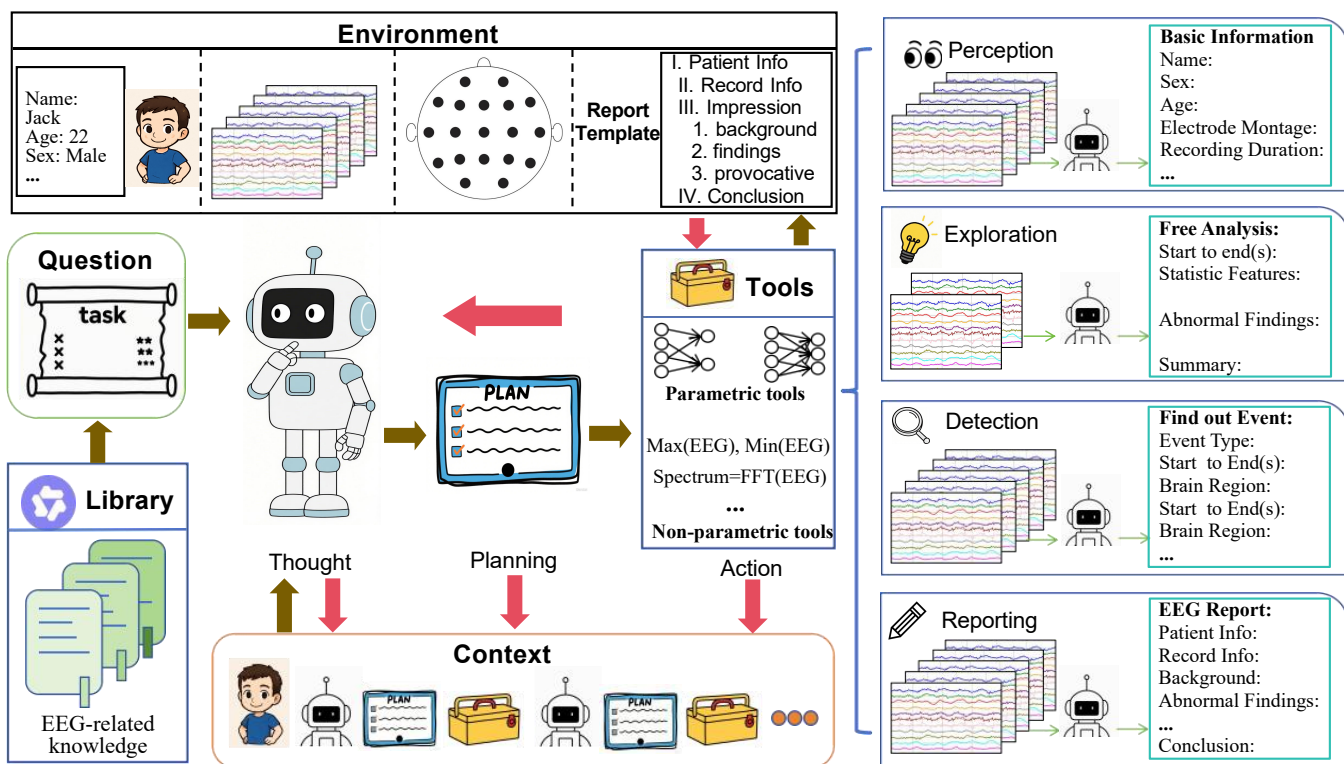


Figure 1: EEGAgent framework

and the temporal span of the recording.

- **Exploration:** The EEGAgent supports in-depth analysis within user-specified or system-identified spatiotemporal ranges. It flexibly schedules appropriate tools—such as frequency analysis, waveform extraction, or statistical characterization—to examine targeted EEG segments from multiple perspectives.
- **Detection:** The EEGAgent can accurately identify the time segments and brain regions where specific EEG events occur and further decode brain states, enabling efficient retrieval and matching.
- **Interaction and Reporting:** The EEGAgent supports interaction with users and automatically generates structured EEG reports based on predefined templates.

To support these capabilities, the EEGAgent integrates a modular toolbox of parametric models and statistical feature extractors. Guided by a large language model, it dynamically plans and sequences tool usage based on task needs and context. This design enables flexible, context-aware analysis—from broad perception to precise localization—along with automated and structured report generation.

EEGAgent Architecture

The EEGAgent is designed to operate in a complex environment consisting of subject information, EEG recording metadata, raw signals, and report templates. It is built on Qwen3-235B (Yang et al. 2025), a large language model

known for its strong reasoning capabilities and seamless integration with external tools, hereafter referred to as Qwen. To provide domain-specific expertise, the Agent incorporates an EEG knowledge base. When a new task arrives, the task description is embedded into a high-dimensional semantic vector using Qwen3-Embedding-8B (Zhang et al. 2025), and relevant knowledge entries are retrieved from the knowledge base via similarity search. The retrieved information is then used to enrich the context for Qwen. The Agent’s toolbox contains two types of tools: parametric tools based on deep learning models and non-parametric tools that utilize handcrafted statistical features. Upon receiving a task, Qwen combines the enriched context and the current EEG data to analyze the problem, formulate a plan, and select the appropriate tools. These tools are then executed to extract the necessary information, which is fed back to Qwen for further reasoning. This iterative cycle continues until the task is satisfactorily completed. **By integrating knowledge retrieval, memory, and a versatile toolbox under the control of Qwen, the EEGAgent provides a flexible and interpretable framework capable of handling diverse EEG analysis tasks in an automated manner.**

Toolbox Design for Capability Support

To realize the core capabilities of the EEGAgent, the toolbox is organized along three independent dimensions: parametric vs. non-parametric nature, temporal granularity, and spatial granularity. These dimensions characterize each tool’s functional attributes and computational cost, enabling flexi-

Tool Name	Type	Time Granularity	Space Granularity	Description
normalAbnormal	Parametric Tool	Full EEG	Whole Channel	Estimates probability of pathological normality/abnormality for the entire EEG.
eyemMuscle	Parametric Tool	Single Second	Single Channel	Classifies eye movement and muscle artifacts within 1-second window per channel.
seizArtiBckg	Parametric Tool	Single Second	Single Channel	Classifies seizure, artifact, and background within 1-second window per channel.
seizNormal	Parametric Tool	Single Second	Single Channel	Detects seizure vs. non-seizure within 1-second window per channel.
slowSeizBckg	Parametric Tool	10 Seconds	Whole Channel	Classifies slow waves, epileptic and background activity in 10-second windows.
baseInfo	Non-parametric Tool	Full EEG	Whole Channel	Extract structured information such as patient demographics and recording metadata from the EEG.
compute_amplitude	Non-parametric Tool	≤ 60 Seconds	Whole Channel	Computes amplitude features (mean abs, RMS, max/min) for selected channels.
compute_psd	Non-parametric Tool	≤ 60 Seconds	Whole Channel	Calculates power spectral density across frequency bands for selected channels.
compute_symmetry	Non-parametric Tool	≤ 60 Seconds	Left-Right Channel Pair	Calculates Pearson correlation for left-right channel pairs to assess symmetry.

Table 1: Tools and their time-space granularity.

ble scheduling and intelligent composition. Parametric deep learning tools capture high-level clinical semantics for complex tasks, whereas non-parametric statistical and signal-processing tools provide efficient extraction of fundamental features. Temporal granularity ranges from coarse long-window analysis to fine short-segment detection, and spatial granularity spans global multi-channel assessment to single-channel localization.

The Agent flexibly schedules and combines these tools to meet task requirements. For instance, non-parametric tools such as `baseInfo` support macro-level perception, while parametric tools such as `seizNormal` and `slowSeizBckg` enable precise event localization. By dynamically integrating tools of different natures and granularities, the Agent achieves multi-perspective analyses and balances accuracy with efficiency. The resulting multidimensional information is further fused to drive interaction and report generation, ultimately producing structured and clinically meaningful EEG outputs. A summary of the tools and their temporal–spatial characteristics is shown in Table 1.

EEG Perception Ability: Comprehending the Fundamental Information of EEG Signals

Perception capability refers to that the EEGAgent can understand the EEG data environment in which it operates. Given the specific characteristics of EEG data, the EEGAgent can not only access basic information but also extract critical advanced features, enabling it to build a comprehensive perceptual model of EEG recordings. EEG data inherently contains rich contextual information, such as electrode configuration, recording duration, and subject status. To enhance the Agent’s understanding of such background details, **we developed a compact knowledge module encapsulated as the `baseInfo` tool, which extracts key metadata upon EEG loading.** This tool is automatically triggered during the initial loading phase (i.e., environment initialization), helping the Agent to construct a foundational perceptual view.

Within this module, the system extracts essential subject information (e.g., name, gender, age), the total recording duration, and the electrode montage. Among these, age plays a particularly crucial role in EEG interpretation. To improve the Agent’s sensitivity to age-related factors, we incorporate summarized insights from clinical literature regarding normative EEG patterns across developmental stages (Liu 2011), allowing the Agent to apply age-appropriate reference standards. Besides, the spatial organization of electrodes and their mapping to anatomical brain regions are presented in a structured tabular format, enabling the Agent to understand spatial relationships among channels, including symmetry and brain regional affiliation. We also provide a concise summary of typical frequency bands and their associated pathological features to serve as cognitive references. Since foundational information such as age-related factors and electrode configurations is independent of specific tasks, it can be preloaded before any analysis begins. **This enables the EEGAgent to form an initial expectation of EEG patterns,** effectively simulating the anticipatory cognition that human experts develop prior to interpreting EEG signals.

EEG Exploration Ability: Analyzing EEG Segments from Multiple Perspectives

The EEGAgent’s exploration ability enables it to conduct flexible and systematic analysis of any task-specified time interval within EEG recordings. To model the EEGAgent’s exploration capability, we consider a temporal interval $[T_{\text{start}}, T_{\text{end}}]$, which is partitioned into N non-overlapping segments of length Δt :

$$X_i = [T_{\text{start}} + (i-1)\Delta t, T_{\text{start}} + i\Delta t], \quad i = 1, \dots, N. \quad (1)$$

For each segment X_i , the Agent selects an analysis plan \mathcal{T}_i from the tool library via a Qwen-based controller π_θ , conditioned on contextual information C (such as patient age, prior events, modality priors):

$$\mathcal{T}_i = \pi_\theta(X_i, C) \subseteq \mathcal{T}_p \cup \mathcal{T}_{np}, \quad (2)$$

where \mathcal{T}_p and \mathcal{T}_{np} represent sets of parametric and non-parametric tools, respectively. Each tool $T \in \mathcal{T}_i$ is treated as a function:

$$T : X_i \mapsto R_i^T, \quad R_i^T \in \mathbb{R}^{d^T} \text{ or } \mathcal{V}^*, \quad (3)$$

where R_i^T denotes the tool’s output, such as feature vectors, classification scores, or symbolic descriptors. To synthesize tool outputs for each segment, a fusion function f_{fuse} integrates intermediate results:

$$R_i = f_{\text{fuse}}(\{R_i^T\}; \phi), \quad \phi = \text{fusion parameters.} \quad (4)$$

Finally, a summarization function $f_{\text{summarize}}$ aggregates results across all segments into a comprehensive semantic summary \mathcal{S} :

$$\mathcal{S} = f_{\text{summarize}}(\{R_i\}_{i=1}^N; C), \quad (5)$$

which includes global assessments, rhythmic patterns, and localized event characterizations. The entire exploration workflow is coordinated by a large language model. It handles every step, from selecting tools to fusing outputs and generating summaries. Rather than relying on fixed rules, the Agent interprets tool results based on context. This approach resolves representational mismatches and allows the system to generate high-level insights through language-based reasoning. As a result, the EEGAgent can explore EEG data like a human expert, with a balance between interpretability and efficiency across different analysis goals.

EEG Detection Ability: Recognizing Targeted Spatiotemporal EEG Event

The EEGAgent is capable of automatically identifying specific clinically relevant EEG event types during review, and precisely localizing these events in both temporal and spatial domains throughout the recording. Given the non-stationary nature of EEG signals and the dynamic variation of their statistical properties over time, analyses conducted at a single temporal scale are insufficient to accurately capture event features. To address this challenge, **we adopt a multi-granularity analysis strategy** that improves detection accuracy and efficiency while reasonably controlling computational costs. In the temporal dimension, the EEGAgent utilizes two time scales: 10 seconds and 1 second. The 10-second window corresponds to the typical duration used in clinical review and serves as a coarse screening tool to determine whether a target event occurs within that period. Upon detection of a potential event, the EEGAgent further switches to a finer 1-second scale to conduct detailed analysis, achieving more precise spatiotemporal localization. In the spatial dimension, the EEGAgent employs both single-channel and multi-channel analyses, enabling it to focus on local signal variations at individual electrodes as well as integrate information across multiple channels to enhance robustness in event detection.

Notably, the entire analysis process is based on a sliding window mechanism and is dynamically planned and scheduled by Qwen. It adjusts the analysis strategy according to task requirements and intermediate results, balancing the trade-off between analysis granularity and computational cost. This flexible guidance enables the EEGAgent

Algorithm 1: EEG Report Generation

Input: Raw EEG data D

Parameter: Tool list \mathcal{L}

Output: Final EEG report R

```

1: Initialize report template  $\mathcal{T}$ 
2:  $\mathcal{S} \leftarrow \text{Segment}(D, 10s)$  {Divide EEG into 10s windows}
3: for each segment  $s_i \in \mathcal{S}$  do
4:    $r_i^{\text{coarse}} \leftarrow \text{Analyze}(s_i, \text{coarse})$  {Coarse-level analysis}
5:   if Qwen_Decide( $r_i^{\text{coarse}}$ ) = fine then
6:     for each subsegment  $f_{i,j} \in \text{Segment}(s_i, 1s)$  do
7:        $r_{i,j}^{\text{fine}} \leftarrow \text{Analyze}(f_{i,j}, \text{fine})$ 
8:        $\mathcal{T} \leftarrow \mathcal{T} \cup r_{i,j}^{\text{fine}}$ 
9:     end for
10:  else
11:     $\mathcal{T} \leftarrow \mathcal{T} \cup r_i^{\text{coarse}}$ 
12:  end if
13: end for
14:  $R \leftarrow \text{GenerateReport}(\mathcal{T})$  {Qwen or template-based synthesis}
15: return  $R$ 

```

to progressively refine its analysis from coarse to fine granularity. As a result, event detection and localization are performed efficiently and economically without compromising accuracy.

Interaction and Reporting Ability: Interacting with Users and Generating Clinical EEG Reports

EEGAgent enables lightweight user interaction by remembering context, allowing users to ask follow-up questions, request specific analyses, or clarify ambiguous results. This conversational ability improves interpretability and aligns with clinical workflows. Building on this interaction, EEGAgent can automatically generate structured reports to assist neurologists, summarizing both global patterns and local abnormalities through a hierarchical, multi-scale analysis pipeline. To support this goal, EEGAgent adopts a modular reporting strategy based on the ACNS standardized template (Tatum et al. 2016), decomposing the report into components such as background activity, abnormal events, and clinical impressions. This design ensures consistency and interpretability while accommodating variability across clinical cases. As shown in Algorithm 1, the input EEG is first segmented into 10-second windows, aligned with standard clinical practice. Each segment is analyzed coarsely to identify salient events. A lightweight Qwen controller determines whether finer 1-second-scale analysis is warranted, allowing the system to focus on diagnostically relevant regions without overreacting to noise or isolated outliers. The resulting observations from both granularities are compiled into a structured template and converted into final text using either template-based generation or Qwen-assisted summarization. This unified workflow enables EEGAgent to generate clinically aligned, robust, and interpretable reports that support multi-level reasoning over complex EEG signals.

Experiments

Dataset

We evaluate the EEGAgent on five widely used EEG datasets, including three representative subsets of the Temple University Hospital (TUH) EEG corpus and two additional public datasets. **TUH Abnormal EEG Corpus (TUAB)** (López et al. 2015) contains 2,993 recordings of about 20 minutes each (10–20 montage, 250 Hz), labeled as normal (1,521) or abnormal (1,472). Recording-level annotations make TUAB suitable for global abnormality classification. **TUH EEG Event Corpus (TUEV)** (Obeid and Picone 2016) includes 518 recordings with 1-second, per-channel annotations for six event types (PLED, GPED, SPSW, EYEM, ARTF, BCKG), enabling fine-grained temporal–spatial localization. It supports tasks such as seizure detection, artifact discrimination, and multi-class event classification. We use TUEV to evaluate the event-level detection capability of EEGAgent. **TUH EEG Slowing Corpus (TUSL)** (Obeid and Picone 2016) contains 300 ten-second segments from 75 sessions (38 patients), labeled for slowing versus background or seizure activity. Data are recorded with the TCP-REF montage at 256 Hz, with segment-level annotations across all channels. **The Sleep-EDF database** (Goldberger et al. 2000) Contains 197 full-night sleep recordings with EEG, EOG, and EMG signals. This dataset features manually created annotations, offering high-quality sleep stage labels for algorithmic research. **MDD Patients and Healthy Controls EEG Data** (Mumtaz 2016) comprises EEG recordings from 34 MDD patients and 30 healthy controls, recruited from the outpatient clinic at Hospital Universiti Sains Malaysia (HUSM), with diagnoses based on DSM-IV criteria (Mumtaz et al. 2017). Together, these datasets provide a comprehensive testbed for multi-scale, multi-task EEG analysis, supporting both global and localized objectives under varying levels of annotation granularity.

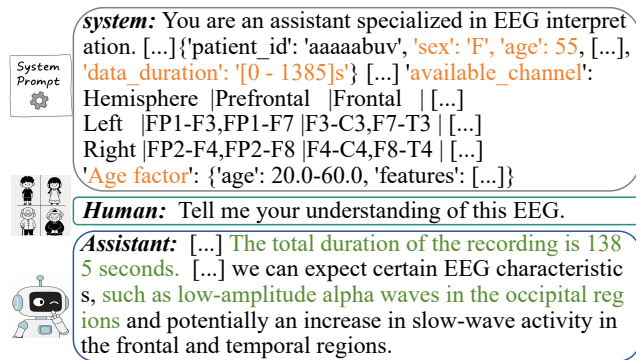


Figure 2: The EEGAgent perceives the environment constructed from EEG data. Solid lines represent processes visible to the user.

EEG Perception Experiment

We evaluate the perceptual capability of the EEGAgent on TUAB. This experiment is designed to assess whether the

Agent can accurately extract contextual metadata and integrate prior neurophysiological knowledge into its interpretation process. As shown in Figure 2, the Agent is provided with structured information, including patient demographics (e.g., sex and age), recording parameters (e.g., start date, start time, and total duration), and spatial channel mappings.

The evaluation focuses on the Agent’s ability to extract key metadata such as the subject’s age and the recording duration, and subsequently apply these contextual cues when interpreting the EEG background activity. In this case, the patient’s age informs the Agent’s expectations regarding typical EEG characteristics, including slowing of background rhythms, reduced fast activity, the presence of low-amplitude occipital alpha rhythms, and a possible increase in frontal and temporal slow-wave activity. This reasoning process leverages well-established age-related EEG patterns (Liu 2011). The objective of this evaluation is to determine whether the EEGAgent can demonstrate context-driven perceptual capability by combining demographic information, spatial channel knowledge, and neurophysiological priors to guide its understanding of the EEG environment.

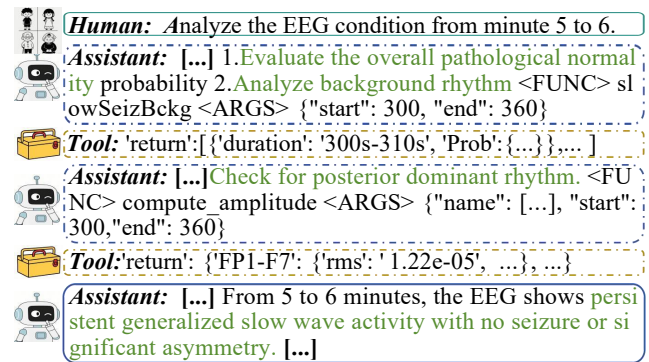


Figure 3: Exploration of EEG segments by the EEGAgent. Solid lines represent processes visible to the user.

EEG Segment Exploration Experiment

To evaluate the Agent’s exploratory capability, we present a case study from the TUAB corpus in response to the non-specific query: “Please analyze the EEG condition from minute 5 to 6.” As shown in Figure 3, the Agent autonomously executes a structured, multi-stage workflow. First, it deploys the parametric tool `slowSeizBckg` for a rapid assessment, which detects slow-wave activity with a high probability. Based on this preliminary result, the Agent initiates a deeper analysis using the non-parametric tool `compute_amplitude` to precisely characterize the spatial distribution of the detected waves. In the final stage, the Agent synthesizes multiple sources of evidence: the high-level classification (slow wave), low-level quantitative features (indicating a generalized pattern), and relevant patient context (elderly). This integrative reasoning produces a clinically meaningful report identifying *persistent generalized slow-wave activity*, which warrants further clinical correlation. This case study **highlights the Agent’s ability to schedule a coarse-to-fine sequence of analytical**

tools, effectively bridging the gap between raw EEG data and expert-level interpretation.

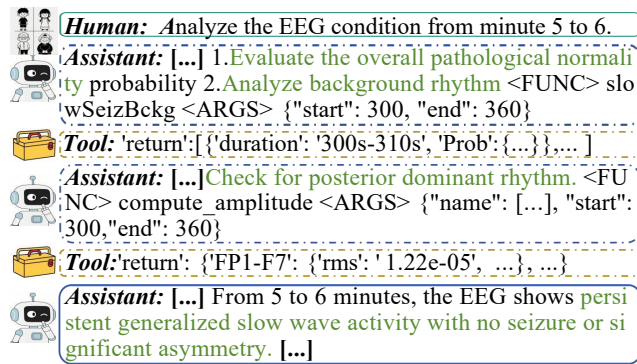


Figure 4: Detection of Epileptic Discharges by the EEGAgent. Solid lines represent processes visible to the user.

EEG Event Detection Experiment

Detection ability involves not only identifying the temporal intervals of EEG events but also localizing their spatial distribution. Because many clinically relevant events are transient and spatially variable, EEGAgent adopts a hierarchical multi-scale strategy to balance accuracy and efficiency. As shown in Figure 4, the system first performs a coarse-grained scan (e.g., 60-s windows) to quickly locate suspicious intervals, and then progressively focuses on these intervals using higher-resolution tools to refine temporal and spatial localization. In the example, an initial elevated seizure probability at 0–10 s was further resolved at the channel level, ultimately pinpointing the right fronto-central region (F4–C4) at 0–1 s as the likely onset site.

We evaluated the EEGAgent on the TUEV dataset for seizure detection, using SPSW, GPED, and PLED as positive classes. Annotations were preprocessed by merging adjacent events occurring within 1 second, and predictions were considered correct if their IoU with reference events exceeded 0.7 on the same channel. To investigate the effect of LLM backbone size, we tested models of different scales: Qwen3 235B achieved a hit rate of 69.30% and a false alarm rate of 44.77%, 32B achieved 60.84% and 55.66%, and 14B achieved 58.04% and 60.02%, respectively. We also measured computational efficiency, showing that the average number of tool calls per task and per-call runtime were as follows: Qwen3-32B (2.01 calls, 12.5 s), and Qwen3-14B (1.93 calls, 10.6 s). These results demonstrate that the step-wise localization strategy enables accurate detection of transient epileptiform events while maintaining low computational cost, even with smaller LLM backbones.

EEG Report Generation Experiment

We conducted a feasibility evaluation of the EEGAgent’s expression capability using the **TUAB dataset**, a large-scale clinical EEG dataset reflecting real-world information distribution. Due to the lack of paired reports, **we designed an automatic report generation task** based on ACNS stan-

dards (Tatum et al. 2016) to assess the system’s ability to translate signal analysis into clinical language.

In this task, the EEGAgent receives a full EEG recording and generates a structured report covering basic information, background activity, abnormal events, and diagnostic conclusions. As shown in Figure 5, the Agent accurately extracts patient metadata, identifies background rhythms and their spatial distribution, and detects abnormal events at a coarse level (e.g., describing occipital alpha slowing, diffuse slow waves, and left fronto-central epileptiform discharges). This experiment demonstrates that **the EEGAgent is capable of generating clinically interpretable EEG reports, covering key information elements and supporting abnormality detection and localization**. Although quantitative evaluation remains challenging, the TUAB-based validation suggests the proposed approach is feasible for automated EEG interpretation.

To assess the generalizability of EEGAgent, we further evaluate it on depression recognition and sleep-stage classification. On the MDD dataset, it achieves a subject-level accuracy of 83.33% and an F1-score of 84.34%. On Sleep-EDF, it attains an accuracy of 82.31% and a Macro-F1 of 76.27%, demonstrating strong adaptability across diverse EEG tasks.

EEG Report generated by EEGAgent					
Patient ID	aaaaabuv	Age	55	Sex	Female
Date	2017/9/12	Duration	1385s	Report Date	
History		State	awake	Montage	10-20
Alpha Activity	The posterior dominant rhythm is a low-amplitude 10 Hz alpha, which is typical for normal aging.				
Waves	Background activity shows intermittent diffuse slowing and moderately asymmetric posterior rhythms.				
Abnormal Findings	Persistent diffuse theta-delta slowing. Focal epileptiform discharges in left fronto-central region, with brief intermittent epileptiform activity elsewhere and low seizure probability.				
Conclusion	Diffuse slowing and left fronto-central epileptiform discharges present without seizures recorded.				
Reported by	EEGAgent	Reviewed by		Review Date	

Figure 5: EEG report generated by the EEGAgent

Conclusion

In this work, we present EEGAgent, the first framework that integrates intelligent agent systems with EEG analysis. It employs a flexible toolbox operating across multiple temporal and spatial scales to support context-aware analysis. Evaluated on public datasets, EEGAgent demonstrates core capabilities including perception, exploration, detection, interaction, and report generation. The framework improves efficiency, scalability, and interpretability, highlighting the potential of LLM-driven agents for automated EEG interpretation and clinical support.

Acknowledgments

This work was supported by STI 2030 Major Projects (2021ZD0200400), the National Natural Science Foundation of China (No.62476240), the Fundamental Research Funds for the Central Universities (226-2025-00122) and the Key Program of the Natural Science Foundation of Zhejiang Province, China (No. LZ24F020004). The corresponding author is Dr. Sha Zhao.

References

- Ashok, S.; and Purushothaman, G. 2016. DWT based Epileptic Seizure Detection from EEG Signals using Naïve Bayes/k-NN Classifiers. *IEEE Access*, 4: 1–1.
- Brown, T. B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; Agarwal, S.; Herbert-Voss, A.; Krueger, G.; Henighan, T.; Child, R.; Ramesh, A.; Ziegler, D. M.; Wu, J.; Winter, C.; Hesse, C.; Chen, M.; Sigler, E.; Litwin, M.; Gray, S.; Chess, B.; Clark, J.; Berner, C.; McCandlish, S.; Radford, A.; Sutskever, I.; and Amodei, D. 2020. Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781713829546.
- Efe, E.; and Ozsen, S. 2023. CoSleepNet: Automated sleep staging using a hybrid CNN-LSTM network on imbalanced EEG-EOG datasets. *Biomedical Signal Processing and Control*, 80: 104299.
- Fu, R.; Han, M.; Tian, Y.; and Shi, P. 2020. Improvement motor imagery EEG classification based on sparse common spatial pattern and regularized discriminant analysis. *Journal of Neuroscience Methods*, 343: 108833.
- Goldberger, A.; Amaral, L.; Glass, L.; Hausdorff, J.; Ivanov, P. C.; Mark, R.; and Stanley, H. E. 2000. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23): e215–e220. RRID:SCR_007345.
- Huang, W.; Abbeel, P.; Pathak, D.; and Mordatch, I. 2022. Language Models as Zero-Shot Planners: Extracting Actionable Knowledge for Embodied Agents. arXiv:2201.07207.
- Khademi, Z.; Ebrahimi, F.; and Kordy, H. M. 2022. A transfer learning-based CNN and LSTM hybrid deep learning model to classify motor imagery EEG signals. *Computers in biology and medicine*, 143: 105288.
- Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.-t.; Rocktäschel, T.; Riedel, S.; and Kiela, D. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems*, volume 33, 9459–9474. Curran Associates, Inc.
- Li, J.; Chen, X.; Shen, F.; Chen, J.; Liu, Y.; Zhang, D.; Yuan, Z.; Zhao, F.; Li, M.; and Yang, Y. 2025. Deep Learning-Powered Electrical Brain Signals Analysis: Advancing Neurological Diagnostics. arXiv:2502.17213.
- Liu, X. Y. 2011. *Clinical Electroencephalography Training Tutorial*. Beijing, China: People's Medical Publishing House, 1st edition. Nov. 2011.
- López, S.; Suarez, G.; Jungreis, D.; Obeid, I.; and Picone, J. 2015. Automated identification of abnormal adult EEGs. In *2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 1–5.
- Mekruksavanich, S.; and Jitpattanukul, A. 2023. Effective detection of epileptic seizures through EEG signals using deep learning approaches. *Machine Learning and Knowledge Extraction*, 5(4): 1937–1952.
- Messaoud, R. B.; and Chavez, M. 2021. Random forest classifier for EEG-based seizure prediction. *arXiv preprint arXiv:2106.04510*.
- Mumtaz, W. 2016. MDD patients and healthy controls EEG data (new). Figshare, Dataset.
- Mumtaz, W.; Xia, L.; Ali, S. S. A.; Yasin, M. A. M.; Hus-sain, M.; and Malik, A. S. 2017. Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD). *Biomedical Signal Processing and Control*, 31: 108–115.
- Obeid, I.; and Picone, J. 2016. The Temple University Hospital EEG Data Corpus. *Frontiers in neuroscience*, 10: 196.
- Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; and Sutskever, I. 2019. Language Models are Unsupervised Multitask Learners.
- Saeidi, M.; Karwowski, W.; Farahani, F. V.; Fiok, K.; Taiar, R.; Hancock, P. A.; and Al-Juaid, A. 2021. Neural decoding of EEG signals with machine learning: a systematic review. *Brain sciences*, 11(11): 1525.
- Santos, E.; San Martin, R.; and Fraga, F. 2023. Comparison of subject-independent and subject-specific EEG-based BCI using LDA and SVM classifiers. *Medical & biological engineering & computing*, 61.
- Shen, Y.-W.; and Lin, Y.-P. 2019. Challenge for Affective Brain-Computer Interfaces: Non-stationary Spatio-spectral EEG Oscillations of Emotional Responses. *Frontiers in human neuroscience*, 13: 366.
- Shinn, N.; Cassano, F.; Gopinath, A.; Narasimhan, K.; and Yao, S. 2023. Reflexion: language agents with verbal reinforcement learning. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Advances in Neural Information Processing Systems*, volume 36, 8634–8652. Curran Associates, Inc.
- Steinmetz, J. D.; Seeher, K. M.; Schiess, N.; Nichols, E.; Cao, B.; Servili, C.; Cavallera, V.; Cousin, E.; Hagins, H.; Moberg, M. E.; et al. 2024. Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet Neurology*, 23(4): 344–381.
- Sun, J.; Xie, J.; and Zhou, H. 2021. EEG Classification with Transformer-Based Models. In *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*, 92–93.
- Tatum, W. O.; et al. 2016. ACNS Guideline 7: Guidelines for EEG Reporting. *Journal of Clinical Neurophysiology*, 33(4): 328–332.

- Wang, G.; Xie, Y.; Jiang, Y.; Mandlekar, A.; Xiao, C.; Zhu, Y.; Fan, L.; and Anandkumar, A. 2023. Voyager: An Open-Ended Embodied Agent with Large Language Models. In *NeurIPS 2023 Foundation Models for Decision Making Workshop*.
- Wang, J.; Zhao, S.; Jiang, H.; Li, S.; Li, T.; and Pan, G. 2024. Generalizable Sleep Staging via Multi-Level Domain Alignment. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(1): 265–273.
- Wang, J.; Zhao, S.; Luo, Z.; Zhou, Y.; Jiang, H.; Li, S.; Li, T.; and Pan, G. 2025a. CBraMod: A Criss-Cross Brain Foundation Model for EEG Decoding. arXiv:2412.07236.
- Wang, Y.; Zhao, S.; Jiang, H.; Li, S.; Li, T.; and Pan, G. 2025b. M-MDD: A multi-task deep learning framework for major depressive disorder diagnosis using EEG. *Neurocomputing*, 636: 130008.
- Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; ichter, b.; Xia, F.; Chi, E.; Le, Q. V.; and Zhou, D. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 24824–24837. Curran Associates, Inc.
- Wei, Y.; Liu, Y.; Li, C.; Cheng, J.; Song, R.; and Chen, X. 2023. TC-Net: A Transformer Capsule Network for EEG-based emotion recognition. *Computers in biology and medicine*, 152: 106463.
- Wu, Y.; Rabe, M. N.; Hutchins, D.; and Szegedy, C. 2022. Memorizing Transformers. arXiv:2203.08913.
- Yang, A.; Li, A.; Yang, B.; Zhang, B.; Hui, B.; Zheng, B.; Yu, B.; Gao, C.; Huang, C.; Lv, C.; Zheng, C.; Liu, D.; Zhou, F.; Huang, F.; Hu, F.; Ge, H.; Wei, H.; Lin, H.; Tang, J.; Yang, J.; Tu, J.; Zhang, J.; Yang, J.; Yang, J.; Zhou, J.; Zhou, J.; Lin, J.; Dang, K.; Bao, K.; Yang, K.; Yu, L.; Deng, L.; Li, M.; Xue, M.; Li, M.; Zhang, P.; Wang, P.; Zhu, Q.; Men, R.; Gao, R.; Liu, S.; Luo, S.; Li, T.; Tang, T.; Yin, W.; Ren, X.; Wang, X.; Zhang, X.; Ren, X.; Fan, Y.; Su, Y.; Zhang, Y.; Zhang, Y.; Wan, Y.; Liu, Y.; Wang, Z.; Cui, Z.; Zhang, Z.; Zhou, Z.; and Qiu, Z. 2025. Qwen3 Technical Report. arXiv:2505.09388.
- Yao, S.; Yu, D.; Zhao, J.; Shafran, I.; Griffiths, T.; Cao, Y.; and Narasimhan, K. 2023a. Tree of Thoughts: Deliberate Problem Solving with Large Language Models. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Advances in Neural Information Processing Systems*, volume 36, 11809–11822. Curran Associates, Inc.
- Yao, S.; Zhao, J.; Yu, D.; Du, N.; Shafran, I.; Narasimhan, K. R.; and Cao, Y. 2023b. ReAct: Synergizing Reasoning and Acting in Language Models. In *The Eleventh International Conference on Learning Representations*.
- Zhang, X.; Chowdhury, R. R.; Gupta, R. K.; and Shang, J. 2024. Large language models for time series: A survey. *arXiv preprint arXiv:2402.01801*.
- Zhang, Y.; Li, M.; Long, D.; Zhang, X.; Lin, H.; Yang, B.; Xie, P.; Yang, A.; Liu, D.; Lin, J.; Huang, F.; and Zhou, J. 2025. Qwen3 Embedding: Advancing Text Embedding and Reranking Through Foundation Models. arXiv:2506.05176.
- Zhao, S.; Li, W.; Wang, X.; Foglia, S.; Tan, H.; Zhang, B.; Hamoodi, A.; Nelson, A.; and Gao, Z. 2024. A Systematic Review of Machine Learning Methods for Multimodal EEG Data in Clinical Application. arXiv:2501.08585.
- Zhao, S.; Yi, S.; Zhou, Y.; Pan, J.; Wang, J.; Xia, J.; Li, S.; Dong, S.; and Pan, G. 2025. Wearable Music2Emotion: Assessing Emotions Induced by AI-Generated Music through Portable EEG-fNIRS Fusion. In *Proceedings of the 33rd ACM International Conference on Multimedia*, MM '25, 5627–5636. New York, NY, USA: Association for Computing Machinery. ISBN 9798400720352.
- Zhou, Y.; Zhao, S.; Wang, J.; Jiang, H.; Li, S.; Li, T.; and Pan, G. 2025a. BrainUICL: An unsupervised individual continual learning framework for EEG applications. In *The Thirteenth International Conference on Learning Representations*.
- Zhou, Y.; Zhao, S.; Wang, J.; Jiang, H.; Li, S.; Li, T.; and Pan, G. 2025b. SPICED: A Synaptic Homeostasis-Inspired Framework for Unsupervised Continual EEG Decoding. *arXiv preprint arXiv:2509.17439*.