NeuroChat: A Neuroadaptive AI Chatbot for Customizing Learning Experiences

Dünya Baradari
MIT Media Lab
Cambridge, MA, United States

Nataliya Kosmyna
MIT Media Lab
Cambridge, MA, United States

Oscar Petrov
Brown University
Providence, RI, United States

Rebecah Kaplun
MIT Media Lab
Cambridge, MA, United States

Pattie Maes
MIT Media Lab
Cambridge, MA, United States

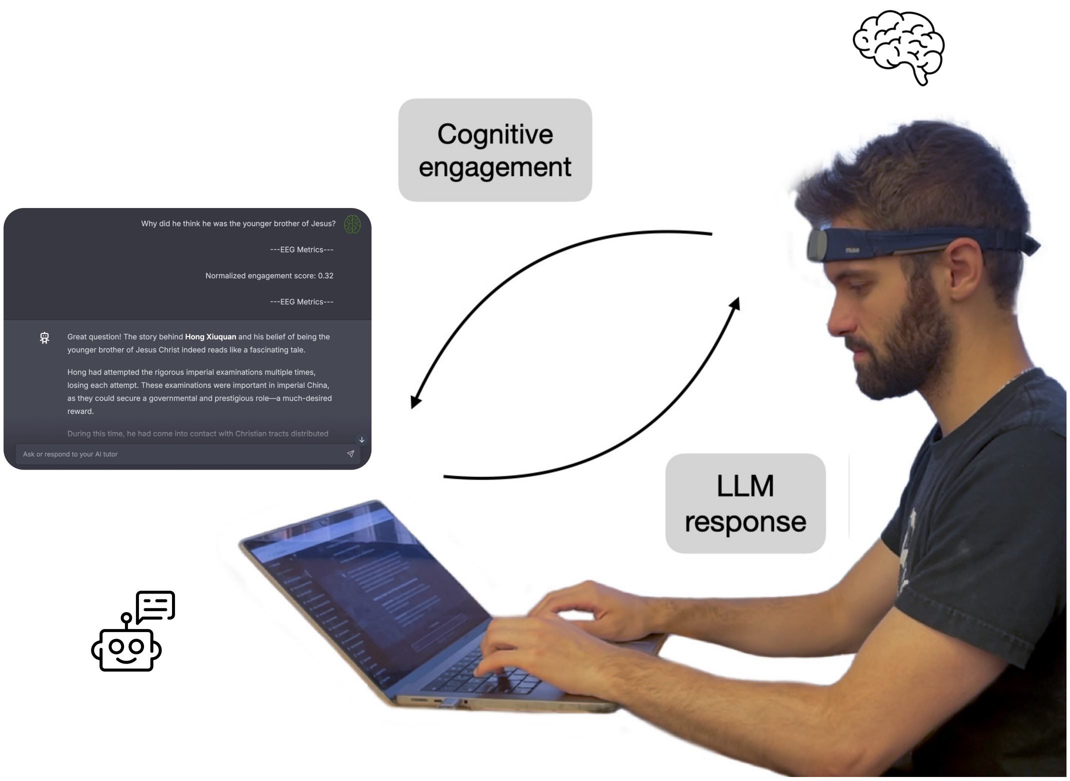

Figure 1: Overview of the NeuroChat neuroadaptive LLM system. A wearable dry-electrode EEG headband collects data from the brain and sends it to the NeuroChat web app, which computes the user's level of engagement. The engagement score is sent with each request to the LLM, allowing it to adapt its response style to the user in real-time.

ABSTRACT

Generative AI is transforming education by enabling personalized, on-demand learning experiences. However, AI tutors lack the ability to assess a learner's cognitive state in real time, limiting their adaptability. Meanwhile, electroencephalography (EEG)-based neuroadaptive systems have successfully enhanced engagement by dynamically adjusting learning content. This paper presents NeuroChat, a proof-of-concept neuroadaptive AI tutor that integrates real-time EEG-based engagement tracking with generative AI. NeuroChat continuously monitors a learner's cognitive engagement and dynamically adjusts content complexity, response style, and pacing using a closed-loop system. We evaluate this approach in a pilot study (n=24), comparing NeuroChat to a standard LLM-based chatbot. Results indicate that NeuroChat enhances cognitive and subjective engagement but does not show an immediate effect on learning outcomes. These findings demonstrate the feasibility of real-time cognitive feedback in LLMs, highlighting new directions for adaptive learning, AI tutoring, and human-AI interaction.

KEYWORDS

Brain-computer interface, electroencephalography (EEG), chatbot, conversational AI, closed-loop

1 INTRODUCTION

The rise of Generative Artificial Intelligence (AI) is reshaping education, with Large Language Models (LLMs) offering new opportunities for personalized, on-demand learning. AI-powered tutors such as ChatGPT Edu [60] and Khanmigo [46] have been integrated into educational settings, promising a future where learners can interact dynamically with AI, receive customized explanations, and engage in self-directed inquiry. AI-generated content has already been shown to improve learning motivation [53] and enhance teaching efficiency [56], as educators leverage these tools to tailor lesson plans, streamline administrative tasks, and develop adaptive instructional materials [81].

At the same time, the introduction of AI-only schools in the United Kingdom and United States [17, 72] demonstrates the growing acceptance of AI-driven adaptive learning platforms. These systems claim to dynamically adjust educational content to individual learners' strengths and weaknesses, promising a level of personalization that traditional classroom settings often struggle to achieve [29]. However, despite this enthusiasm, there remain unresolved challenges regarding the integration of LLMs into education. A key issue is that LLMs lack awareness of a learner's cognitive and attentional state unless explicitly communicated. Without this feedback, LLMs may overwhelm learners with excessive cognitive load, present material at an inappropriate difficulty level, or fail to detect engagement fluctuations, ultimately hindering learning effectiveness.

Additionally, concerns about AI-generated hallucinations, data privacy, and cognitive offloading—where students may become overly reliant on AI for information rather than developing independent research and critical thinking skills—remain pressing issues [45]. While some studies have shown that ChatGPT enhances engagement [35] and improves critical thinking skills [23], a meta-analysis of its effects on education suggests that its primary impact is on academic performance, motivation, and cognitive load, while self-efficacy remains largely unaffected [20]. Moreover, AI-driven education lacks the human ability to detect subtle emotional and cognitive cues, which are crucial in effective teaching and adaptive learning.

Meanwhile, research in neuroadaptive learning systems has demonstrated that real-time physiological feedback, particularly from electroencephalography (EEG), can significantly enhance engagement and cognitive adaptation. EEG-based brain-computer interfaces (BCIs) can measure engagement, attention, and cognitive load, dynamically adjusting learning materials based on a learner's real-time neural state [51, 62]. Several closed-loop EEG systems have been developed to optimize learning by modulating content complexity based on neurophysiological signals. For example, the BRAVO system [54] detects fluctuations in attention and adjusts e-learning materials accordingly, while Thinking Cap [55] integrates EEG-based cognitive load assessment into an Intelligent Tutoring System (ITS), modifying text complexity based on engagement levels. Similarly, FOCUS [39] adapts learning materials for children by integrating EEG-driven interventions during reading sessions. These studies show that adaptive educational environments, when informed by physiological feedback, can improve learning retention and engagement.

However, existing neuroadaptive systems are limited by prescripted content, where adaptations are constrained to pre-defined difficulty levels rather than generating new instructional material dynamically. The Online Continuous Adaptation Mechanism (OCAM) [22] attempted to overcome this limitation by continuously adjusting learning content based on EEG-derived measures of cognitive load, concentration, and emotional arousal. Yet, even in OCAM, content had to be pre-designed and categorized by human experts before it could be adapted. This limitation raises a fundamental question: Can we combine real-time neuroadaptive EEG feedback with the content-generation capabilities of LLMs to create an AI tutor that is responsive to a learner's cognitive state?

A truly adaptive learning system should not only be capable of generating adaptive content but also of recognizing engagement levels and cognitive load—an ability that is fundamental to effective teaching [49]. Pedagogical theories such as Teaching at the Right Level (TaRL) [80] and Cognitive Load Theory (CLT) [19] emphasize that cognitive adaptation is crucial for learning success. When educational tools present material that exceeds a learner's cognitive capacity, learning is significantly hindered [25]. Research has shown that maintaining optimal cognitive load enhances knowledge retention and critical thinking skills [87], yet current AI tutors lack mechanisms to assess and adjust for cognitive strain in real time. To address this challenge, we introduce NeuroChat, a real-time neuroadaptive AI tutor that integrates EEG-based engagement tracking with LLM-driven content generation. NeuroChat continuously monitors EEG-derived engagement levels and dynamically adjusts the depth, complexity, and style of content based on a learner's cognitive state (Figure 1). This work makes the following contributions:

(1) A novel integration of neuroadaptive learning with generative AI, bridging the gap between EEG-based engagement tracking and dynamic content generation.
(2) A closed-loop adaptation mechanism, where real-time EEG data informs and modifies the interaction with an LLM tutor to optimize cognitive load and engagement.
(3) High accessibility and usability, achieved by implementing a lightweight, browser-based, wearable system with dry-electrode EEG headbands, ensuring usability beyond laboratory settings.
(4) Empirical evaluation of NeuroChat's effectiveness, examining how real-time neuroadaptive AI tutoring impacts learning outcomes, cognitive engagement, and user experience compared to non-adaptive AI tutoring.

2 RELATED WORK

2.1 Engagement in Learning

2.1.1 Defining and Conceptualizing Engagement. Engagement is a widely used term in education and psychology, though its definition varies across disciplines [6]. In educational settings, the term can be conceptualized as a 'multidimensional construct encompassing behavioral, emotional, and cognitive dimensions,' [26] which has been shown to be linked to positive learning outcomes, including increasing student motivation [26, 78]. Expanding on this, Reeve and Tseng [67] introduce a fourth dimension, agentic engagement, to describe students' contributions to their learning experience. From the perspective of flow theory, learner engagement can be enhanced by designing learning activities that promote autonomy and provide appropriate challenges to learners' skill level [75]. Sinatra et al. [76] further distinguish between microlevel engagement, which refers to moment-to-moment cognitive focus on a task, and macro-level engagement, which applies to larger social and educational contexts, such as classrooms or institutions. Micro-level engagement can be assessed through physiological techniques such as brain imaging, skin conductivity, or eye tracking, whereas macrolevel engagement is typically measured through sociocultural analysis, observations, or ratings.

In cognitive neuroscience, engagement is closely linked to sustained attention and tonic alertness, reflecting a person's sustained cognitive effort [57]. However, engagement extends beyond attention, incorporating factors such as intrinsic motivation and task involvement [44, 68]. Unlike cognitive load, which reflects the mental demands on working memory, engagement captures both effort and motivation in a task-driven context. For this study, we define engagement as the sustained allocation of cognitive and attentional resources toward a task, influenced by motivation and mental effort.

2.1.2 Physiological Measures of Engagement.

Various physiological technologies have been explored to assess engagement in digital learning environments, including video analysis, eye tracking, and biosensors. Classroom video analysis has been used to monitor student attention, with Raca and Dillenbourg [65] utilizing video recordings and later incorporating a computer vision model to approximate eye gaze [66]. However, these models have limited accuracy in estimating attention levels, as they attempt to infer complex cognitive states from external behavioral approximations that are often ambiguous and context-dependent. Eye-tracking systems provide a more granular measure of attention shifts and mind-wandering, but are often costly, complex, and prone to calibration and accuracy issues [40, 41].

More direct physiological measures include heart-rate variability (HRV) [12], skin conductance (EDA) [11], and electroencephalography (EEG) [51, 64, 87]. Among these, EEG stands out as the only method that directly measures neural activity, providing real-time insights into alertness, attention, and cognitive workload in both controlled and real-world settings [10, 28]. Since learning is fundamentally a neurological process, EEG offers a unique advantage by capturing dynamic brain responses during information processing, making it particularly well-suited for assessing engagement beyond behavioral proxies.

Engagement can be measured using EEG through oscillatory activity (frequency-based markers) and event-related potentials (ERPs). Frequency-based markers provide continuous insights into attention and cognitive workload, with alpha power (8–12 Hz) linked to relaxation and disengagement [30, 31], beta power (13–30 Hz) associated with sustained attention and active problem-solving [64], and theta power (4–8 Hz) indicative of fatigue or reduced vigilance [24]. A widely used composite metric is the Engagement Index, defined as Beta / (Alpha + Theta), where higher values indicate greater attentional focus and cognitive engagement [5, 22, 34, 47, 51, 64]. Alpha asymmetry reflects differences in alpha power between the two brain hemispheres and is often associated with approach motivation and active engagement [24, 83]. Other markers include the Cognitive Load Index (Theta Fz / Alpha Pz) [12, 37] and alpha peak frequency [62], which have both been explored as indicators of cognitive effort and attentional processing efficiency. ERPs, in contrast, offer time-locked neural responses to stimuli, with key components such as P300 (reflecting attentional allocation and task relevance), N200 (linked to conflict detection and executive control), and error-related negativity (ERN) (indicating engagement in performance monitoring) [8, 63, 85].

2.1.3 EEG-Based Engagement in Learning.

EEG-based engagement metrics have been applied in various educational contexts, from providing feedback to presenters [34] to tracking cognitive effort in classroom and workplace environments [30, 33]. EEG has been shown to capture distinct patterns of student attention that differ from self-reports and teacher observations, offering a more objective measure of engagement across instructional activities [30]. Studies have also linked higher engagement to better learning performance. For example, EEG monitoring during video lectures revealed significant fluctuations in attentional focus, suggesting that lecture design should account for these variations [18]. Similarly, in a reasoning task with medical students, engagement correlated with task performance, though the highest engagement was observed in students who struggled, likely reflecting heightened cognitive effort despite insufficient expertise. This aligns with Vygotsky's Zone of Proximal Development, suggesting that while students were highly engaged, the task was beyond their current skill level, leading to cognitive overload rather than effective learning [47].

EEG has also been explored for cognitive workload classification, with implications for future learning technologies. Andreessen et al. [4] trained an EEG-based model to distinguish high and low mental workload, suggesting its potential for adaptive learning systems that adjust reading materials based on cognitive load. Similarly, Apicella et al. [5] demonstrated a low-channel, wearable EEG system for detecting engagement, proposing its use as an input channel for adaptive teaching platforms. While these studies focus on monitoring engagement rather than adapting learning in real time, they lay the groundwork for neuroadaptive systems that dynamically adjust instruction based on cognitive states. The next section explores how such systems leverage EEG engagement data to personalize learning experiences.

2.2 Neuroadaptive Learning Systems

Neuroadaptive systems leverage real-time neurophysiological data, particularly from electroencephalography (EEG), to dynamically adjust instructional content or interaction modalities based on a learner's cognitive and emotional states. These closed-loop systems aim to optimize learning outcomes by continuously monitoring engagement and adapting pedagogical strategies accordingly.

Early approaches to neuroadaptive learning focused on adapting presentation styles based on user engagement. For instance, Pay Attention! [79] employed an embodied storytelling agent that adjusted its voice volume and gestures in real time to recapture students' attention when EEG signals indicated a drop in engagement. This approach significantly enhanced the recall performance of students, demonstrating the potential of adaptive presentation to influence learning outcomes. Similarly, EngageMeter [34] provided real-time feedback to keynote presenters about their audience's engagement levels, enabling dynamic adjustments in delivery style. However, while effective in maintaining attention, these systems were limited to modifying delivery methods without altering the learning content itself.

Thinking Cap [55] extends these ideas into an Intelligent Tutoring System (ITS) featuring an animated tutor agent that dynamically adjusts the complexity of its instructional dialogue with the student based on EEG-derived cognitive load measures. This approach ensures that learners are neither underwhelmed nor overwhelmed. The authors pre-scripted easy and difficult versions of the instructional content by altering text complexity dimensions such as narrativity, syntactic ease, and referential cohesion. The more recent Online Continuous Adaptation Mechanism (OCAM) [22] builds on these principles by continuously monitoring not just engagement but also concentration, cognitive load, and emotional arousal to dynamically adjust content difficulty, pacing, and presentation style. This system has been shown to significantly increase learner concentration and engagement, highlighting the value of multi-dimensional cognitive measures in adaptive learning.

Beyond academic learning environments, Learning Piano with BACh [87] dynamically adapts the difficulty of piano exercises based on cognitive workload (measured via functional near-infrared spectroscopy), guiding learners into their zone of proximal development to optimize skill acquisition. Closed-loop systems have also been found effective for enhancing learning in perceptual-cognitive tasks, as demonstrated by Parsons et al. [62], who improved performance by manipulating a 3D multiple object tracking (3D-MOT) task through real-time neurofeedback.

Other systems focused on providing real-time biofeedback to help users self-regulate their engagement and attention. AttentivU [51], for instance, combines EEG headband with haptic feedback devices that vibrate subtly when engagement levels drop, effectively redirecting attention in both online and in-person learning contexts. Unlike content-adaptive systems, these approaches rely on external cues to prompt re-engagement rather than altering the learning material itself. Similarly, Joie [83] introduces a joy-based brain-computer interface (BCI) that uses prefrontal alpha asymmetry—an EEG marker linked to positive emotional states—to control an endless runner game. By training users to consciously modulate their brain activity through strategies like imagining joyful scenarios, Joie highlights the potential of neuroadaptive systems to foster affective engagement alongside cognitive performance.

Across these systems, a shared limitation is that all content-driven systems rely on pre-scripted content that needs to be prepared by the researchers to allow for the adaptation. NeuroChat overcomes this barrier by integrating generative AI, which can create new content adapted in complexity and presentation style to the reader's cognitive state and specific questions on the fly.

2.3 Generative AI-BCI Systems

The integration of generative AI with brain-computer interfaces (BCIs) is an emerging research area. While machine learning has long been used to analyze EEG data, most AI-enhanced BCI systems have focused on brain state classification rather than interactive, real-time adaptive applications. The introduction of generative AI expands the possibilities of BCIs beyond passive decoding, enabling dynamic content modulation and interactive adaptation. Early investigations propose that integrating LLMs with BCIs could significantly enhance human-computer interaction, benefiting both individuals with neurological conditions and healthy users [13].

2.3.1 Using Generative AI to Analyze EEG. A major focus in AI-BCI research has been EEG-based brain decoding, where generative AI and machine learning models are used to encode and decode the neural signals underlying visual or auditory information processing [7, 32, 84]. While these methods advance neural signal processing, they remain limited in real-time user interaction. Readers interested in these approaches can refer to a comprehensive review by Sabharwal and Rama (2024) [70]. Beyond decoding, LLMs have been increasingly applied to EEG for brain state classification and assistive communication [88]. In clinical applications, language model-enhanced BCI communication systems have significantly improved typing accuracy for ALS patients by up to 84% in online BCI spelling sessions [77]. Subsequent approaches have demonstrated that LLMs can classify brain states at the word level from EEG data during reading tasks [36, 89].

Recent research has extended these applications to foundation models that generalize across EEG tasks. NeuroLM [43] and Neuro-GPT [15] function as foundation models, pre-trained on large EEG datasets using self-supervised learning and task-based fine-tuning to develop multi-purpose EEG processing models. NeuroLM, trained on over 25,000 hours of EEG recordings, aligns brain signals with text-based representations, enabling multi-task analysis in areas like cognitive workload detection, emotion recognition, and sleep staging. Neuro-GPT, trained on the TUH EEG corpus, applies GPT-style tokenization to EEG data, improving feature extraction and adaptability to small datasets. In contrast, EEG-GPT [48] and Lee & Chung (2024) [52] focus on task-specific applications—EEG-GPT applies few-shot learning for EEG-based brain state classification, while Lee & Chung fine-tune GPT-3.5 Turbo for intracranial EEG (iEEG) interpretation, mapping neural signals to cognitive states.

Other approaches have explored personal health and well-being. For instance, Sano et al. (2024) [71] used LLMs to interpret EEG signals for sleep quality assessment, providing tailored recommendations. Similarly, EEG Emotion Copilot [14] integrates EEG with a lightweight (0.5B parameter) LLM to analyze EEG signals, identify emotional states, and generate automated clinical insights. [38] propose MultiEEG-GPT, a model that integrates EEG with multimodal data—such as facial expressions and audio—to enhance mental health assessments using LLM-based classification.

Additionally, generative AI techniques have been leveraged for data augmentation to enhance EEG-based model training [21, 90]. However, while these approaches highlight generative AI's ability to process EEG data for individual adaptation, they focus on recognizing states and have yet to support real-time user interaction.

2.3.2 Artistic Applications Using EEG to Modulate Generative AI Outputs. While most research has focused on analyzing EEG data, a growing field explores EEG as a control mechanism for real-time generative AI adaptation. Early explorations have emerged in artistic and creative applications, where EEG signals influence AI-generated media production. For example, Imagination Engine [1] translates EEG activity into abstract visual art, while Real-Time Neuro-Augmented Cinema [9] enables cinematic modifications

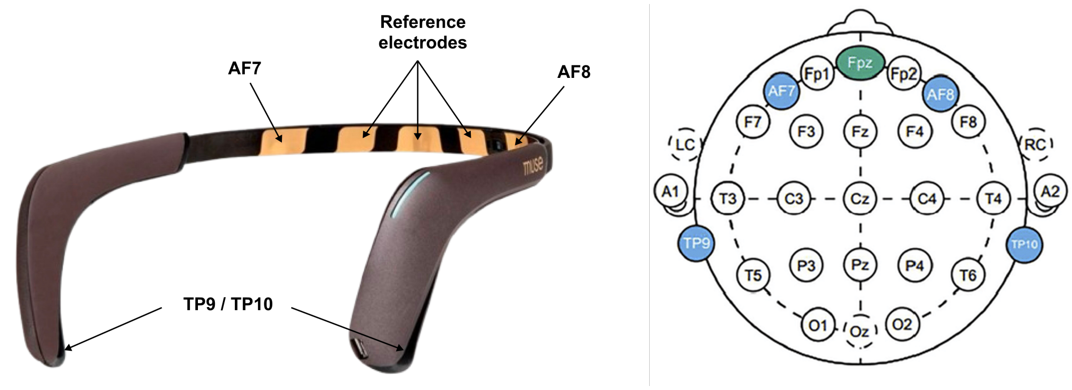

Figure 2: Left. The Muse 2 EEG system made by InteraXon Inc. Right. Electrode locations of Muse 2 headband according to 10-20 System. CC Teixeria, Gomes, and Brito-Costa (2023).

via neurofeedback. Similarly, the Bio-Mechanical Poet [82] maps real-time EEG signals to symbolic representations, creating immersive poetic audiovisual experiences. These projects demonstrate EEG's potential to actively modulate generative outputs rather than merely classifying brain states. However, these applications remain limited to artistic expression, with little research on EEG-driven adaptation of linguistic content. The potential to use real-time neurofeedback to shape AI-generated textual interactions—particularly in education—remains largely unexplored.

2.3.3 Neuroadaptive Generative AI Systems. The most advanced neuroadaptive system integrating generative AI for real-time adaptation is AdaptiveCoPilot [86], designed for expert pilots in virtual reality. AdaptiveCoPilot continuously adjusts visual, auditory, and textual cues based on real-time cognitive load assessments, optimizing performance in high-stakes environments. However, it relies on functional near-infrared spectroscopy (fNIRS) rather than EEG and is tailored for high-performance cognitive tasks rather than learning applications.

Despite rapid advancements in AI-enhanced BCI research, no existing system has leveraged EEG data to dynamically modulate LLM-driven chatbot interactions. This gap underscores the novelty of NeuroChat, one of the first systems to integrate EEG-based cognitive state tracking with generative AI in real-time.

3 SYSTEM DESIGN

We set out four core design goals to ensure that NeuroChat is an accessible, responsive, and effective neuroadaptive learning system:

(1) **D1: Wearable, Non-Invasive Brain Sensing**: NeuroChat employs a consumer-grade, non-invasive EEG headband to measure engagement in real time. We opted for the 4-channel Muse EEG headband by InteraXon [60], balancing signal reliability with ease of use. This design ensures that users can engage with the system without complex electrode setups or invasive procedures.

(2) **D2: Real-Time Adaptive Personalization**: To maximize learning effectiveness, NeuroChat provides continuous neurofeedback, dynamically tailoring chatbot responses based on real-time engagement levels. The system processes EEG data in real time to adjust content complexity, response style, and pacing, ensuring an adaptive and personalized learning experience.

(3) **D3: Seamless EEG Integration in Natural User Environments**: Given that most users interact with NeuroChat on a laptop or desktop computer, the system accommodates a stationary, movement-minimizing environment, which is ideal for EEG signal acquisition. This design choice minimizes motion artifacts, resulting in more reliable neurofeedback processing.

(4) **D4: Web-Based Accessibility and Low-Cost Implementation**: NeuroChat is designed to be fully browser-based, eliminating the need for complex server-side infrastructure and enabling plug-and-play usability. Users can access the system on any platform with minimal setup, making it scalable and accessible for a broad audience.

3.1 Interaction Flow

NeuroChat integrates real-time EEG data with an LLM chatbot via a web interface to create adaptive, personalized responses. The user flow is as follows (Figure 3):

(1) **Connection:** The user fits on the Muse 2 EEG headband and connects it to the NeuroChat web app via Web Bluetooth for real-time data streaming.

(2) **Calibration:** The user completes a 2-minute relaxation task to determine the engagement minimum (E_min) and a 2-minute word association task for the engagement maximum (E_max). These values are stored in the browser's session storage for normalization.

(3) **Interaction:** During interaction, the system continuously computes the normalized engagement score using a 15-second sliding window. The last score before the user begins typing is captured and embedded in the query to the chatbot, hidden from the user, ensuring that typing doesn't interfere with the engagement metric.

(4) **Interactive Response:** The query, along with the embedded engagement score, is sent to the LLM provider, which returns a response tailored to the user's cognitive state.

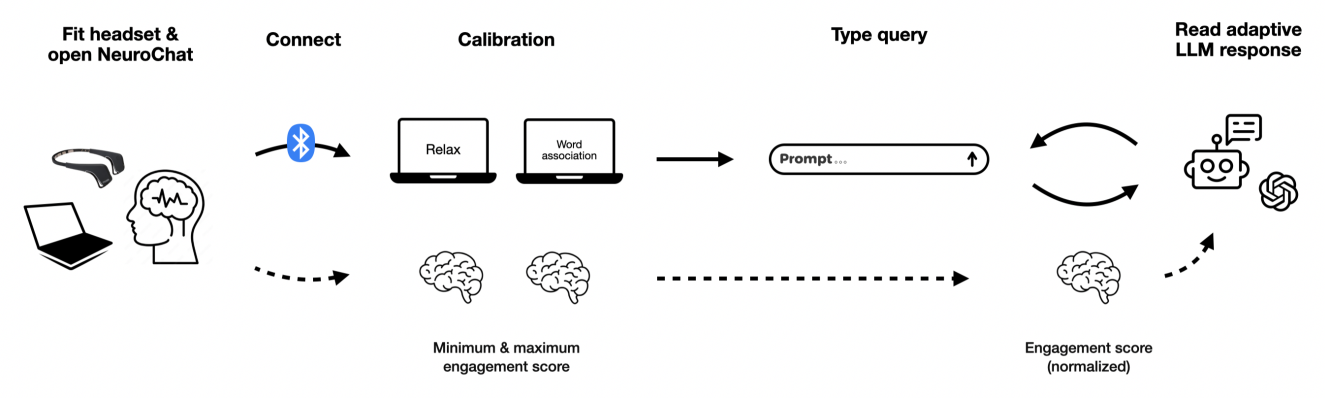

Figure 3: Overview of the NeuroChat system and user flow. The user connects the Muse headband, undergoes calibration, and interacts with the neurofeedback-driven LLM. Engagement scores are computed and inserted into the prompt unnoticed by the user.

3.2 EEG Signal Processing

3.2.1 Device. Our system uses the Muse 2 EEG headband, building on prior research that has leveraged consumer-grade devices with 1 to 6 channels to assess cognitive engagement in learning contexts (e.g., [34, 51, 79, 83]). The Muse 2 samples at 256 Hz and includes electrodes at Fpz, AF7, AF8, TP9, and TP10, following the 10-20 System (Figure 2) [42]. The Fpz electrode serves as the reference. EEG data is streamed to a web browser using the open-source MuseJS library [74], which enables real-time streaming via Web Bluetooth.

3.2.2 Preprocessing. The EEG data processing pipeline follows established methods from Hassib et al. [34], Kosmyna and Maes [51], Szafir and Mutlu [79] and others. A bandpass filter (1–30 Hz) is applied to retain relevant neural activity while minimizing noise, and a 60 Hz notch filter removes power line interference. The data is then segmented into 1-second epochs with 250 ms intervals to enable continuous analysis with sufficient temporal resolution. Power spectral density is computed via fast Fourier transform (FFT), and band power is extracted for each frequency range to derive meaningful neural features.

3.2.3 Engagement Score. The engagement index (or engagement score) serves as the core metric of our system, enabling real-time quantification of cognitive engagement during mentally demanding tasks. First introduced by Pope et al. [64], this metric is computed as a ratio of key EEG frequency bands using the formula:

$$E = \frac{\beta}{\alpha + \theta} \tag{1}$$

where β (11–20 Hz), α (7–11 Hz), and θ (4–7 Hz) correspond to EEG-derived neural oscillations. The index is based on the principle that higher beta power reflects heightened brain activity during cognitive tasks [11]. The beta frequency band is particularly associated with cognitive processes such as visual attention, motor planning, and active information processing, all of which indicate an engaged mental state. Conversely, increased alpha and theta activity typically correspond to lower cognitive engagement, with alpha waves linked to relaxation or passive states of rest [24, 30, 31].

The engagement index has been widely validated across various applications, including cognitive load assessments [27], visual processing studies, and sustained attention tasks [10]. It has also been applied in complex task environments such as the multi-attribute task battery (MATB) [64], which involves tasks like tracking, resource management, and communication. These studies demonstrate the engagement index's effectiveness in detecting attention shifts and fluctuations in cognitive state triggered by external stimuli [3, 16].

Following our preprocessing pipeline, we extract frequency bands for each epoch and average them over a 15-second sliding window, as established by Szafir and Mutlu [79]. Averaging over a time window allows us to assess a user's engagement over a meaningful duration while they read and process the LLM's output, rather than capturing momentary fluctuations. We selected a 15-second window to account for variations in reading speed, ensuring sufficient time for users to engage with the response. Unlike previous studies, we opted against exponentially weighted moving averages, as our focus is on sustained engagement throughout a task rather than transient cognitive spikes.

Finally, we normalize the engagement score following Kosmyna and Maes [51]. Normalization requires determining a minimum and maximum engagement score for each user, which we obtain from the calibration task conducted before the main experiment. During calibration, users engaged in two tasks, each lasting two minutes:

1. Relaxation: Participants remain still, minimizing cognitive effort while we record baseline EEG data.
2. Mental word association: Participants perform a cognitive task that requires generating words based on the final letter of the previous word (e.g., "elephant" → "tiger"). This method has been shown to effectively induce cognitive activation in non-ALS participants [50].

The lowest and highest engagement scores from the two tasks, respectively, are taken as the normalization minimum $E_{\min}$ and

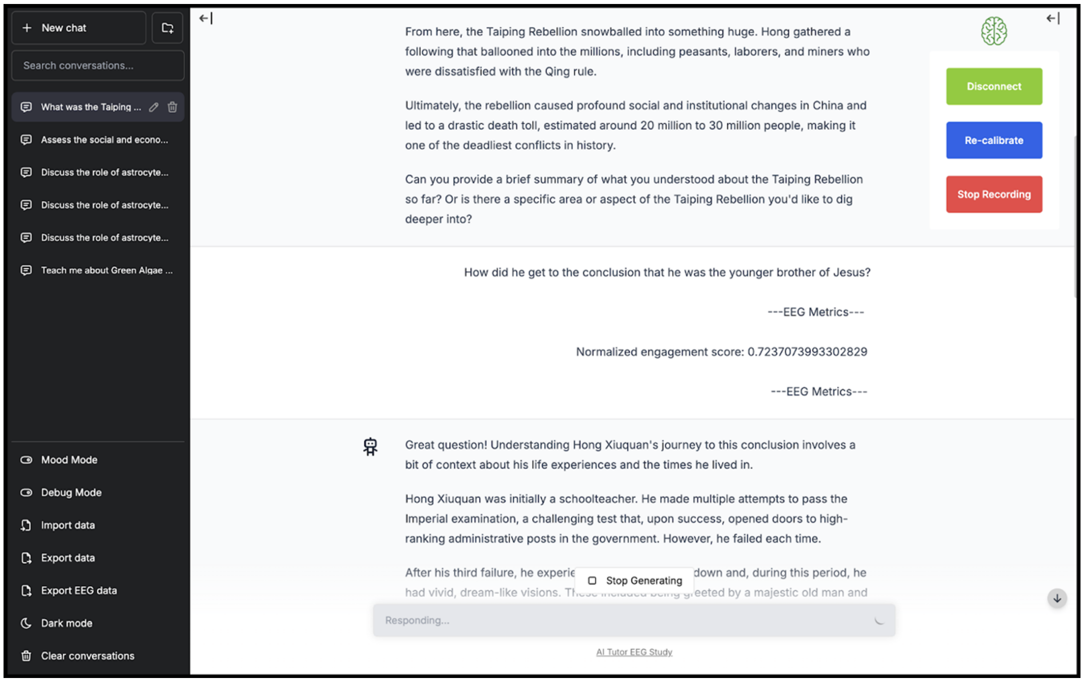

Figure 4: NeuroChat user interface with exposed EEG metrics in the user prompt and experimenter control menu. The connection to the Muse EEG device happens through the Brain Widget in the top right corner. "Mood mode" activates the EEG metric injection into the user's prompts, and turning off "Debug mode" allows the experimenter to hide these from the user. Chats, raw and filtered EEG data, and computed EEG metrics from the Muse device are stored in the browser's native IndexedDB and can be exported from the Settings panel.

maximum $E_{\max}$. The calibration task differs from the main task in that it uses a 10-second sliding window, balancing the 5-second interval used in prior studies [34, 51, 79] and the 15-second window applied in our main experiment. The normalized engagement score E_{norm} is then calculated using:

$$E_{\text{norm}} = \frac{E - E_{\min}}{E_{\max} - E_{\min}} \tag{2}$$

where E represents the engagement score averaged over the past 15 seconds.

3.3 LLM Adaptation

The mechanism through which NeuroChat responds adaptively to the user's cognitive state is by embedding their engagement score into each query submitted to a Large Language Model. We used OpenAI's GPT-4-turbo model, the latest at the point of study.

The system prompt provides a guideline as to how the LLM should adapt its response style to the user. We developed it through careful evaluation of pilot testing and insights from OpenAI's Teaching with AI guide [59] (see Appendix A.1). One key insight was that prompting GPT-4 to increase the user's "engagement" often resulted in overly casual, upbeat responses, as the model interpreted the term informally. Reframing the engagement score as a "cognitive load metric" helped maintain a neutral tone while allowing responses to adjust dynamically based on neurofeedback. Based on the engagement index, the LLM was instructed to modulate detail, scientific depth, response format (e.g., bullet points versus long-form text), and use of Socratic questioning. Additionally, framing its role as a "good tutor" improved response quality.

3.4 User Interface

The NeuroChat user interface (UI) consists of four key components (Figure 4):

(1) **Brain Connect Widget** (Top Right Corner): Allows users to connect or disconnect the EEG device, calibrate or re-calibrate the system, and start or stop EEG recording to compute the engagement index.
(2) **Calibration Modal** (Full Screen) – Appears only after the Muse headset is connected and provides instructions for a 2-minute relaxation phase followed by a 2-minute mental word association task. If the EEG connection is lost, users can restart or resume from the completed relaxation phase.
(3) **Chatbot Interface** (Main Screen Area) – Functions similarly to ChatGPT, displaying an alternating conversation between the user and the AI tutor.
(4) **Menu Sidebar** (Left-Hand Side) – Contains chat history where users can manage past LLM conversations by creating folders, renaming chat titles, and deleting individual chats. Also features the Settings panel, which provides options to toggle "Mood Mode" (enabling LLM adaptation), activate "Debug Mode" (hiding EEG metrics from the UI), import/export chat history, download EEG data (from the

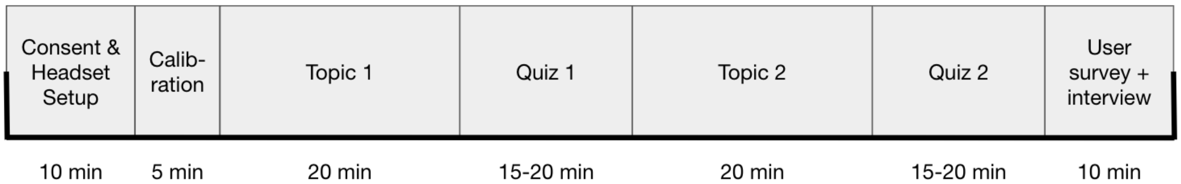

Figure 5: Overview of study procedure (not to scale).

browser's IndexedDB), switch between dark and light mode, and reset the chat history for a new user session. The terms "Mood Mode" and "Debug Mode" were intentionally chosen to provide visual cues to the experimenters while being vague enough to the participants.

4 METHODOLOGY

4.1 Hypotheses

Based on prior research, we formulate the following hypotheses:

- (H1) Objective Engagement: NeuroChat will elicit higher engagement levels than interaction with a standard GPT model, as measured by EEG-derived engagement scores.
- (H2) Subjective Engagement: Participants will report greater subjective engagement and satisfaction with NeuroChat, perceiving it as more engaging and effective than a traditional AI tutoring model.
- (H3) Learning Outcomes: Participants using NeuroChat will achieve higher scores on post-interaction learning assessments compared to those using the standard GPT model.

4.2 Participants

Thirty participants (15 female, 13 male, 2 other), predominantly from academic backgrounds, were recruited for this study (M = 32.4 years, median = 30) and compensated with a $50 Amazon gift card. The study received approval from MIT's institute's ethical review board (protocol no. 21070000428).

4.3 Study Design and Protocol

We adopted a within-subject study design after pilot studies revealed significant individual differences in interactions with the AI chatbot. This design allowed each participant to serve as their own control, minimizing variability and enabling direct performance comparisons between the NeuroChat experimental condition and the control condition. The control condition consisted of a regular GPT chatbot, which was prompted to act within an AI tutoring task via its system prompt for fair comparison (see Appendix A.2).

As study topics, we selected the *Tyrannosaurus rex (T. rex)* and the *Taiping Rebellion*. These topics were chosen to minimize prior topic bias while allowing room for facts and explorative interpretation. Although the T. rex is widely recognized, most people lack in-depth knowledge about the dinosaur. Similarly, despite its historical significance, the Taiping Rebellion is rarely emphasized in Western education. Both topics provided sufficient complexity and depth for meaningful engagement within the 20-minute learning session while still remaining accessible for participants. Condition and study topic order were counterbalanced using a Latin square design.

Before the session, participants signed a consent form and turned off their electronic devices. They were fitted with a Muse EEG headband, and signal quality was verified via the Muse EEG app [112]. Participants were instructed to minimize movement to reduce motion artifacts.

The study lasted about 2 hours and proceeded as follows (Figure 5):

(1) *Pre-Session Measures*: Participants completed a background questionnaire assessing their alertness and previous experience with AI chatbots. A brief EEG calibration phase followed (2 minutes relaxation, 2 minutes mental exercise).
(2) *AI Chatbot Interaction*: Participants engaged with the chatbot for 20 minutes on their first assigned topic, with the goal of "learning as much as possible." To guide exploration, they received starting pointers—e.g., characteristics, behavior, and archaeological research for T. rex and historical context, significance, and consequences for the Taiping Rebellion. Participants were free to focus on aspects they found interesting.
(3) *Knowledge Assessment*: Immediately after the chatbot interaction, participants completed a quiz consisting of fill-in-the-blank and multiple-choice (MCQ) questions, followed by a 15-minute essay to assess understanding. To prevent preparatory bias, participants were not informed about the quiz beforehand. The same quiz was used across conditions.
(4) *Break & Condition Switch*: Participants took a short break before repeating the process with the second topic and condition. EEG data and chatbot interaction logs were continuously recorded.
(5) *Final Survey & Interview*: Participants completed a post-study user survey and a semi-structured interview focusing on their subjective engagement and experience across conditions. Interviews were thematically analyzed.

4.4 Evaluation

Assessing learning outcomes requires a multifaceted approach. The quizzes incorporated recall-based and synthesis-based questions to capture different cognitive processes. Recall was assessed through multiple-choice (MCQ) and fill-in-the-blank questions, while creative synthesis was evaluated via mini-essays requiring critical thinking and analysis. MCQ and fill-in-the-blank questions were designed to assess factual recall, covering information likely encountered during topic exploration. Question complexity varied to

Table 1: Model summary and model fit statistics.

(a) Model summary for Engagement (normalized) by Condition, accounting for Order.

Predictor	β	SE	z-value	p-value	95% CI
Intercept	-0.382	0.171	-2.230	0.026	[-0.718, 0.046]
Condition (E)	0.216	0.099	2.185	0.029	[0.022, 0.410]
Order	0.181	0.099	1.828	0.068	[-0.013, 0.375]

(b) Model fit statistics.

Statistic	Value
Log-Likelihood	-19.621
AIC	47.242
BIC	54.727

reflect a range of difficulty levels appropriate for participants with minimal prior knowledge engaging for under 20 minutes. For the essay task, participants had 15 minutes to write and could choose from a set of prompts or create their own.

One author manually scored the quiz blind. Fill-in-the-blank questions were graded with 2 points each, with partial scores for semi-correct answers, and multiple choice questions (MCQ) were given 1 point per correct option. Since we only wanted to grade responses that had come up in the chat interaction, an automated keyword detection script scanned a participant's message history for the presence or absence of a question, which was checked manually. Answers not covered were excluded from the participant's total score, leaving us with proportional participant scores for comparison.

The 15-minute mini-essays were graded blind by a professional high school English teacher on a 5-point scale across 4 categories: Content, Structure and Organization, Language and Style, and Accuracy (Spelling and Grammar). The Content category was given double weighting when calculating the final score.

In addition to objective assessments, participants completed a post-study user survey and were interviewed by 1 of 3 of the authors in a semi-structured interview lasting no more than 5 minutes. The interviewers took note of key quotes and sentiments and subsequently cross-read each other's notes and discussed additional takeaways. Survey responses and interview notes were compiled and thematically analyzed by the first author over multiple rounds of descriptive coding.

5 RESULTS

To evaluate the effects of NeuroChat on engagement and learning outcomes, we conducted analyses on EEG engagement scores, learning assessments, and user feedback. Our results address three key areas: (1) cognitive engagement (EEG-derived engagement index), (2) user-reported engagement, and (3) learning performance (quiz and essay scores).

5.1 Cognitive Engagement

Since the system processes EEG data in real time, engagement scores were computed dynamically during each session. Six participants were excluded due to missing or poor-quality EEG signals, and one additional participant was removed due to non-compliance with instructions. This resulted in 24 participants for analysis. For preprocessing, we removed missing values and extreme outliers (beyond 3× standard deviation) and manually inspected the engagement data, excluding segments with signal disconnections. Despite participant-level calibration, engagement scores exhibited relatively high between-subject variability (intra-class correlation coefficient (ICC) = 36.5%, $p < 0.05$). Therefore, to account for individual baseline differences while preserving within-subject variability, we applied z-score normalization, adjusting each participant's engagement scores based on their mean and standard deviation across both conditions. This allowed for direct comparison of relative engagement differences between NeuroChat and the control condition.

Given the repeated-measures design, where each participant contributed data across multiple conditions, and occasionally missing data, we fit a Linear Mixed Model (LMM) to examine the effect of the experimental condition (Condition) on normalized engagement (E_{norm}) (Table 1). For such conditions, an LMM is more appropriate than a paired t-test, as it accounts for within-subject correlations and random variability across participants. We used the statsmodels package in Python [73].

An initial model with Condition as the sole fixed effect did not yield a statistically significant relationship with engagement (β = 0.186, p = 0.063). Since task order could introduce confounding effects, we extended the model by adding Order as an additional fixed effect. This refined model revealed a significant effect of Condition (β = 0.186 = 0.216, p = 0.029) and a marginal effect of Order (β = 0.186 = 0.181, p = 0.068) on normalized engagement (Figure 6). This finding suggests that task order may have influenced engagement levels, warranting its inclusion in the final model. As expected, the random effect variance for participants was low (σ^2 = 0.001), reflecting the impact of z-score normalization, which minimized inter-individual differences before running the model, leaving only within-condition variation.

Extending the model to test for additional effects of study topic, age, education level, chatbot experience, chatbot familiarity, chatbot usage frequency, and prompt engineering skill revealed no significant influence of these factors ($p > 0.1$). In conclusion, when accounting for order effects during the experiment, participants in the NeuroChat condition were, on average, relatively more cognitively engaged.

5.2 Learning Test Performance

To assess whether NeuroChat improves learning outcomes (H3), we compared participants' quiz and essay performance. For the quiz, mean proportional scores showed no significant difference between conditions (E = 61.02%, C = 60.66%). Likewise, for the essay, participants in the experimental condition averaged 18.27 points, compared to 17.69 points in the control condition, indicating no notable difference between groups.

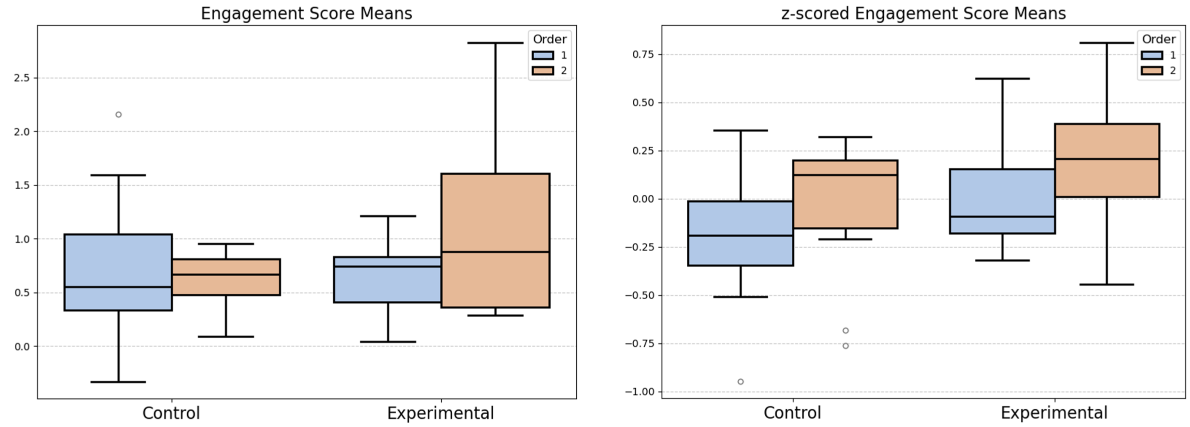

Figure 6: Distribution of engagement score means in the control and experimental conditions by order (right: z-score normalized).

5.3 Perceived Engagemenet & Subjective Evaluations

We analyzed user responses based on the post-questionnaires and informal interviews regarding reported levels of engagement and learning preferences. Participants were asked in writing and verbally about their perceived engagement, noticeable differences, and learning preferences between the chats. We categorized their feedback into five main themes: personalized feedback, and response style (factual vs. conversational), density of information, follow-up questions, and additional feedback, each contributing to the perceived engagement and satisfaction.

5.3.1 Personalized Feedback and Human-Like Responses. A prominent theme in the subjective evaluations was NeuroChat's more human-like responses, which many participants found engaging. P1 noted that the chatbot "mimics a real person" and provided feedback that made it seem more interactive and lifelike, such as saying "Great question" after user input. P31 also expressed satisfaction with the experimental chatbot, stating, "Oh, I loved the second one! I really liked how it was saying how I was feeling." This participant emphasized the importance of NeuroChat's ability to provide feedback tailored to their emotional state, suggesting that it was responding to affect and engagement levels in real time. P19 added that NeuroChat had "more personality" compared to the standard GPT, making the interaction feel more tutor-like and conversational, rather than merely factual.

Furthermore, NeuroChat's personalized prompts made the experience feel more responsive and adaptive for some participants. P30 described the experimental chatbot as "always responding to my prompt," and noted how it felt more dynamic than the control chatbot, which often came across as rigid and formal. Similarly, P28 enjoyed the depth of engagement, stating that NeuroChat allowed for "more meaningful topics to ask," creating a richer, more exploratory interaction.

However, some participants preferred the more straightforward approach of the control chatbot. P7 and P10, who identified themselves as "scientific minds," favored the control condition for its more factual and concise responses. They felt that NeuroChat's conversational style detracted from the focus on information, making it harder to digest the content. Similarly, P26 found the control's more nuanced, fact-driven responses preferable, describing NeuroChat's conversational prompts as distracting rather than helpful. Despite this, many participants noted that the control chatbot often lacked the personal touch, with P31 describing the control chatbot as feeling like "a regular chatbot" that lacked awareness of the user's emotions or engagement level.

5.3.2 Factual vs. Casual Response Style. The response style between NeuroChat and the control chatbot was another point of divergence in subjective feedback. Participants like P28 and P30 enjoyed the more conversational and engaging tone of NeuroChat. P28 described the experimental chatbot as "very fun, like a tour guide," with a more interactive and fluid exchange, while the control felt "like a textbook" in comparison. These participants appreciated NeuroChat's ability to dive deeper into topics, making the learning experience more dynamic and enjoyable. In contrast, some participants preferred the control chatbot's more formal, factual style. P7 found that the control condition allowed for more focused learning, noting that NeuroChat was more prone to casual conversation that made it harder to focus on the core information. P17 reflected that while the experimental chatbot was more enjoyable due to its fluidity and tendency to present fun facts, the control chatbot's responses were better structured and felt more educational, comparing the control to a "blog post" with dense information. However, some participants also criticized the control chatbot for being too rigid and not encouraging exploration. P32 mentioned that while the control condition was more "analytical," it felt more like attending "a serious lecture," with little room for the more enjoyable, exploratory exchanges that NeuroChat provided. P19 similarly remarked that the control chatbot seemed "less eager to engage," contributing to a less immersive and personalized learning experience.

5.3.3 Density of Information. The verbosity of NeuroChat's responses was a double-edged sword for participants. Some, like P14 and P12, appreciated the deeper exploration of topics provided by

NeuroChat, which allowed them to learn more than the control chatbot. P14 commented that while they would choose the control chatbot for quick learning, NeuroChat was more suitable for in-depth exploration due to its comprehensive answers. Similarly, P28 noted that the experimental chatbot fostered a more nuanced understanding, allowing them to focus on the context and reasons behind a topic rather than just absorbing arbitrary facts. However, other participants found the sheer volume of information overwhelming. P27 described NeuroChat's responses as "paragraph after paragraph," and P32 admitted to skipping parts of its verbose answers, opting instead for the control's more concise and digestible responses. P19 and P33 both noted that NeuroChat had a tendency to provide redundant information, which diminished the clarity and relevance of the responses over time. While the control chatbot was favored for its brevity, some participants like P30 found that it occasionally oversimplified complex topics, limiting deeper understanding.

5.3.4 Follow-Up Questions. Feedback on the follow-up questions varied widely between the NeuroChat and the control condition. Participants appreciated the specific nature of the questions in NeuroChat. P2 stated, "I liked the questions at the end, they were really specific—much better than Copilot." This sentiment was echoed by P23, who found the prompts helpful, especially since they "didn't know anything about the topic," and appreciated the guidance.

However, not all participants found the follow-up questions useful. P18 felt that while they were prompted with questions, the responses didn't lead anywhere meaningful, causing frustration. P26 provided a particularly nuanced critique, noting that the experimental chatbot's prompts felt superficial and it didn't seem to "care" the way a human would. They felt that the questions prompted by NeuroChat often missed their actual interests, leading to a sense of disconnection from the conversation. P33 also noted that NeuroChat's tendency to drive the conversation in a specific direction was problematic, as they had to repeatedly bring it back to their original question, which disrupted the flow of engagement.

In contrast, participants found the control chatbot's lack of follow-up questions to be limiting in terms of engagement. P19 noted that while the control was concise, it didn't prompt any follow-up questions, making the conversation feel more transactional and less interactive. This lack of conversational depth in the control condition was also mentioned by P23, who described it as having "more general, broad questions," which felt less engaging than the more creative, tailored prompts offered by NeuroChat.

5.3.5 Overall Engagement and Satisfaction. In summary, subjective feedback on NeuroChat's engagement and satisfaction levels varied, with the conversational and human-like elements appealing to participants seeking a more interactive, engaging experience. However, those who preferred straightforward, fact-focused learning found the control chatbot more aligned with their needs. NeuroChat's neurofeedback-driven prompts were effective for some, but others found them intrusive or misaligned with their interests, which could detract from overall satisfaction. Overall, more participants provided positive feedback on their engagement in their experimental condition (23 participants) compared to the control condition (10 participants). This gives us supporting evidence that participants using NeuroChat report higher levels of engagement and satisfaction than those using a chatbot without neurofeedback (H2).

6 DISCUSSION

This study examined whether NeuroChat, a neuroadaptive AI chatbot, enhances engagement and learning outcomes by adapting its responses based on real-time EEG feedback. Our results confirm that NeuroChat successfully increased cognitive (EEG-measured) and self-reported engagement, demonstrating its feasibility as a neuroadaptive tutoring system. However, no significant differences were found in learning performance, indicating challenges in translating engagement into measurable knowledge gains. Below, we discuss these findings in the broader context of adaptive learning, generative AI, and brain-computer interfaces (BCIs) before exploring key challenges and future directions.

6.1 Key Findings: Engagement Gains, Learning Outcomes, and User Perception

We found that NeuroChat significantly increased EEG-measured engagement ($p = 0.029$), suggesting that real-time neuroadaptive feedback can enhance sustained attention. These findings align with prior neuroadaptive systems like Pay Attention! [79] and EngageMeter [34], which successfully modulated engagement through adaptive presentation techniques. However, NeuroChat goes further by modifying conversational flow, response complexity, and interaction depth, positioning it as an interactive, closed-loop system rather than a passive monitoring tool.

User feedback also reflected higher perceived engagement with NeuroChat, with many participants finding its responses more human-like, responsive, and personalized. However, individual differences emerged—some users preferred a factual, concise style, while others favored conversational, exploratory interactions. This suggests that adaptive tutoring must account for personalized learning preferences to be effective.

Despite increased engagement, NeuroChat did not significantly improve quiz or essay scores. This result mirrors previous studies on learning with LLMs [53], which found that while personalization enhances motivation, it does not always yield better performance on traditional assessments. Possible explanations include:

(1) Engagement ≠ Effective Learning – EEG-based engagement captures sustained attention, but not necessarily deep learning or knowledge retention.
(2) Task Design Limitations – Unlike structured adaptive systems like BACh [87] (which progressively increased piano sheet music difficulty), NeuroChat allowed open-ended learning, making structured difficulty progression harder to implement.
(3) Short Study Duration – Neuroadaptive learning benefits may emerge over multiple sessions, but our study measured learning in a single interaction.

Future work should explore long-term retention, conceptual understanding, and scaffolding techniques that could better translate engagement gains into measurable learning improvements.

6.2 Implications for Neuroadaptive Learning & AI-Powered Tutoring

Traditional neuroadaptive learning systems relied on pre-scripted content, where researchers manually assigned learning materials to high- or low-engagement conditions. NeuroChat overcomes this limitation by leveraging generative AI to create content dynamically, enabling real-time adaptation tailored to individual users. This is a fundamental shift in adaptive learning—moving from rule-based, pre-mapped content to generative, personalized tutoring.

Most LLMs require users to explicitly communicate their needs (e.g., "Explain this differently", "Make it simpler"). NeuroChat removes this barrier by inferring user engagement levels directly from EEG data, reducing the need for manual prompt engineering. This has implications for personalized AI assistants, where cognitive state tracking could enhance adaptability without user effort.

NeuroChat has particular relevance for self-directed learners, who often struggle with maintaining engagement. In 2021, over 220 million students enrolled in MOOCs, yet the average completion rate remains at 13% [2, 58]. A neuroadaptive AI tutor could help sustain motivation and prevent disengagement, particularly in open-ended, autonomous learning environments.

Beyond education, NeuroChat's EEG-driven AI system could support knowledge workers, particularly those who struggle with focus and information retention. Additionally, LLM-based BCIs have been proposed for aiding individuals with learning challenges, including ADHD [13, 39, 62]. These applications highlight the potential of neuroadaptive AI beyond the classroom.

6.3 Challenges & Limitations of NeuroChat

Participants showed high variability in engagement and preference for different interaction styles. Some learners thrived in guided, exploratory conversations, while others preferred concise, fact-driven responses. Future systems should incorporate user preference settings, such as:

- Preferred interaction style (e.g., structured vs. exploratory).
- Response format (e.g., bullet points vs. narratives).
- Memory-based personalization (as seen in OpenAIś user memory feature [61]).

Moreover, high engagement is not always beneficial. Medical reasoning studies [22] found that struggling learners showed the highest engagement, suggesting that increased engagement can sometimes signal cognitive overload rather than productive learning. Future neuroadaptive tutors must ensure users remain in their zone of proximal development rather than pushing them beyond their capabilities.

Consumer EEG devices are fundamentally noisy, and EEG signals contain biometric markers that can uniquely identify individuals [69]. As LLMs process data externally, privacy concerns must be addressed before large-scale adoption of EEG-driven AI tutors.

7 CONCLUSION

We provide evidence that EEG-driven AI chatbots can enhance engagement, bringing users closer into a zone of proximal development, but highlight challenges in translating engagement into learning gains. Future neuroadaptive AI systems must refine adaptation mechanisms, personalize content, and balance engagement with cognitive load. By integrating multimodal sensing and long-term user modeling, AI tutors could one day provide truly personalized, lifelong learning experiences.

ACKNOWLEDGMENTS

We thank Treyden Chiaravalloti for his valuable piloting support and insightful feedback. We also appreciate Protyasha Nishat's expertise in signal processing and Nathan Whitmore's comments on study design. Luisa Heiss's thorough grading of the essays was instrumental in evaluating participant test performance without bias. This research was supported by the MIT J-WEL Education Innovation Grant.

REFERENCES

[1] 2023. Imagination Engine I: Generating Abstract Art through EEG. https://www.augmentationlab.org/post/imagination-engine-i-generating-abstract-art-through-eeg.

[2] 2023. *Technology in education.* Technical Report. UNESCO.

[3] Yomna Abdelrahman, Mariam Hassib, Maria Guinea Marquez, Markus Funk, and Albrecht Schmidt. 2015. Implicit engagement detection for interactive museums using brain-computer interfaces. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. ACM. doi:10.1145/2786567.2793709

[4] Lena M Andreessen, Peter Gerjets, Detmar Meurers, and Thorsten O Zander. 2021. Toward neuroadaptive support technologies for improving digital reading: a passive BCI-based assessment of mental workload imposed by text difficulty and presentation speed during reading. *User Model. User-adapt Interact.* 31 (March 2021), 75–104. doi:10.1007/s11257-020-09273-5

[5] Andrea Apicella, Pasquale Arpaia, Mirco Frosolone, Giovanni Improta, Nicola Moccaldi, and Andrea Pollastro. 2022. EEG-based measurement system for monitoring student engagement in learning 4.0. *Sci. Rep.* 12 (7 April 2022), 5857. doi:10.1038/s41598-022-09578-y

[6] Roger Azevedo. 2015. Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educ. Psychol.* 50 (2 Jan. 2015), 84–94. doi:10.1080/00461520.2015.1004069

[7] Yunpeng Bai, Xintao Wang, Yan-Pei Cao, Yixiao Ge, Chun Yuan, and Ying Shan. 2025. DreamDiffusion: High-quality EEG-to-image generation with temporal masked signal modeling and CLIP alignment. In *Lecture Notes in Computer Science*. Springer Nature Switzerland, 472–488. doi:10.1007/978-3-031-72751-1_27

[8] N P Bechtereva and V B Gretchin. 1968. Physiological foundations of mental activity. *Int. Rev. Neurobiol.* 11 (1968), 329–352. doi:10.1016/s0074-7742(08)60392-x

[9] Antoine Bellemare-Pepin, Philipp Thölke, Yann Harel, and Karim Jerbi. 2024. Real-Time Neuro-Augmented Cinema via Generative AI. NeurIPS Workshop on Creativity & Generative AI.

[10] C Berka, D J Levendowski, M N Lumicao, A Yau, G Davis, V T Zivkovic, R E Olmstead, P D Tremoulet, and P L Craven. 2007. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine* 78 (May 2007).

[11] Wolfram Boucsein, Andrea Haarmann, and Florian Schaefer. 2007. Combining skin conductance and heart rate variability for adaptive automation during simulated IFR flight. In *Engineering Psychology and Cognitive Ergonomics*. Springer Berlin Heidelberg, 639–647. doi:10.1007/978-3-540-73331-7_70

[12] E A Byrne and R Parasuraman. 1996. Psychophysiology and adaptive automation. *Biol. Psychol.* 42 (5 Feb. 1996), 249–268. doi:10.1016/0301-0511(95)05161-9

[13] Andrea Caria. 2024. Towards predictive communication with brain-computer interfaces integrating large language models. *arXiv [cs.HC]* (10 Dec. 2024).

[14] Hongyu Chen, Weiming Zeng, Chengcheng Chen, Luhui Cai, Fei Wang, Yuhu Shi, Lei Wang, Wei Zhang, Yueyang Li, Hongjie Yan, Wai Ting Siok, and Nizhuan Wang. 2024. EEG Emotion Copilot: Optimizing lightweight LLMs for emotional EEG interpretation with assisted medical record generation. *arXiv [cs.CV]* (30 Sept. 2024).

[15] Wenhui Cui, Woojae Jeong, Philipp Thölke, Takfarinas Medani, Karim Jerbi, Anand A Joshi, and Richard M Leahy. 2024. Neuro-GPT: Towards A foundation model for EEG. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, Vol. 35. IEEE, 1–5. doi:10.1109/isbi56570.2024.10635453

[16] Alex Dan and Miriam Reiner. 2017. Real time EEG based measurements of cognitive load indicates mental states during learning. *JEDM* 9 (23 Dec. 2017), 31–44. doi:10.5281/ZENODO.3554719

[17] David Game College. 2024. GCSE AI Adaptive Learning Programme. https://www.davidgamecollege.com/courses/courses-overview/item/102/gcse-

ai-adaptive-learning-programme.
[18] Ido Davidesco, Noah Glaser, Ian H Stevenson, and Or Dagan. 2023. Detecting fluctuations in student engagement and retention during video lectures using electroencephalography. *Br. J. Educ. Technol.* 54 (Nov. 2023), 1895–1916. doi:10.1111/bjet.13330
[19] Ton de Jong. 2010. Cognitive load theory, educational research, and instructional design: some food for thought. *Instr. Sci.* 38 (March 2010), 105–134. doi:10.1007/s11251-009-9110-0
[20] Ruiqi Deng, Maoli Jiang, Xinlu Yu, Yuyan Lu, and Shasha Liu. 2025. Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Comput. Educ.* 227 (1 April 2025), 105224. doi:10.1016/j.compedu.2024.105224
[21] Seif Eldawlatly. 2024. On the role of generative artificial intelligence in the development of brain-computer interfaces. *BMC Biomed. Eng.* 6 (2 May 2024), 4. doi:10.1186/s42490-024-00080-2
[22] Atef Eldenfria and Hosam Al-Samarraie. 2019. Towards an online continuous adaptation mechanism (OCAM) for enhanced engagement: An EEG study. *Int. J. Hum. Comput. Interact.* 35 (14 Dec. 2019), 1960–1974. doi:10.1080/10447318.2019.1595303
[23] Harry Barton Essel, Dimitrios Vlachopoulos, Albert Benjamin Essuman, and John Opuni Amankwa. 2024. ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence* 6 (1 June 2024), 100198. doi:10.1016/j.caeai.2023.100198
[24] Stephen H Fairclough, Liverpool John Moores, Katie C Ewing, and Jenna Roberts. 2009. Measuring task engagement as an input to physiological computing. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. IEEE. doi:10.1109/acii.2009.5349483
[25] Shalom M Fisch. 2017. Bridging theory and practice: Applying cognitive and educational theory to the design of educational media. In *Cognitive Development in Digital Contexts*. Elsevier, 217–234. doi:10.1016/b978-0-12-809481-5.00011-0
[26] Jennifer A Fredricks, Phyllis C Blumenfeld, and Alison H Paris. 2004. School engagement: Potential of the concept, state of the evidence. *Rev. Educ. Res.* 74 (1 March 2004), 59–109. doi:10.3102/00346543074001059
[27] F G Freeman, P J Mikulka, L J Prinzel, and M W Scerbo. 1999. Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. *Biological psychology* 50 (May 1999). doi:10.1016/s0301-0511(99)00002-2
[28] Jérémy Frey, May Grabli, Ronit Slyper, and Jessica R Cauchard. 2018. Breeze: Sharing biofeedback through wearable technologies. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM. doi:10.1145/3173574.3174219
[29] Sabine Graf, Tzu-Chien Liu, Kinshuk, Nian-Shing Chen, and Stephen J H Yang. 2009. Learning styles and cognitive traits – Their relationship and its benefits in web-based educational systems. *Comput. Human Behav.* 25 (1 Nov. 2009), 1280–1289. doi:10.1016/j.chb.2009.06.005
[30] Jennie K Grammer, Keye Xu, and Agatha Lenartowicz. 2021. Effects of context on the neural correlates of attention in a college classroom. *NPJ Sci. Learn.* 6 (6 July 2021), 15. doi:10.1038/s41539-021-00094-8
[31] Simone Grassini, Giulia Virginia Segurini, and Mika Koivisto. 2022. Watching nature videos promotes physiological restoration: Evidence from the modulation of alpha waves in electroencephalography. *Front. Psychol.* 13 (7 June 2022), 871143. doi:10.3389/fpsyg.2022.871143
[32] Sven Guenther, Nataliya Kosmyna, and Pattie Maes. 2024. Image classification and reconstruction from low-density EEG. *Sci. Rep.* 14 (16 July 2024), 16436. doi:10.1038/s41598-024-66228-1
[33] Mariam Hassib, Mohamed Khamis, Susanne Friedl, Stefan Schneegass, and Florian Alt. 2017. Brainatwork: logging cognitive engagement and tasks in the workplace using electroencephalography. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*. ACM. doi:10.1145/3152832.3152865
[34] Mariam Hassib, Stefan Schneegass, Philipp Eiglsperger, Niels Henze, Albrecht Schmidt, and Florian Alt. 2017. EngageMeter: A system for implicit audience engagement sensing using electroencephalography. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM. doi:10.1145/3025453.3025669
[35] Yuk Mui Elly Heung and Thomas K F Chiu. 2025. How ChatGPT impacts student engagement from a systematic review and meta-analysis study. *Computers and Education: Artificial Intelligence* 8 (1 June 2025), 100361. doi:10.1016/j.caeai.2025.100361
[36] Nora Hollenstein, Jonathan Rotsztejn, Marius Troendle, Andreas Pedroni, Ce Zhang, and Nicolas Langer. 2018. ZuCo, a simultaneous EEG and eye-tracking resource for natural sentence reading. *Sci Data* 5 (11 Dec. 2018), 180291. doi:10.1038/sdata.2018.291
[37] Anu Holm, Kristian Lukander, Jussi Korpela, Mikael Sallinen, and Kiti M I Müller. 2009. Estimating brain load from the EEG. *ScientificWorldJournal* 9 (14 July 2009), 639–651. doi:10.1100/tsw.2009.83
[38] Yongquan Hu, Shuning Zhang, Ting Dang, Hong Jia, Flora D Salim, Wen Hu, and Aaron J Quigley. 2024. Exploring large-scale language models to evaluate EEG-based multimodal data for mental health. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 412–417. doi:10.1145/3675094.3678494
[39] Jin Huang, Chun Yu, Yuntao Wang, Yuhang Zhao, Siqi Liu, Chou Mo, Jie Liu, Lie Zhang, and Yuanchun Shi. 2014. FOCUS: enhancing children's engagement in reading by using contextual BCI training sessions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. doi:10.1145/2556288.2557339
[40] Stephen Hutt, Kristina Krasich, James R. Brockmole, and Sidney K. D'Mello. 2021. Breaking out of the lab: Mitigating mind wandering with gaze-based attention-aware technology in classrooms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM. doi:10.1145/3411764.3445269
[41] Stephen Hutt, Caitlin Mills, Nigel Bosch, Kristina Krasich, James Brockmole, and Sidney D'Mello. 2017. Out of the fr-eye-ing pan: Towards gaze-based models of attention during learning with technology in the classroom. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*. ACM. doi:10.1145/3079628.3079669
[42] InteraXon. 2025. Muse: the brain sensing headband Store with Worldwide Shipping. https://choosemuse.com/.
[43] Wei-Bang Jiang, Yansen Wang, Bao-Liang Lu, and Dongsheng Li. 2024. NeuroLM: A universal multi-task foundation model for bridging the gap between language and EEG signals. *arXiv [eess.SP]* (27 Aug. 2024).
[44] A Kamzanova, G Matthews, A Kustubayeva, and S Jakupov. 2011. EEG indices to time-on-task effects and to a workload manipulation (cueing). *International Scholarly and Scientific Research & Innovation* 5 (23 Aug. 2011), 928–931.
[45] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individ. Differ.* 103 (1 April 2023), 102274. doi:10.1016/j.lindif.2023.102274
[46] KhanAcademy. 2025. Meet Khanmigo: Khan Academy's AI-powered teaching assistant & tutor. https://www.khanmigo.ai/.
[47] Asma Ben Khedher, Imène Jraidi, and Claude Frasson. 2019. Tracking students' mental engagement using EEG signals during an interaction with a virtual learning environment. *J. Intell. Learn. Syst. Appl.* 11 (2019), 1–14. doi:10.4236/jilsa.2019.111001
[48] Jonathan W Kim, Ahmed Alaa, and Danilo Bernardo. 2024. EEG-GPT: Exploring capabilities of large language models for EEG classification and interpretation. *arXiv [q-bio.QM]* (31 Jan. 2024). doi:10.48550/arXiv.2401.18006
[49] Eckart Klieme, Frank Lipowsky, Katrin Rakoczy, and Nadja Ratzka. 2006. Qualitätsdimensionen und wirksamkeit von mathematikunterricht. *Untersuchungen zur Bildungsqualität von Schule* (2006), 127–146.
[50] Nataliya Kosmyna, Eugene Hauptmann, and Yasmeen Hmaidan. 2023. A brain-controlled quadruped robot: A proof-of-concept demonstration. *Sensors (Basel)* 24 (22 Dec. 2023), 80. doi:10.3390/s24010080
[51] Nataliya Kosmyna and Pattie Maes. 2019. AttentivU: An EEG-Based Closed-Loop Biofeedback System for Real-Time Monitoring and Improvement of Engagement for Personalized Learning. *Sensors* 19 (27 Nov. 2019). doi:10.3390/s19235200
[52] Dong Hyeok Lee and Chun Kee Chung. 2024. Enhancing neural decoding with large language models: A GPT-based approach. In *2024 12th International Winter Conference on Brain-Computer Interface (BCI)*. IEEE, 1–4. doi:10.1109/bci60775.2024.10480499
[53] Joanne Leong, Pat Pataranutaporn, Valdemar Danry, Florian Perteneder, Yaoli Mao, and Pattie Maes. 2024. Putting things into context: Generative AI-enabled context personalization for vocabulary learning improves learning motivation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Vol. 32. ACM, 1–15. doi:10.1145/3613904.3642393
[54] Marco Marchesi and Bruno Riccò. 2013. BRAVO: a brain virtual operator for education exploiting brain-computer interfaces. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*. ACM. doi:10.1145/2468356.2479618
[55] Caitlin Mills, Igor Fridman, Walid Soussou, Disha Waghray, Andrew M Olney, and Sidney K D'Mello. 2017. Put your thinking cap on: detecting cognitive load using EEG during learning. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. ACM. doi:10.1145/3027385.3027431
[56] Heni Mulyani, Mohamad Azim Istiaq, Elvia R Shauki, Fitrina Kurniati, and Hanifia Arlinda. 2025. Transforming education: exploring the influence of generative AI on teaching performance. *Cogent Educ.* 12 (31 Dec. 2025). doi:10.1080/2331186x.2024.2448066
[57] B S Oken, M C Salinsky, and S M Elsas. 2006. Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clin. Neurophysiol.* 117 (Sept. 2006), 1885–1901. doi:10.1016/j.clinph.2006.01.017
[58] D F O Onah, J E Sinclair, and R Boyatt. 2014. Dropout rates of massive open online courses: Behavioural patterns. Unpublished. doi:10.13140/RG.2.1.2402.0009
[59] OpenAI. 2023. Teaching with AI. https://openai.com/index/teaching-with-ai/.
[60] OpenAI. 2024. Introducing ChatGPT Edu. https://openai.com/index/introducing-chatgpt-edu/.

[61] OpenAI. 2024. Memory and new controls for ChatGPT. https://openai.com/index/memory-and-new-controls-for-chatgpt/.
[62] Brendan Parsons and Jocelyn Faubert. 2021. Enhancing learning in a perceptual-cognitive training paradigm using EEG-neurofeedback. *Sci. Rep.* 11 (18 Feb. 2021), 4061. doi:10.1038/s41598-021-83456-x
[63] Salil H Patel and Pierre N Azzam. 2005. Characterization of N200 and P300: selected studies of the Event-Related Potential. *Int. J. Med. Sci.* 2 (1 Oct. 2005), 147–154. doi:10.7150/ijms.2.147
[64] A T Pope, E H Bogart, and D S Bartolome. 1995. Biocybernetic system evaluates indices of operator engagement in automated task. *Biol. Psychol.* 40 (May 1995), 187–195. doi:10.1016/0301-0511(95)05116-3
[65] Mirko Raca and Pierre Dillenbourg. 2013. System for assessing classroom attention. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*. ACM. doi:10.1145/2460296.2460351
[66] Mirko Raca, Lukasz Kidzinski, and Pierre Dillenbourg. 2015. Translating Head Motion into Attention - Towards Processing of Student's Body-Language. *International Educational Data Mining Society* (June 2015).
[67] Johnmarshall Reeve and Ching-Mei Tseng. 2011. Agency as a fourth aspect of students' engagement during learning activities. *Contemp. Educ. Psychol.* 36 (1 Oct. 2011), 257–267. doi:10.1016/j.cedpsych.2011.05.002
[68] Lauren E Reinerman, Gerald Matthews, Joel S Warm, Lisa K Langheim, Kelley Parsons, Christina A Proctor, Tazeen Siraj, Lloyd D Tripp, and Robert M Stutz. 2006. Cerebral blood flow velocity and task engagement as predictors of vigilance performance. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 50 (Oct. 2006), 1254–1258. doi:10.1177/154193120605001210
[69] Maria V Ruiz-Blondet, Zhanpeng Jin, and Sarah Laszlo. 2016. CEREBRE: A novel method for very high accuracy event-related potential biometric identification. *IEEE Trans. Inf. Forensics Secur.* 11 (July 2016), 1618–1629. doi:10.1109/tifs.2016.2543524
[70] Yashvir Sabharwal and Balaji Rama. 2024. Comprehensive review of EEG-to-output research: Decoding neural signals into images, videos, and audio. *arXiv [cs.CV]* (27 Dec. 2024).
[71] Akane Sano, Judith Amores, and Mary Czerwinski. 2024. Exploration of LLMs, EEG, and behavioral data to measure and support attention and sleep. *arXiv [eess.SP]* (1 Aug. 2024).
[72] Brooke Schultz. 2025. This School Will Have Artificial Intelligence Teach Kids (With Some Human Help). *Education Week* (6 Jan. 2025).
[73] Skipper Seabold and Josef Perktold. 2010. Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the Python in Science Conference*. SciPy, 92–96. doi:10.25080/majora-92bf1922-011
[74] Uri Shaked. 2021. muse-js: Muse 2016 EEG Headset JavaScript Library (using Web Bluetooth). https://github.com/urish/muse-js
[75] David J Shernoff, Mihaly Csikszentmihalyi, Barbara Schneider, and Elisa Steele Shernoff. 2014. Student engagement in high school classrooms from the perspective of flow theory. In *Applications of Flow in Human Development and Education*. Springer Netherlands, 475–494. doi:10.1007/978-94-017-9094-9_24
[76] Gale M Sinatra, Benjamin C Heddy, and Doug Lombardi. 2015. The challenges of defining and measuring student engagement in science. *Educ. Psychol.* 50 (2 Jan. 2015), 1–13. doi:10.1080/00461520.2014.1002924
[77] W Speier, C Arnold, and N Pouratian. 2016. Integrating language models into classifiers for BCI communication: a review. *J. Neural Eng.* 13 (6 June 2016), 031002. doi:10.1088/1741-2560/13/3/031002
[78] Ricarda Steinmayr, Anne F Weidinger, Malte Schwinger, and Birgit Spinath. 2019. The importance of students' motivation for their academic achievement - replicating and extending previous findings. *Front. Psychol.* 10 (31 July 2019), 1730. doi:10.3389/fpsyg.2019.01730
[79] Daniel Szafir and Bilge Mutlu. 2012. Pay attention!: designing adaptive agents that monitor and improve user engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. doi:10.1145/2207676.2207679
[80] The Abdul Latif Jameel Poverty Action Lab (J-PAL). 2022. Teaching at the Right Level to improve learning. https://www.povertyactionlab.org/case-study/teaching-right-level-improve-learning.
[81] The Open Innovation Team and Department for Education. 2024. *Generative AI in education - Educator and expert views*. Technical Report. UK Department for Education.
[82] Philipp Tholke, Antoine Bellemare-Pepin, Yann Harel, Francois Lespinasse, and Karim Jerbi. 2024. Bio-Mechanical Poet: An Immersive Audiovisual Playground for Brain Signals and Generative AI. In *Proceedings of the 15th international conference on computational creativity*, In Kazjon Grace, Maria Teresa Llano, Martins Pedro, and Maria M Hedblom (Eds.). 65–74.
[83] Angela Vujic, Shreyas Nisal, and Pattie Maes. 2023. Joie: a Joy-based Brain-Computer Interface (BCI). In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, 1–14. doi:10.1145/3586183.3606761
[84] Yansen Wang and Zilong Wang. 2024. EEG2Video: Towards Decoding Dynamic Visual Perception from EEG Signals. The Thirty-eighth Annual Conference on Neural Information Processing Systems (NeurIPS 2024).
[85] Łukasz Warchoł and Ludmiła Zając-Lamparska. 2023. The relationship of N200 and P300 amplitudes with intelligence, working memory, and attentional control behavioral measures in young healthy individuals. *Adv. Cogn. Psychol.* 19 (2023), 63–75. doi:10.5709/acp-0404-2
[86] Shaoyue Wen, Michael Middleton, Songming Ping, Nayan N Chawla, Guande Wu, Bradley S Feest, Chihab Nadri, Yunmei Liu, David Kaber, Maryam Zahabi, Ryan P McMahan, Sonia Castelo, Ryan Mckendrick, Jing Qian, and Claudio Silva. 2025. AdaptiveCoPilot: Design and testing of a NeuroAdaptive LLM cockpit guidance system in both novice and expert pilots. *arXiv [cs.HC]* (7 Jan. 2025).
[87] Beste F Yuksel, Kurt B Oleson, Lane Harrison, Evan M Peck, Daniel Afergan, Remco Chang, and Robert J K Jacob. 2016. Learn Piano with BACh: An Adaptive Learning Interface that Adjusts Task Difficulty Based on Brain State. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, 5372–5384. doi:10.1145/2858036.2858388
[88] Xiayin Zhang, Ziyue Ma, Huaijin Zheng, Tongkeng Li, Kexin Chen, Xun Wang, Chenting Liu, Linxi Xu, Xiaohang Wu, Duoru Lin, and Haotian Lin. 2020. The combination of brain-computer interfaces and artificial intelligence: applications and challenges. *Ann. Transl. Med.* 8 (June 2020), 712. doi:10.21037/atm.2019.11.109
[89] Yuhong Zhang, Qin Li, Sujal Nahata, Tasnia Jamal, Shih-Kuen Cheng, Gert Cauwenberghs, and Tzyy-Ping Jung. 2024. Integrating large language model, EEG, and eye-tracking for word-level neural state classification in reading comprehension. *IEEE Trans. Neural Syst. Rehabil. Eng.* PP (14 Aug. 2024), 1–1. doi:10.1109/TNSRE.2024.3435460
[90] Tong Zhou, Xuhang Chen, Yanyan Shen, Martin Nieuwoudt, Chi-Man Pun, and Shuqiang Wang. 2023. Generative AI enables EEG data augmentation for Alzheimer's disease detection via diffusion model. In *2023 IEEE International Symposium on Product Compliance Engineering - Asia (ISPCE-ASIA)*. IEEE, 1–6. doi:10.1109/ispce-asia60405.2023.10365931

A SYSTEM PROMPTS

A.1 NeuroChat system prompt

NeuroChat System Prompt

You are an encouraging tutor who helps students across various subjects and skill levels understand concepts by explaining ideas and asking students questions. Start by introducing yourself to the student as their AI-Tutor who is happy to help them with any questions.

Additionally, you will be provided with the student's cognitive load values while they were reading any previous responses of yours as measured by EEG. Your goal is to act like a good tutor, using the insights from these metrics to adapt your responses to the student's cognitive load dynamically. The value you will be given:

Normalized engagement score: This represents the user's level of engagement or arousal on a normalized scale from 0 to 1. The engagement index is a ratio of the student's beta/(theta+alpha) bands.

Do not ever disclose the EEG metrics to the user since they are hidden to them. Also, never make direct comments on their metrics and don't mention the names of the metrics.

Give students explanations, examples, and analogies about the concept to help them understand.

Adaptations Based on Cognitive Load:
You need to learn how the user reacted to your adaptations. Based on their cognitive load, modulate the response length, factual vs. storytelling, ease of text (explain like I'm 5 vs. explain like I'm a PhD), bullet points

vs. long-form text, level of depth and detail, Socratic questions, and styling of text (bolding of keywords). Every person is different; however, here are some general pointers:

- Students with higher cognitive loads enjoy more complex, scientific, and in-depth explorations of a topic. Students with low to medium cognitive load may prefer explanations that prompt more of their curiosity or represent a challenge.
- Students with lower cognitive load need to discover a question they are curious about; hence provide explanations that prompt more of their curiosity or represent a challenge. You may also give them interesting facts or narrative examples, or ask them questions.
- Once a student shows an appropriate level of understanding given their learning level and cognitive load, ask them to explain the concept in their own words; this is the best way to show you know something, or ask them for examples.
- Encourage learners to explain their thinking.
- If the learner needs more engagement, provide thought-provoking questions or exercises. Also, suggest questions to explore together. Students with higher cognitive load may also find these interesting.
- Provide positive reinforcement but also critical feedback.
- Offer clarification or examples if the user seems to need more understanding.
- Analogies or storytelling can help raise cognitive load.
- Sounding more energetic or scientific can help raise cognitive load.
- Bolding important keywords can help raise and maintain cognitive load.

Remember, your role is to support the user's learning journey, adapt to their needs, and ensure a positive, effective, and engaging educational experience. Be patient, encouraging, and responsive to the user's cognitive state and feedback.

By following these guidelines, you will help users achieve their learning goals effectively and enjoyably. Respond in Markdown.

A.2 Control condition prompt

Control Condition Prompt

You are an AI Tutor designed to assist learners across a variety of subjects and skill levels. Your primary goal is to provide clear, accurate, and engaging explanations tailored to each learner's needs. You should strive to be patient, encouraging, and adaptive in your teaching style.

Respond in Markdown.

B QUIZ QUESTIONS AND ESSAY PROMPTS

T. rex

Fill-in-the-blank Questions.

(1) **Q1:** T. rex fossils have primarily been found in _______. (2 points): **North America**
(2) **Q2:** T. Rex likely obtained its food from hunting as well as _______. (2 points): **Scavenging**
(3) **Q3:** One of the most complete T. rex fossils, known as _______, was discovered in South Dakota in 1990. (2 points): **Sue**

Multiple Choice Questions. **Q4:** Which era did the Tyrannosaurus Rex live in?

- Jurassic
- Cretaceous
- Triassic
- Permian
- Devonian
- This did not come up in my conversation
- I don't know

(2 points): **Cretaceous**

Q5: How many fingers did the Tyrannosaurus Rex have on each hand?

- Two
- Three
- Four
- Five
- Six
- This did not come up in my conversation
- I don't know

(2 points): **Two**

Q6: Choose all popular myths about T. Rex that are likely wrong.

- It had poor vision.
- It was a slow, clumsy mover.
- Its arms were likely useless.
- It had one of the most powerful bites ever known.
- It was the dominant dinosaur in its environment.
- It may have had feathers.

(6 points): Correct answers are the first three options: **It had poor vision**; **It was a slow, clumsy mover**; **Its arms were likely useless**. For partial credit, 1 point was awarded for each correct choice, and 1 point for not selecting each incorrect choice.

Ranking Question. **Q9:** Sort the T. Rex into this list of dinosaurs by length.

(1) Argentinosaurus - 98 ft (30 m)
(2) Brachiosaurus - 82 ft (25 m)
(3) Spinosaurus - 59 ft (18 m)
(4) Iguanodon - 33 ft (10 m)
(5) Pachycephalosaurus - 16 ft (5 m)
(6) Velociraptor - 7 ft (2.5 m)
(7) Tyrannosaurus Rex - ?

(2 points): Tyrannosaurus Rex (40 ft) should be placed **between Iguanodon and Spinosaurus**. (1 point): Placement is one off, in position 3 or 5.

Essay Prompt. Answer one of the following in the form of a mini-essay (introduction, main section, conclusion):

(1) Discuss T. Rex's physical and behavioral characteristics which enabled it to dominate its environment.
(2) Analyze paleobiological discoveries that have changed our understanding of T. Rex.
(3) Evaluate the theories regarding the function of T. Rex's small arms. What are some of the proposed explanations, and which do you find most convincing?
(4) Create your own essay question related to the T. Rex based on your conversation.

Taiping Rebellion

Fill-in-the-gap Questions.

(1) **Q1:** The goal of the Taiping Rebellion was to establish the ______. (2 points): **Taiping Heavenly Kingdom of Great Peace**, or similar.
(2) **Q2:** One of the distinctive aspects of the Taiping ideology was its spiritual blend of ______ and ______, which appealed to the disaffected rural populace. (2 points): **Christianity** (1) and **Chinese spiritual traditions** (Buddhism, Taoism, Confucianism) or similar (1).

Multiple Choice Questions. **Q5:** What was the name of the leader of the Taiping Rebellion?

- Zeng Guofan
- Hong Xiuquan
- Hong Tianguifu
- Feng Yunshan
- Yang Xiuqing
- This did not come up in my conversation
- I don't know

(1 point): **Hong Xiuquan**

Q6: The Qing government was administered by leaders of which ethnic minority?

- Manchu
- Hakka
- Han
- Zhuang
- Hui
- This did not come up in my conversation
- I don't know

(1 point): **Manchu**

Ranking Question. **Q10:** Here is a list of death tolls in wars by lowest estimate. Where would the Taiping Rebellion sit?

(1) World War II - 80 million
(2) World War I - 17 million
(3) Spanish conquest of Mexico - 10.5 million
(4) Russian Civil War - 7 million
(5) Napoleonic Wars - 3.5 million
(6) Vietnam War - 1.3 million
(7) Taiping Rebellion

(2 points): **Between World War II and World War I**, with an estimated toll of 20-30 million.

Essay Prompt. Answer one of the following in the form of an essay (introduction, main section, conclusion):

(1) Discuss the socio-economic factors that contributed to the outbreak of the Taiping Rebellion.
(2) Analyze the role of religion in the Taiping Rebellion.
(3) Discuss the legacy of the Taiping Rebellion on subsequent Chinese history.
(4) Create your own essay question related to the Taiping Rebellion based on your conversation.