TU Delft

Toward Emotion-Aware AI Agents:
A General Framework for Real-Time EEG-Driven Human–Agent Collaboration

Author: Ziyu Wei

MSc Integrated Product Design
Faculty of Industrial Design Engineering
Delft University of Technology

Supervisory Team: Dr. P. Wang (Chair) & Dr. R.S.K. Chandrasegaran (Mentor)

Toward Emotion-Aware AI Agents: A General Framework for Real-Time EEG-Driven Human–Agent Collaboration

Abstract

Generative AI reshapes creative workflows through mixed-initiative collaboration, requiring users to rapidly make decisions among alternatives and fluidly shift roles as creators and critics. During exploration, users often rely on emotional intuition for swift judgments. However, most current systems depend on explicit prompts, limiting their adaptivity across iterations. Meanwhile, EEG-based neuroadaptive systems dynamically adjust interactions, improving human–machine alignment. EmotivChat addresses this gap by continuously monitoring users' immediate emotional responses. Within a closed-loop architecture, the system optimizes prompting strategies across rounds. The system was evaluated across two paired studies on image co-creation(N=31) and text ideation (N=29) each contrasting the EEG-adaptive system with a rigorously matched non-adaptive baseline. Results indicate EmotivChat improved perceived teaming and iteration efficiency in image co-creation, while in text ideation, it improved perceived collaboration but produced limited changes in idea outcomes. These findings validate real-time emotional feedback in LLM-based interactions, pointing the way toward cross-modal, emotion-adaptive agents.

1 Introduction

Human-AI collaboration is rapidly evolving from simple command-response interactions to dynamic, co-creative partnerships [37]. As generative AI systems become more sophisticated, they increasingly participate as creative partners in tasks ranging from artistic design to scientific discovery [29, 40, 51]. However, current collaboration paradigms remain fundamentally asymmetric: while AI agents can generate, iterate, and refine outputs at machine speed [45, 47], they lack awareness of users' evolving emotional and cognitive states during the collaborative process.

This limitation becomes particularly pronounced in creative domains, where human collaborators frequently experience rapid shifts in inspiration, frustration, and satisfaction [17, 20]. During iterative design sessions, users may feel excited by a promising direction, frustrated by repeated failures, or cognitively overloaded when exploring complex solution spaces. These emotional dynamics profoundly influence creative decision-making [32], determining when to persist with refinements versus when to explore entirely new approaches [16]. Yet most AI systems remain oblivious to these crucial human signals, relying solely on explicit textual feedback to guide collaboration.

While other biosignals such as heart rate variability, electrocardiograms, or eye-tracking have been explored for affect-adaptive interfaces [43], they are often too slow, too shallow, or too context-dependent for real-time creative collaboration. Electroencephalography (EEG) offers unique advantages: millisecond-scale temporal resolution, direct access to neural dynamics tracking emotion and cognitive load, and signals that are difficult to suppress consciously [14, 53]. The emergence of consumer-grade EEG devices has made real-time neural sensing technically feasible [56], providing a continuous stream of implicit feedback that complements explicit user instructions. This capability has seeded applications across adaptive learning [11], psychological support [52], and creative co-creation such as EEG-to-image diffusion and affect-aware music generation [33, 61].

However, integrating real-time neural feedback into AI collaboration presents significant technical and design challenges. Raw EEG signals are noisy and require sophisticated processing pipelines [50]. More fundamentally, it remains unclear how physiological signals should influence AI behavior. Should neural data directly modulate model parameters, or should it gate higher-level interaction policies? When should systems adapt, and when might adaptation prove disruptive to user workflow?

Recent advances in large language models (LLMs) provide a promising foundation for neuroadaptive AI collaboration [15, 19]. Unlike previous generation systems that required extensive retraining for personalization, LLMs can modify their behavior through prompt-level interventions without parameter updates [18]. This architectural property enables lightweight, real-time adaptation while maintaining system stability and auditability [30]. By treating EEG signals as discrete policy selectors rather than continuous control inputs, we can create neuroadaptive systems that remain interpretable and debuggable [21, 31].

However, despite these promising prototypes, a reusable cross-task EEG-agent interaction framework is still missing, and the closed-loop dynamics of 'EEG affect trajectory → agent decision → user behavior' remain poorly understood. [49]. To address this gap, this research introduces and validates a unified framework, which covers signal decoding, interface information flow, and experimental evaluation, that systematically examines when and how

---

EEG-derived emotion signals enhance open human-agent interaction and distills design guidelines for future multimodal affect adaptive systems [46].

- **RQ1: What is the effect of prompt adaptation driven by real-time EEG on the quality of human-AI collaboration?** More specifically, I ask the following questions:
  - How does real-time EEG-based detection of cognitive-emotional states (excitement, stress) influence prompt adaptation strategies?
  - How do EEG-adaptive systems compare to fixed-prompt systems in terms of user satisfaction and perceived collaboration quality?
  - What is the relationship between EEG-detected emotional states and subsequent user behavior in collaborative tasks?
- **RQ2: How does the effectiveness of EEG-adaptive prompting vary between different creative tasks and contexts?** This question can be broken down into the following components:
  - What are the differences in how EEG-adaptive prompting affects image creation compared to text-based ideation ?
  - What makes EEG-adaptive prompting work better in some creative tasks than others?
  - When do users find EEG-based prompt adaptation helpful, and when do they find them annoying or disruptive?
- **RQ3:What are the key design considerations for building reliable EEG-adaptive prompting systems?** This translates to the following:
  - What technical approaches (EEG signal filtering, stress/excitement detection thresholds) are most reliable for real-time prompt adaptation?
  - How should the system decide when a user's stress or excitement level is high enough to trigger a prompt change?
  - What are the key implementation considerations for deploying such systems across different creative workflows?

We make three primary contributions:

- **A real-time EEG-to-prompt adaptation framework** We propose the systematic framework that integrates EEG emotional signals into large language model prompting strategies through discrete tag mapping. The framework employs event-anchored windows and explicit policy branches to ensure stable, auditable real-time adaptation.
- **Empirical validation across two modalities** We conducted rigorous paired studies across image co-creation (N=31) and text ideation (N=29) domains, finding that EEG adaptation significantly improves subjective experience in image tasks while enhancing collaboration perception but showing limited impact on objective outcomes in text tasks.
- **Transferable design principles for neuroadaptive prompting** Based on experimental results, we propose five design principles extensible to other physiological signals.

2 Related Work

2.1 Emotion-Aware Human-Computer Interaction

2.1.1 Emotional Dynamics in Creative Collaboration. In psychology and cognitive science, emotion is regarded as a key driving force for human creativity and decision-making. As a core regulatory variable in cognitive processing, emotional states can alter our attentional focus, memory organization, and information processing strategies [42]. A substantial body of research has shown that emotions shape people's patterns of creative thinking and the quality of their output. In general, positive emotions, particularly high-arousal, approach-motivated states such as joy and excitement, help stimulate divergent thinking and imagination, thereby enhancing creative performance [9]. A meta-analysis by Baas et al. further supports this conclusion: positive emotions (especially high-arousal, approach-oriented pleasant moods) significantly enhance creativity, while negative emotions as a whole show no clear facilitative effect [9]. However, experimental evidence indicates that compared to neutral states, positive emotions broaden the attentional scope of people toward visual and cognitive signals, while negative emotions tend to narrow and focus thought [27].

Emotion influences not only static cognitive styles, but also the tendencies of people to make decisions between 'exploration' (trying new options) and 'exploitation' (sticking to existing ones). High-arousal emotions often increase the urgency and boldness of decisions. Positive high-arousal emotions (such as excitement and joy) are typically accompanied by a stronger motivation for the approach, which makes individuals more willing to experiment with new strategies and explore unknown possibilities [48]. Therefore, in systems designed for creative human–AI collaboration, neglecting the role of emotion may lead to misalignment with the actual decision-making pathways of users, thus introducing additional costs to the use of the system.

2.1.2 Evidence in HCI: Emotional input has been employed to enhance user experience, regulate cognitive load, and deepen immersion. In the field of human-computer interaction, extensive research has shown that incorporating users' emotions as input signals can significantly improve interaction experience and system performance. For example, in intelligent tutoring systems, the system dynamically adjusts teaching content and feedback through real-time emotion recognition (such as facial expressions and physiological signals) to adapt to students' emotional states, thereby enhancing engagement and effectiveness in the learning process [60]. In the entertainment domain, emotion-driven games and virtual reality applications dynamically adjust game difficulty or virtual environments based on players' emotions (detected through EEG, heart rate, etc.). This adaptive mechanism has been proven more effective than fixed settings in maintaining player engagement and preventing fatigue [41]. Similarly, adaptive user interfaces adjust interface complexity or content based on users' stress or emotions (such as capturing expressions and postures through cameras), resulting in significantly higher user productivity and satisfaction compared to using static interfaces [8]. These research cases demonstrate that integrating emotional signals into interaction design is not only

Table 1: Affect sensing modalities for in-the-loop adaptation (typical properties).

Modality	Latency	Temporal res.	Common artifacts	Best-fit tasks
EEG	0.1–1s	ms–10ms	motion, EMG	rapid, event-anchored reactions; round-based co-creation
Pupil / Eye	0.2–0.6s	ms-level	lighting, gaze shifts	arousal spikes; visual attention shifts
GSR (EDA)	1–5s	0.1–1s	temperature, movement	sustained arousal/engagement trends
HRV	10s–1min	beat-level	respiration, posture	workload/fatigue over longer windows
Facial/Voice	0.3–1s	frame-level	masking, context	explicit affect display, social cues

practically feasible but can also significantly enhance users' sense of immersion, satisfaction, and usage efficiency.

Unlike human-computer interaction, in human-AI co-creation, both human and machine take the initiative and influence the result. [57, 58]. Empirical studies show that users are most creative when treating the AI as a true co-creator rather than passively editing its outputs: McGuire et al. (2024) found that collaborative poetry writing with AI produced more original results, while participants who only edited AI-generated drafts were less creative than when writing alone [38]. In this collaboration, users must rapidly switch between different roles, both generating ideas and evaluating or providing feedback on ideas. This rapid change of role often relies on the emotional intuition of users: satisfaction typically means accepting current AI output and ending exploration, while dissatisfied calls for continued iteration in search of better results. However, most current AI collaboration systems primarily rely on explicit textual instruction and cannot perceive users' implicit emotional states, which easily leads to discontinuous collaboration and users feeling that the AI "doesn't understand them." Experiments show that when using AI brainstorming that differs from users' values, users often experience greater difficulty and frustration [35]. This gap highlights the necessity of introducing emotion perception mechanisms into human–AI co-creation—for example, capturing users' emotions in real time through physiological signals such as EEG, enabling AI to adaptively adjust strategies and ensure smoother role switching and more harmonious collaboration between humans and AI.

2.1.3 *Candidate Signals and the Case for EEG.* Human emotions can be inferred from multiple modalities, including speech prosody, facial expressions, body posture, heart rate variability (HRV), galvanic skin response (GSR), and electroencephalography (EEG) [3, 4]. Each of these modalities offers distinctive advantages but also has important limitations. Speech and facial expressions are intuitive and well supported by mature recognition techniques, yet they rely on explicit user behavior and are easily shaped by context or self-control, which makes them less reliable for subtle or suppressed affective states. Physiological measures such as HRV and GSR are more implicit and difficult to manipulate voluntarily, but their dynamics unfold over several seconds, which limits their ability to capture the rapid emotional fluctuations that often characterize iterative human–AI collaboration [22]. EEG, in contrast, provides millisecond-level temporal resolution and can reflect both affective arousal (e.g., excitement, stress) and higher-order cognitive states such as attention and workload [6]. Event-related potentials like the N170 emerge within a few hundred milliseconds of stimulus onset, and consumer-grade EEG headsets (e.g., Emotiv, Muse) now support real-time use outside laboratory settings [4]. Considering the trade-offs, EEG offers the most favorable balance of implicitness and timeliness, motivating its adoption as a foundation for neuroadaptive systems.

2.2 Neuroadaptive Systems and Their Applications

Neuroadaptive systems are interactive systems that adapt their behavior based on real-time neural or physiological signals, without requiring explicit user commands. The concept originates from passive brain–computer interfaces (pBCI) [59], in which spontaneous brain activity is leveraged to modulate system behavior. Such systems typically follow one of two approaches: rule- or threshold-based control, where predefined physiological thresholds trigger adaptations such as lowering task difficulty [2]; and reinforcement-based control, where neural signals act as implicit rewards to optimize adaptive policies through trial and error [5]. The first is more interpretable and stable, while the second provides flexibility but introduces greater opacity and noise sensitivity.

Over the past decade, neuroadaptive principles have been applied across diverse domains of HCI. In learning, adaptive tutoring systems dynamically adjust pacing or complexity when EEG signals indicate overload, yielding improved outcomes [13, 44]. In therapeutic contexts, meditation VR games harness EEG biofeedback to foster relaxation [34], and exposure therapies modulate the intensity of phobic stimuli based on EEG or related physiological signals [10, 55]. In training and assistance, flight simulators adjust difficulty using fNIRS workload measures [36], while collaborative robotics adapt to real-time EEG error-related potentials to optimize coordination [5]. These examples illustrate the breadth of neuroadaptive applications, yet they also underscore persistent challenges: commercial EEG systems update at relatively low frequencies (e.g.,

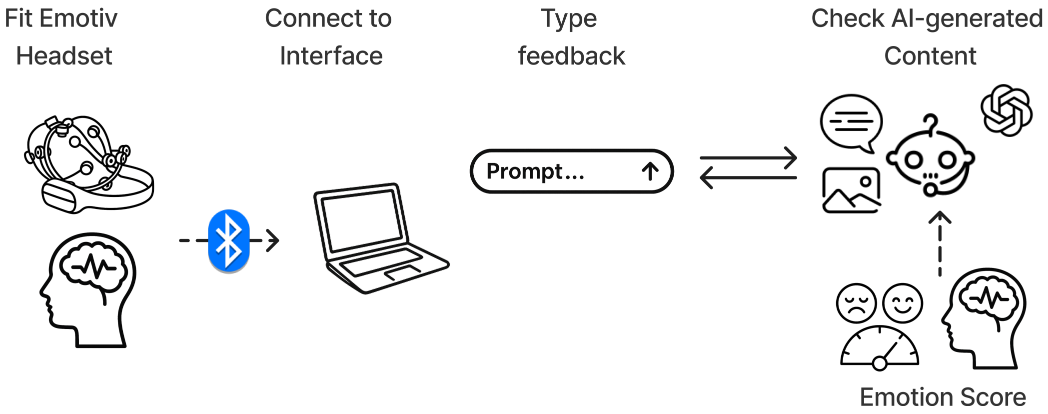


Figure 1: System design flowchart: EEG → Emotion inference → Strategy generation → User feedback

2 Hz), proprietary metrics often operate as black boxes [22], and long-term headset use raises issues of comfort, privacy, and monitoring [1]. Best practices in the literature recommend mitigating these challenges through event-anchored analyses that align neural windows with interface events, conservative adaptation thresholds, and rigorous control conditions to ensure reproducibility [7].

2.3 Toward Neuroadaptive Generative AI

Building on these foundations, recent research has begun to integrate neuroadaptive methods with generative AI. Some approaches inject neural signals directly into generative models: BrainDreamer aligns EEG with text and image embeddings and integrates them into diffusion models for controllable image synthesis [54], while Thought2Text couples an EEG encoder with an instruction-tuned LLM to generate natural language directly from brain activity [39]. Other studies treat physiology as feedback within adaptive policies. For instance, EEG has been used as an implicit reward for reinforcement learning in extended reality haptics [28], and recent frameworks fuse EEG, eye tracking, and generative models to create emotion-aware XR scenes [26]. In conversational contexts, EEG-driven tutoring systems such as NeuroChat dynamically adjust style, pacing, and complexity, thereby enhancing engagement and learner satisfaction [12]. Together, these advances demonstrate the feasibility of linking neural signals with adaptive generative systems. Yet current efforts remain fragmented, and systematic frameworks for neuroadaptive human–AI collaboration in creative tasks are still underdeveloped, leaving an important open space for future research.

3 System Design

3.1 Architecture Overview

We established five design objectives to make EmotivChat an accessible, low-latency, and reproducible neuroadaptive system capable of supporting both text and image workflows.

D1: Minimal Interface Deviation and Alignment with Mainstream LLM Interaction Paradigms To effectively evaluate EmotivChat and assess the impact of additional EEG channels on user experience, we adopted a design philosophy that closely mirrors mainstream chat model UI interactions, including ChatGPT, Claude, and Gemini, implementing only minimal modifications. This approach reduces user learning costs while minimizing the potential emotional impact of novel interface elements on user affective states.

D2: Event-Anchored Low-Latency Analysis In both Study A (Image Design Session) and Study B (Idea Generation), we anchored our analysis to specific UI events, employing brief post-stimulus temporal windows to capture users' immediate reactions to displayed content, rather than conducting continuous averaging over extended temporal windows for EEG analysis. This design choice stems from considerations regarding emotional signal sensitivity, as second-scale post-event windows optimally reflect users' emotional fluctuations.

D3: Prompt-Layer Neuroadaptation via Discrete Tag Mapping Through pilot studies, we determined that mapping continuous EEG emotional signals to discrete tags and subsequently utilizing these tags to govern prompt strategies yields superior AI behavioral stability and auditability compared to direct transmission of raw numerical values. The system employs these tags to select well-defined branching prompt-level strategies (e.g., fine-grained modification versus complete image regeneration; inspirational provision versus mild challenge prompts).

D4: Hide EEG Emotional Data from users Pilot studies revealed that users exhibited a tendency to modify their self-perception of emotional states after observing emotional scores generated by the EEG device. To prevent system-generated emotional metrics from influencing users' subjective emotional self-awareness, we employed only simple button indicators for successful connection throughout the experimental process, deliberately omitting actual user EEG emotional data from all UI interfaces.

D5: Control Conditions and Reproducible Logging To ensure experimental rigor and eliminate potential bias interference following user awareness of experimental distinctions, both control and experimental groups maintained identical UI, timing, sampling and logging protocols. The sole differentiation resided in the AI's operational mode: the control condition employed a singular default mode without the flexibility to select strategies based on EEG emotional signals. EmotivChat integrates real-time EEG with an

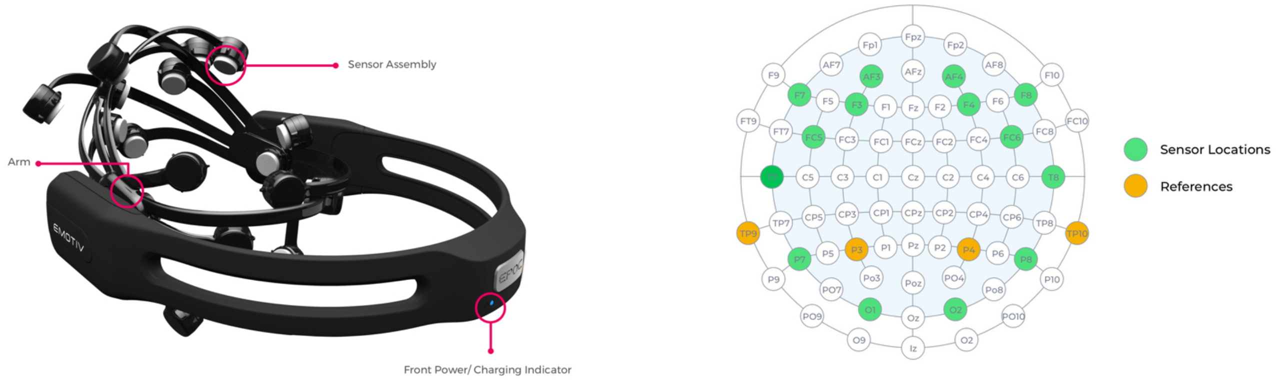


Figure 2: Left – the Emotiv EPOC$^+$ EEG system manufactured by Emotiv. Right – electrode layout of the device showing 14 active electrodes and 2 reference electrodes (P3, P4) positioned according to the 10–20 (extended 10–10) system. Illustration adapted from Emotiv official technical documentation(2021).

LLM assistant through a web UI and a Socket.IO back end, implementing a closed-loop neuroadaptive system. Figure 3 overviews the workflow for both creative tasks. The system operates on a fundamental principle: rather than directly injecting continuous EEG values into prompts, we convert physiological signals into discrete behavioral tags that gate predefined prompt-processing pipelines. This design ensures stable, auditable adaptation while maintaining real-time responsiveness.

3.2 EEG-to-Adaptation Pipeline

3.2.1 Signal Acquisition and Processing. Our system uses the Emotiv EPOC+ EEG headband. EPOC + has a fixed saline electrode cap layout with 14 electrodes placed according to 10–20 system locations covering frontal, temporal, parietal and occipital sites. The sensor positions are AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 [24]. The EPOC+ streams data wirelessly using a proprietary 2.4 GHz digital link (with a USB receiver dongle) by default.

Before each session we verify contact quality and seating of the mastoid references, rehydrating saline felts as needed and managing hair around the ear hooks per EMOTIV guidance. We conduct the study in a quiet meeting room with uniform lighting and no audio playback, and restart any session with sustained dropouts. These steps align with vendor recommendations on Contact Quality and EEG Quality.

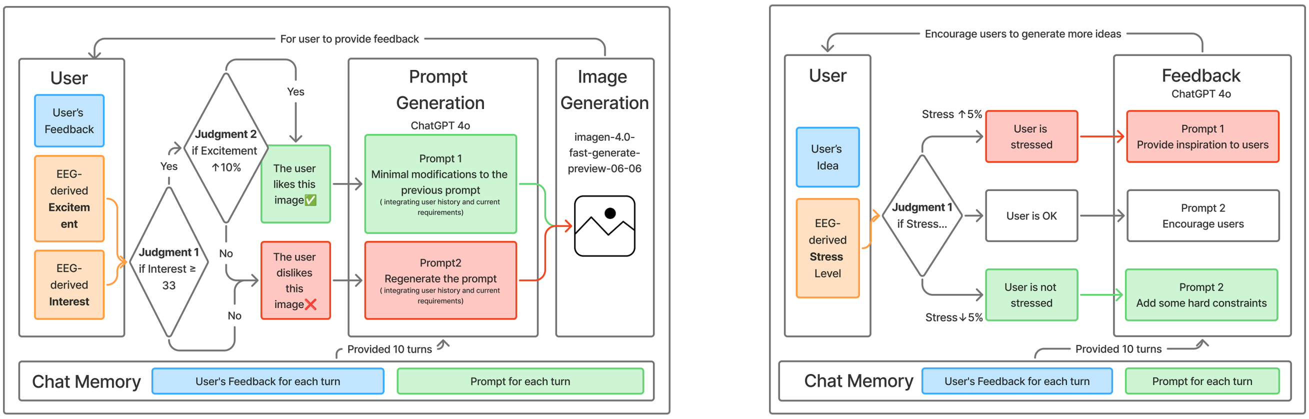
We treat EMOTIV's PM as black-box affective observables to reduce engineering overhead and latency for real-time human–agent loops. Consumer-grade EEG (including EPOC/EPOC+) has shown encouraging validity for ERPs and online interfaces, though PM algorithms remain proprietary and have been evaluated as such in prior reliability studies.

The participant wears the Emotiv EPOC$^+$ headset. Upon first connection, the server launches a single Performance Metrics (PM) streaming thread that continuously processes raw EEG signals into affective metrics at 2 Hz frequency. This dedicated thread appends preprocessed emotional state frames {timestamp, excitement, interest, stress, engagement, focus, relaxation} to a session log every 0.5 s. We verify electrode contact quality and stabilize headset fit before beginning experimental tasks to ensure reliable signal acquisition throughout the session.

3.2.2 Affective Metrics and Event Anchoring. For our neuroadaptive framework, we utilize Emotiv EPOC+ Performance Metrics (PM), which provide real-time affective state estimates derived from multichannel EEG signals. We selected this consumer-grade system because it offers validated emotional classification algorithms suitable for real-time interaction without requiring extensive signal processing expertise or laboratory-grade equipment setup.

The Emotiv system provides six affective metrics normalized to 0-1 scales: excitement (physiological arousal), interest (valence toward stimuli), stress (frustration with current challenge), engagement, attention, and relaxation. Based on validation studies and vendor documentation, we selected three primary metrics for our adaptation framework:

- **Excitement**: Measures physiological arousal and activation of the sympathetic nervous system. This metric responds to positive high-arousal states and captures immediate reactions to stimuli within seconds. We use excitement in Study A (image co-creation) because visual content typically elicits rapid arousal responses that indicate user engagement with generated images.
- **Interest**: Reflects valence (attraction vs. aversion) toward current stimuli or activities. Interest provides a complementary signal to excitement, helping distinguish between high-arousal positive states (excitement + high interest) and high-arousal negative states (excitement + low interest). This arbitration improves the reliability of our discrete tag classification.



Workflow of Study A: Image Co-creation

Workflow of Study B: Text Ideation

Figure 3: System workflow diagrams illustrating EEG-driven prompt adaptation in two creative tasks. Left: Image co-creation workflow where EEG-derived excitement and interest levels guide binary decisions between prompt modification (green path) and regeneration (red path). Right: Text ideation workflow where EEG-derived stress levels trigger adaptive feedback responses through three-branch prompt selection. Both systems maintain chat memory of user feedback and generated prompts across interaction turns.

- **Stress**: Indicates frustration or discomfort with current task demands. High stress often signals cognitive overload or dissatisfaction with system outputs. We employ stress as the primary adaptation signal in Study B (text ideation) because reading and processing AI-generated text suggestions can produce measurable stress responses when content misaligns with user expectations.

These three metrics were selected based on their relevance to creative collaboration scenarios and their reported classification accuracy. According to Emotiv's validation studies, excitement achieves approximately 80% classification accuracy across diverse user populations, while stress and interest show similar reliability for detecting significant state changes [23]. This accuracy level is sufficient for our discrete threshold-based adaptation approach.

We allow a brief relaxation period to stabilize electrode contact and signal quality, but deliberately avoid computing per-user min–max normalization. This design choice prevents potential bias from individual baseline differences and ensures that our adaptation logic relies on within-subject, within-session relative changes rather than absolute threshold values. All EEG-based adaptation uses short post-stimulus windows (design principle D2) to capture immediate emotional responses to system outputs rather than longer-term mood shifts.

We use post-stimulus windows anchored at rendering events to minimize motor-typing contamination. In the text study, the anchor is the moment an AI message is displayed on screen; we take the subsequent 10 s window of EEG Stress and compare its mean to the previous round. A relative change of **±5%** yields `FRU_TAG`. The tag is injected into the system prompt to select a nudge branch (adaptive), or computed but ignored (control). Sampling is 2 Hz with 0.5 s logging cadence to JSON lines.

In the image study, the anchor is the image display event. We take the mean values of EEG Excitement and Interest within the**0–5s** post-display window. First, the relative change of Excitement compared to the previous round is computed, and a **±10%** threshold is applied to classify`EXC_TAG`∈{`UP`,`FLAT`,`DOWN`}. Next, Interest values arbitrate the final tag: if Interest≥ 66, the tag is forced to UP; if Interest lies between 33–66, the original Excitement-based classification is preserved; if Interest ≤ 33, the tag is forced to DOWN. The final tag determines the prompt policy: the first two rounds always explore; thereafter, UP triggers a two-stage modification pipeline (intent analysis ⟶ precise replace or smart addition), while otherwise the image is regenerated. The control condition preserves the same processing and logging pipeline but ignores tags in strategy selection.

We do not apply subject-level min–max normalization of PM. Instead, we use within-subject, within-session relative change anchored to the previous round's post-stimulus window, which compensates for slow drifts and avoids overfitting to a single pre-task baseline. Practically, we allow a brief 2-minute relaxation period before tasks to stabilize the fit and contact quality, but the trace is not used for calibration. (This choice differs from studies that compute engagement indices over longer windows—e.g., 15 s for reading—because our tasks require rapid, round-based adaptation.)

3.2.3 Discrete Tag Generation. Given the vendor's PM update rate of 2 Hz and internal detection windows, end-to-end latency is dominated by the length of our post-stimulus windows (0 ˜5 s for images; 10 s for text). We therefore sample and persist PM at 2 Hz to match the upstream signal while avoiding redundant interpolation.

For Study A we expose `EXC_TAG` and the Interest band to the image-prompt policy; for Study B we expose `FRU_TAG` to the chat policy. In the control conditions we compute the same tags and logs but do not feed them into the policy, keeping the UI/timing identical across conditions.

Hi, welcome to Create! You can describe your design using style words + theme words, such as 'Victorian style + chair' or 'modern style + lamp'.

✨ Prompt:

Create a photorealistic portrayal of a Victorian-style chair, intricately carved with floral and scroll patterns, upholstered in rich emerald velvet. Bathed in soft, golden hour sunlight, the setting exudes elegance and nostalgia. Capture in a medium-close shot, with a sha depth of field to softly blur the ornate wallpaper in the background, evoking a mo timeless sophistication.

Figure 4: Example of the user interface for Study A (Image Co-Creation) experimental research. It demonstrates the workflow from initial prompt input, through a mandatory 5-second full-screen image display for EEG data collection, back to the chat interface for iterative refinement. The interface for Study B (Text Ideation) is identical to Study A and is therefore not shown separately.

Using vendor PM avoids heavy signal processing, but it also inherits two limitations: (i) the mapping from raw EEG to PM is proprietary (black-box), and (ii) PM latency is constrained by the detector's internal window and 2 Hz update rate. We mitigate these issues by using relative changes in short post-stimulus windows and by reporting full logs (PM frames, event anchors, window boundaries, tags) to support reproducibility and sensitivity analyses.

3.3 Prompt Adaptation Strategies

3.3.1 Rule-Based Policy Gating. Unlike prior neuroadaptive systems that embed continuous EEG values directly into the prompt and expect the model to infer adaptation rules, our design follows a rule-based gating approach. EEG signals are first converted into discrete tags through hard-coded thresholds on post-stimulus windows (±5% for stress, ±10% for excitement). These tags do not directly alter the wording style of the model; instead, they determine which of several predefined prompt-processing pipelines is executed.

This design guarantees that adaptation remains stable, auditable, and reproducible. The LLM itself only sees the branch-specific instructions; it is never directly exposed to raw EEG values or asked to reason about "engagement" as a concept. In effect, our adaptation logic is implemented as hard-coded policy gating, with the model serving as a generation module under externally imposed rules.

3.3.2 Task-Specific Implementations. In the text ideation study, the system always begins each turn by summarizing all accumulated user ideas into a numbered list (hard-coded rule). After this list, the tag `FRU_TAG` selects one of three fixed nudge templates: HIGH → empathy + micro-scaffold; LOW → lightweight challenge; NORMAL → short encouragement. The control condition computes the same tag but ignores it, always appending the same baseline encouragement.

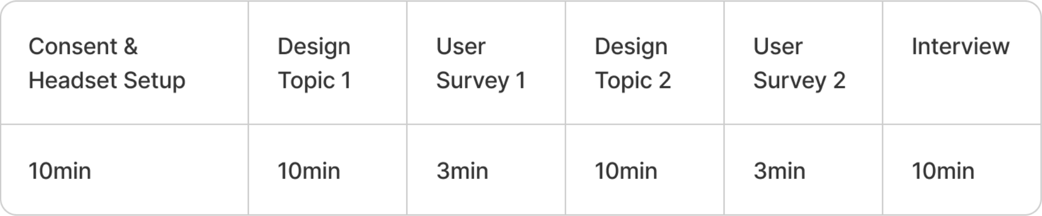
In the image co-creation study, the tag `EXC_TAG` decides whether the system will refine or regenerate prompts. The first two rounds always explore. Thereafter, if `EXC_TAG`=UP (possibly re-written by Interest arbitration), the pipeline calls a two-stage modification function: (i) analyze user intent and locate target spans, (ii) apply precise replacement or a smart addition. If `EXC_TAG`=DOWN/FLAT, the system bypasses modification and instead regenerates a new prompt via the simple optimizer. All of these branches are enforced by explicit if–else structures in the backend code rather than inferred implicitly by the LLM.

3.4 Implementation Details

3.4.1 Technical Infrastructure. For text, we use an LLM (Chat Completions API) to render summaries and nudges with the tag as a system hint. For images, we use Google Imagen (imagen-4.0-fast-generate-preview-06-06) for fast, consistent quality; prompt modification is LLM-assisted.

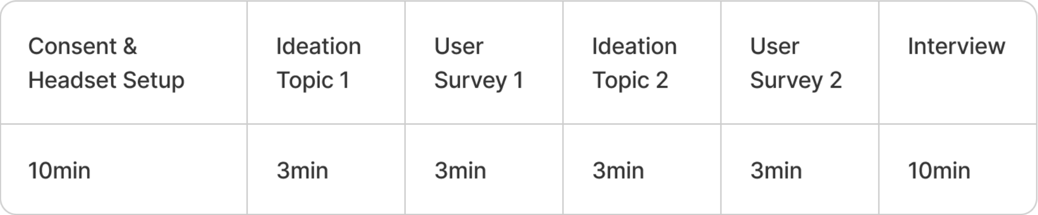
To ensure experimental rigor, both EEG-adaptive and control conditions maintain identical interface timing, logging protocols, and user interactions. The only difference lies in whether EEG-derived tags influence prompt selection policies. This parity design allows us to attribute any observed differences directly to neuroadaptive prompt strategies rather than system variations. Additionally, the framework supports manual tag override modes for ablation studies and technical validation, ensuring robust experimental methodology across both creative tasks.

3.4.2 User Interface Design. The experimental interface was implemented as a lightweight web application. Although both studies share a common streaming back-end, their user interfaces differed



Consent & Headset Setup	Design Topic 1	User Survey 1	Design Topic 2	User Survey 2	Interview
10min	10min	3min	10min	3min	10min

Study A: Image co-creation



Consent & Headset Setup	Ideation Topic 1	User Survey 1	Ideation Topic 2	User Survey 2	Interview
10min	3min	3min	3min	3min	10min

Study B:Text ideation

Figure 5: Experimental procedure timeline for both studies. Each study followed a within-subjects design with counterbalanced conditions. Study A (Image co-creation) consisted of two 10-minute design sessions separated by brief surveys, while Study B (Text ideation) used two 3-minute brainstorming sessions. Both studies included initial consent and EEG headset setup, followed by post-session interviews.

according to task requirements. As shown in Figure 4, the interface consists of several key components:

(1) **Conversation History** This panel displays the user's conversation history, allowing participants to review previous dialogue content and return to earlier conversations. This feature facilitates precise discussion of specific AI interactions during post-session interviews.
(2) **Welcome Message** Since each participant completes two similar sessions with different topics, the AI provides a welcome message before starting each session. This message briefly outlines the main task and includes an initial prompt example.
(3) **AI Response Display** The AI response format differs between studies: In Study A (Image Co-creation), generated images were displayed full-screen for exactly 5 seconds to ensure clean EEG signal capture while participants remained still and immersed; the system then returned to chat mode, showing the optimized prompt with the generated image embedded in the conversation history. In Study B (Text Ideation), by contrast, the system collected free-text ideas and returned structured responses in a fixed format, first listing cumulative ideas followed by contextual encouragement.
(4) **Chat Window** This is the main interaction area where users communicate with the AI system. The interface includes a text input field and send button for user input.
(5) **Timer and Session Controls** In Study A, participants completed a visible countdown (5–10 minutes) with enforced 5-second full-screen image displays, whereas Study B used a shorter 3-minute countdown with automatic session termination.
(6) **Post-Task Rating Panel** After each session, participants were redirected to an evaluation interface: in Study A, they selected their preferred generated image, whereas in Study B, they rated each idea on four dimensions (originality, feasibility, interest, relevance) using a radar style widget, with the ratings stored for subsequent analysis.

Together, these four components ensured that participants could complete tasks with minimal friction, while the experimenter could monitor session timing and EEG logging in real time.

4 Method

We evaluate two interaction modes under matched UIs and timing. The only difference between conditions is whether EEG tags influence prompt policy.

4.1 Hypotheses

We evaluate two interaction modes under matched UIs and timing. The only difference between conditions is whether EEG tags influence prompt policy.

4.1.1 Study A Image Co-Creation Related Hypotheses.

- **(H1) Objective Excitement - Image task**: In the image co-creation task, the experimental group using EEG-adaptive support will exhibit higher average excitement levels during the ideation stage compared to the control group, as measured by Emotiv devices.
- **(H2) Subjective Experience - Image Task**: In the image co-creation task, compared to the control group, the experimental group will show higher average ratings in subjective excitement, satisfaction, and other experience dimensions, as measured by questionnaire.

4.1.2 *Study B Text Ideation Related Hypotheses.*

- **(H3) Objective Stress - Text Task**: In the text ideation task, monitoring with Emotiv devices will reveal that the experimental group exhibits more stable stress levels (lower variability) during the ideation phase compared to the control group.
- **(H4) Subjective Experience - Text Task**: In the text ideation task, compared to the control group, the experimental group will show more positive evaluations in subjective satisfaction, stress perception, and other experience dimensions, as measured by questionnaire.

4.2 Participants

We recruited volunteers from the Delft University of Technology campus through mailing lists, poster advertisements, and word of mouth. To ensure that they could safely wear the EEG headset and complete both design tasks, participants had to (1) be between 18 and 45 years of age, (2) report normal or corrected normal vision, (3) have no history of neurological disorders or scalp injuries, and (4) be familiar with basic computer-aided sketching tools.

Following the guidelines of Faul et al [25], we performed an a priori power analysis with G*Power 3.1, adopting the medium paired-sample effect size cited from Cohen ($d_z = 0.50$). Under a two-tailed test with a significance level of $\alpha = 0.05$ and a target power of $1 - \beta = 0.80$, the analysis indicated that a minimum of 34 participants would be required. To account for potential attrition and unusable EEG data, we initially recruited 38 volunteers.

Most of the participants were master or doctoral students in design, architecture, or computer science. All gave their written informed consent; The study was approved by the Human Research Ethics Committee of the TU Delft (application no. 5380, 21 May 2025). Each participant completed two 20-minute brainstorming sessions, one with adaptive EEG support and one without, on a counterbalanced AB / ABA crossover schedule (Latin square order). The sessions were separated by a 5-minute neutral video break. The entire visit lasted about 70 minutes, for which the participants received a €3 coffee coupon.

4.3 Study Design and Protocol

Study A (Image Co-Creation). Participants worked on two classic design briefs, a chair and a lamp, that support open-ended exploration while providing a shared reference point. In each round, the assistant generated an image based on the current prompt and the user's textual feedback. In the adaptive condition, we compute the mean Excitement and Interest values in the window 0–5 s following image display. Relative changes in ±10% produced an EXC_TAG (UP/FLAT/DOWN), with Interest serving as an arbitration channel (LOW⟶DOWN; HIGH⟶UP). Policy mapping followed a staged design: the early rounds emphasized exploration, while the later rounds used tags to decide between two strategies. When UP was detected, the assistant performed a two-stage modification pipeline - intent analysis with precise replacement followed by smart addition - while in all other cases the assistant regenerated the image. The control condition used the identical generation stack and logging infrastructure, but always applied the same fixed prompt-optimization routine regardless of EEG input. All emotion frames (2 Hz), display events, window analyzes, prompts, and generated images were persistently logged in the session directory.

Study B (Text Ideation). The second study examined the generation of text-based ideas in remote work settings. Participants responded to two prompts: sustaining focus after 45 minutes of working from home, and unwinding during a 15-minute end-of-day transition. Ideas were entered in a chat interface, where the assistant first summarized and refined the accumulated list and then offered a nudge. In the adaptive condition, we computed the 10 s post-display Stress mean after each AI response. Relative changes of ±5% yielded a FRU_TAG (HIGH/LOW/NORMAL), which was injected into the system prompt to select a predefined nudge branch. The control condition preserved the identical UI, timing, sampling rate, and logging pipeline, but the computed tag was not used and the policy remained fixed. Logging again captured the full set of emotion metrics at 2 Hz (attention, engagement, excitement, stress, relaxation, interest), as well as window analyses and complete chat transcripts with timestamps, ensuring reproducibility.

4.4 Evaluations

4.4.1 Objective EEG. Study A uses post-display 0–5 s Excitement means per round; Study B uses post-display 10 s Stress means per round, with Interest recorded in the same window. **Subjective experience.** After each condition, participants rated excitement/engagement (image study) or perceived tension/relaxation and usefulness of nudges (text study) on 7-point Likert items; free-text comments captured qualitative impressions.

4.4.2 Questionnaires. To evaluate the demographic background, subjective experience and qualitative impressions of the participants, we designed a set of questionnaires consisting of three parts.

- **Demographics & Background Survey**: age of the participants collected, academic / professional field, previous experience with generative AI, familiarity with EEG, hours of sleep before the experiment, and recent caffeine intake. This was used as covariates for subsequent analyses.
- **Post Text-Session User Survey**: a six-item Likert questionnaire administered after each comfort conversation. It assessed relaxation, stress reduction, communication comfort, ease of understanding, perceived empathy, and willingness to continue interaction.
- **Post Image-Session User Survey**: a six-item Likert questionnaire administered after each image co-creation session. Measured excitement, absorption, satisfaction with the results, iteration efficiency, AI understanding of the intent of the user, and reuse intention.
- **Interviews**: two open-ended questions were asked at the end of the text session, and two at the end of the image session. These questions focused on moments of felt empathy, suggestions for improvement, peak inspiration events, and reflections on emotion-adaptive support.

Full questionnaires (English version) are provided in the Appendix (Section C) for reproducibility.

Table 2: OLS regression results examining the effects of study condition, topic, and order on user ratings and EEG metrics in Study A (N=31). The EEG condition significantly improved user satisfaction ratings, while order effects dominated EEG excitement measures and topic influenced interest levels.

Outcome variable	EEG vs Normal	Topic	Order	R^2	Outcome variable	EEG vs Normal	Topic	Order	R^2
User_rate_score	$\beta = +0.50^{*}$ [0.10, 0.91]	n.s.	n.s.	0.012	E_peak (peak)	n.s.	n.s.	$\beta = +8.61^{***}$ [5.47, 11.75]	0.059
E_level (mean)	n.s.	n.s.	$\beta = +6.30^{***}$ [3.78, 8.83]	0.049	I_level (mean)	n.s.	$\beta = -2.28^{**}$ [-3.69, -0.87]	n.s.	0.027

Note: Entries are coefficient estimates (β) with 95% CIs in brackets. n.s. = not significant. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

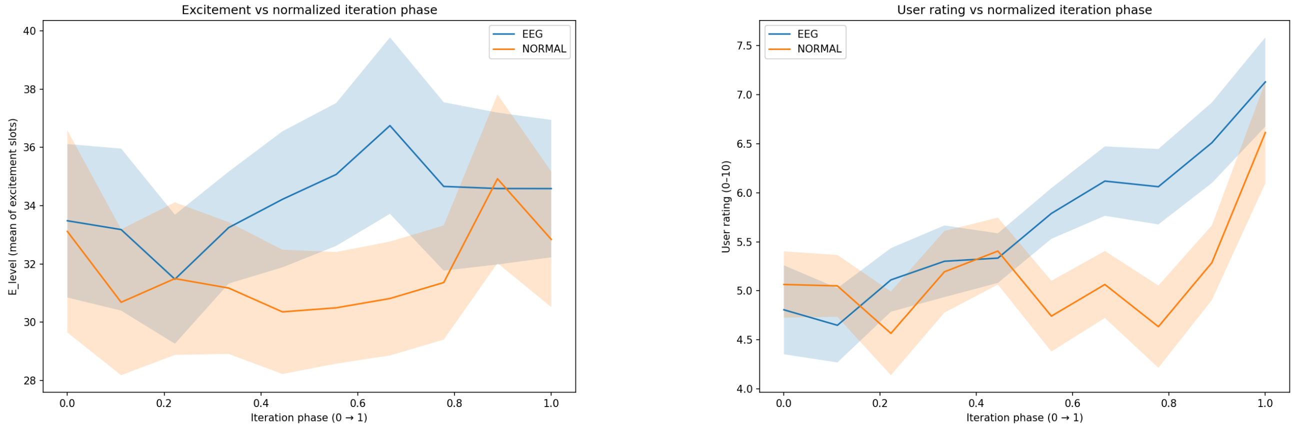


Figure 6: Group-level trajectories across normalized iteration phases. Left: mean Excitement (E_level). Right: mean user ratings. Shaded areas represent standard errors.

5 Results

5.1 Study A: Image Co-creation

5.1.1 EEG Analysis. A total of 38 participants completed the study. During data preprocessing, we first selected EEG signals within the first 5 seconds after image presentation. This time window was chosen because it is most strongly associated with participants' immediate reactions, whereas signals from other periods were considered less relevant and were excluded.

We then performed data cleaning. Specifically, we removed segments with values of 0 and those with two consecutive identical values. The rationale is that the Emotiv device repeats the last value when the signal is disconnected, and given the two-decimal precision of the output, such repetitions are unlikely to occur under normal conditions. Since Emotiv records at 2 Hz, each 5-second window should contain 9–10 valid data points. Participants with missing or duplicated values across multiple slots were excluded. After this procedure, data from 31 participants remained. As Emotiv normalizes all affective outputs in the range 0 to 100, we directly used these values to calculate emotional indices: mean Excitement (E_level), peak Excitement (E_peak), and mean Interest (I_level). In addition, participants' subjective ratings of image satisfaction (0–10 scale) were included in the analysis together with EEG signals.

Macro Analysis. At the macro level, we fitted ordinary least squares (OLS) regression models to test the effects of EEG condition, topic, and order on user ratings and emotional indices. The results showed that the EEG condition significantly increased user ratings ($\beta = +0.50$, 95% CI [0.10, 0.91], $p = 0.016$), suggesting that EEG feedback may improve subjective experience. In contrast, changes in Excitement were mainly driven by order: the second round elicited significantly higher values than the first, while EEG condition and topic were not significant. The interest was primarily affected by the topic, with lower values in the condition of the lamp compared to the condition of the chair. In general, the EEG condition had the most robust effect on user ratings, while order and topic were the main factors influencing Excitement and Interest, respectively.

Micro Analysis. Given the limited macro-level effects, we further examined the dynamics of user responses across iterative cycles, exploring how ratings and EEG-based affective responses evolved over time. Because participants varied in the number of iterations, we applied linear interpolation and resampling, normalizing each participant's sequence to the range [0,1] and dividing it into 10 phases. Group means and standard errors were then calculated for the EEG and control conditions.

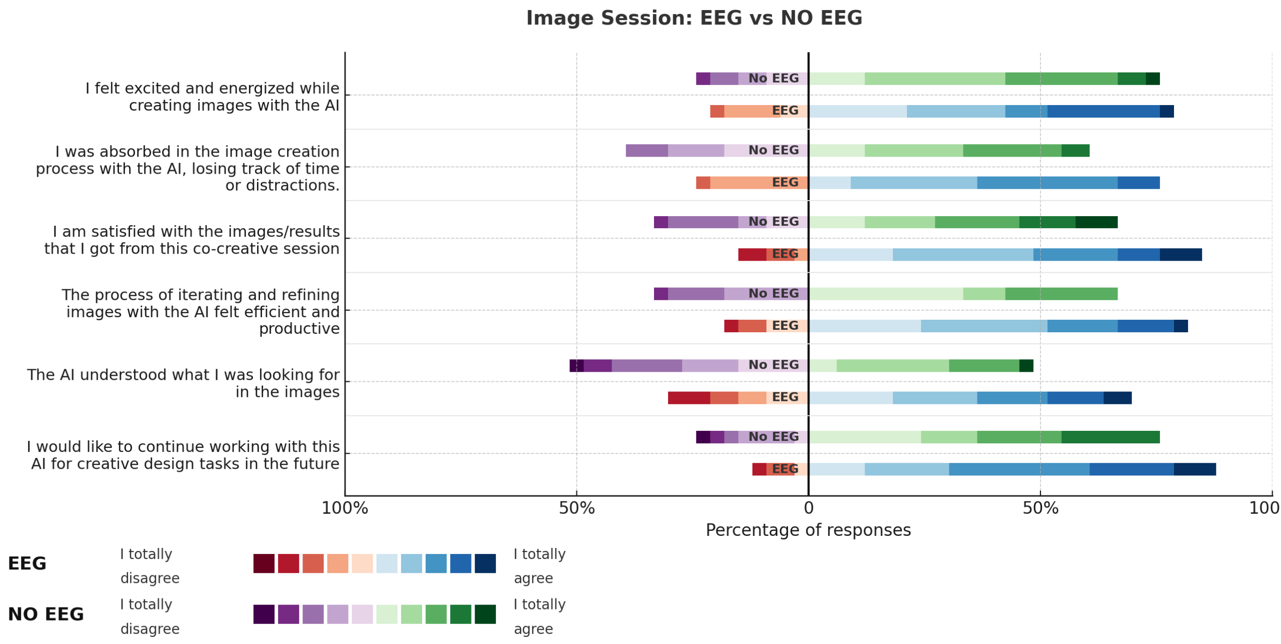


Figure 7: Questionnaire results for Study A (Image co-creation) comparing EEG-adaptive and non-adaptive conditions. The diverging bar chart shows response distributions for six Likert-scale items (N=31). Purple/red bars represent the EEG condition while green/blue bars represent the No EEG control condition, with darker colors indicating stronger agreement. The EEG-adaptive system consistently received more positive ratings across all dimensions, with particularly notable improvements in perceived AI understanding, iteration efficiency, and future usage intention. Response percentages are displayed as horizontal stacked bars extending from a central axis.

The analysis revealed that for Excitement, both groups showed an initial decline, possibly reflecting reduced arousal after initial engagement. However, in later phases, the EEG group diverged from the control group, showing faster increases and earlier peaks, indicating more sustained arousal under EEG feedback. The trajectory of the scores closely matched that of Excitement: little difference in the early phases, but faster and more stable increases in the EEG group during later phases, further supporting the positive impact of the EEG feedback on subjective experience. In contrast, interest remained relatively flat around 50 in both groups, reflecting a neutral state without strong preference or dislike. This may be because we only captured EEG signals within 5 seconds after image presentation, a window in which Interest shows limited dynamics. In this context, Excitement is a more sensitive indicator of user preference and engagement than Interest.

5.1.2 Questionnaire results. Figure 7 illustrates the distribution of responses for each questionnaire item in **Study A** under the two experimental conditions (EEG vs. Non-EEG). As shown, the purple and orange bars represent the proportion of participants who selected *"Totally Disagree"* in the Meaningful and Non-meaningful conditions, respectively, while the blue and green bars correspond to the proportion of participants who selected *"Totally Agree"* in the two conditions. This visualization allows a direct comparison of how the agreement levels shifted between conditions for each item.

Figure 7 presents the response distributions for the image co-creation session, comparing the *EEG-adaptive* condition with the *No EEG* control. Across all six survey items, participants consistently rated the EEG-adaptive condition more positively.

For item *"I would like to continue working with this AI for creative design tasks in the future"*, the EEG-adaptive condition received substantially higher agreement compared to the control. Similarly, the participants reported that *"The AI understood what I was looking for in the images"* more strongly in the EEG-adaptive condition. The same trend was observed for *"The process of iterating and refining images with AI felt efficient and productive"*, where the adaptive condition was rated as more effective.

In addition, participants expressed greater satisfaction with the results of the co-creative session (*"I am satisfied with the images / results that I obtained from this co-creative session"*) under the EEG-adaptive condition. They also reported being more deeply absorbed in the creative process (*"I was absorbed in the image creation process with AI, losing track of time or distractions"*). Finally, for the item *"I felt excited and energized while creating images with the AI"*, the EEG-adaptive condition again outperformed the control, though the difference was comparatively smaller.

Overall, the results suggest that the incorporation of EEG signals into the image co-creation workflow led to higher levels of user satisfaction, perceived alignment, and engagement across all dimensions surveyed.

Table 3: Qualitative themes from post-session interviews comparing EEG-adaptive and non-adaptive *text ideation* experiences. EEG offered more sparks and divergence but sometimes felt opaque; Normal was clearer and calmer but less stimulating.

Theme	EEG (Experimental group)	Normal (Control group)
Personalization & Stimulation	Tailored cues; frames; sparked ideas	Generic encouragement; fewer sparks
Clarity & Relevance	Mixed clarity; some abrupt/opaque prompts	Clearer but generic; supportive
Depth & Divergence	Promoted divergence; added new angles	Incremental refinement; risk superficiality
User Experience	Engaging but added stress/flow resets	Calmer, lower friction; flatter engagement
Satisfaction & Preference	More novelty/breakthroughs; stimulus-seeking users preferred	Safer, steadier; clarity-seeking users preferred
Trade-offs	Benefits shaped by task fit, timing, expectations	Calmer but less discovery; same moderators

5.1.3 *Qualitative Analysis (Condition–Centered).*

We analyzed post-session interview transcripts from the image study using a condition–centered approach. For each participant (S#), we inserted a condition header (e.g., `[Condition] EEG_Chair; Normal_Lamp`) and, within the narrative, automatically labeled every mention of task objects with the corresponding tag (“Chair (EEG)”, “Lamp (Normal)”). The *same participant*’s experiences under the two conditions—EEG (assisted) vs. Normal (non-assisted)—served as the primary unit of comparison; object-level differences (chair vs. lamp) were intentionally backgrounded.

We followed a three-step thematic procedure. First, *open coding*: we tagged condition-relevant fragments with short, experience-oriented labels grounded in participants’ language (e.g., *continuity/jumps*, *remembers/aligns*, *inspiring/satisfying*, *iteration cost/redo*, *odd outputs*). Second, *axial coding*: we aggregated labels per condition (EEG vs. Normal) and contrasted the two along shared dimensions. Third, *selective coding*: we consolidated high-frequency, cross-participant themes (while noting counterexamples) into a small set of reportable findings. All coding was performed on a single labeled master document; references to S# point to those passages.

T1: Continuity vs. Jumps. Across participants, EEG runs were often described as smoother and easier to carry forward, whereas Normal runs were more likely to “jump” or “drift”. Participants repeatedly framed EEG as “picking up where the last one left off”, while the Normal condition “changed too fast” from one image to the next. For example, S01 remarked that in the Normal session images “got close to my ideal and then ran away… not continuous”, whereas the EEG session “made me want to keep going” (translated). S06 similarly contrasted a “more jumping” Normal run with an EEG run that “still felt coherent and aware of what I’d said before” (translated). S32 noted that brief prompts tended to be respected and incrementally applied under EEG, while equivalent brief prompts in Normal would “jump off” the track (translated). These accounts characterize EEG as building upon prior state, and Normal as more reset-like between steps.

Still, discontinuity was not universally negative. A minority valued Normal’s unpredictability for novel directions: S27 described abandoning a shape tweak mid-way, yet finding an unexpected base style “I actually liked” and decided to stop there; S25 also reported that while Normal jumped and occasionally produced “weird” frames, the final lamp variation was the one they preferred (translated). These counterexamples position Normal’s jumps as sometimes creatively productive, albeit at the risk of derailment.

Mini-synthesis. Overall, more participants associated EEG with continuity and Normal with jumps; a smaller group explicitly appreciated Normal’s stylistic leaps for serendipity. Task and taste moderated these preferences.

T2: Remembering/Alignment and Sense of Control. A second, closely related theme concerns memory for prior constraints. Many participants reported that EEG runs better “remembered” what had been established, reducing the need to restate constraints at every step. S10, for instance, found that EEG eventually “remembered both ergonomics and warmth” simultaneously, whereas in the Normal lamp session “I had to say *everything* again each time—otherwise it forgot and couldn’t add on top” (translated). S16 similarly described the need to “keep reminding” the Normal run to retain earlier elements, while the EEG run converged into a more unified behavior later. S33 observed a concrete “memory forgot” moment in the Normal chair sequence after a strong intermediate design, forcing a return to earlier prompts.

Memory was not perfect in EEG either. S14 noted that switching languages mid-session (English/Chinese) disrupted reference continuity; S12 found that even small tweaks (e.g., just changing color) sometimes “reset” the Normal run into a different style, making it hard to revert; S09 reported that both conditions could drift late in the sequence, though EEG was more likely to “come back” to the desired style after reminders. Together these accounts frame EEG as *more likely* to preserve constraints across turns, with Normal more prone to forgetting.

Mini-synthesis. Participants often linked EEG’s better remembering with a stronger sense of control; Normal’s forgetfulness diminished this sense and increased overhead. Still, memory was situational and not guaranteed in either condition.

T3: Inspiration/Satisfaction and Carry-On Potential. Participants frequently tied satisfaction to whether a run delivered a “good place to continue." EEG sessions were said to produce those carry-on moments more consistently. S01 reported feeling "happy" and eager to proceed under EEG; S09 found the EEG lamp had a "good starting point," needed fewer turns, and "kept the style" early on; S15 characterized EEG chair as "smoother" overall (translated). These moments of alignment appeared to reduce friction and support momentum.

At the same time, Normal could be inspiring in a different way—by throwing out a surprising variant. S27 described an unforeseen base style that was "interesting enough to stop on" even though it diverged from the requested tweak. S35 similarly accepted a non-conforming lamp direction because it was "impressive" as a design object despite departing from the intended specification (translated). Such cases suggest that Normal's exploration can surface aesthetically compelling but less controlled outcomes.

Mini-synthesis. EEG more often produced incremental, momentum-preserving progress; Normal more often produced jumps that, while disruptive, sometimes yielded appealing side paths. Participants' satisfaction tracked these two routes to inspiration differently.

T4: Iteration Cost and Rework. Relative to Normal, EEG sessions were described as involving fewer "fix–drift–refix" loops and less re-statement overhead. S09 contrasted the difficulty of steering a Normal chair ("very hard to get the output I wanted") with the EEG lamp's shorter path and fewer iterations; S38 said EEG tended to "follow the prompt more tightly," whereas the Normal lamp was "stuck fixing one issue and never quite finishing—back and forth again and again" (translated). Participants also reported that EEG missteps felt more "recoverable"—it could be nudged back to an earlier, better state—while Normal more often required restating the whole set of constraints.

Still, low iteration cost depended on prompt clarity and target specificity. S28 explicitly contrasted a vivid mental image (EEG chair) that led to steady, compliant changes versus a vague target (Normal lamp) that scattered into unhelpful directions; S20 noted that material preferences (e.g., wood vs. metal) sometimes "stuck" and had to be re-asserted. These cases indicate that the system's path length is co-determined by the user's articulation strategy.

Mini-synthesis. Participants generally perceived iteration to be cheaper under EEG and costlier under Normal; however, costs rose under either condition when prompts were underspecified or when edits seemed to reset style/memory.

T5: Failure Modes and Odd Outputs. Both conditions produced occasional "odd" or irrelevant outputs—e.g., S03 mentioned unexpected "lions" or structurally implausible chairs; S33's EEG lamp began as "a space station"; S18's EEG chair took on a whimsical, creature-like form mid-way (translated). Participants consistently reported such oddities as *more* common in Normal sessions, with EEG's odd outputs seen as less frequent and more readily steered back. S21, S24, and S27 each encountered Normal jumps to off-topic or malformed variants; several noted that returning to a previously satisfactory state was harder under Normal than under EEG.

Mini-synthesis. Odd outputs occurred in both conditions; participants more often associated them with Normal and described EEG oddities as more recoverable.

Summary and Boundaries. Across themes, participants tended to perceive EEG runs as more continuous, better at remembering prior directions, and less costly to iterate, while Normal runs were more prone to abrupt jumps, forgetfulness, and longer rework. However, boundary conditions mattered. First, *prompt clarity*: when participants had a crisp mental image (S28), either condition worked better; when targets were vague, Normal's exploration sometimes helped discovery (S27, S35) but also wasted turns. Second, *task structure*: some reported chairs as harder to steer (S09), others found lamps more volatile (S38), hinting at object–model priors. Third, *language switches*: mixing English/Chinese mid-stream occasionally disrupted continuity (S14). These nuances qualify the trend and explain individual differences in satisfaction.

5.2 Study B: Text Ideation

5.2.1 User EEG Analysis. We visualized the evolution of stress throughout 180 s (sampled at 2 Hz; 0.5 s per point) for the conditions EEG and NORMAL. For each participant and condition, we plotted the raw series of time points as thin semitransparent lines to show individual variability, and overlaid a group trend using a 7-point moving average (approximately 3.5 s) to highlight the overall trajectory. This visualization aggregates all available trials while remaining robust to missing values (NaNs were ignored in the moving average computation).

To assess the variability within participants in stress responses, we computed three dispersion metrics from each participant's 3-minute stress time series (`stress_001–stress_360`): (1) the standard deviation (SD), (2) the median absolute deviation (MAD), and (3) the coefficient of variation (CV = SD / mean). Because each participant completed both the EEG and the NORMAL conditions, we conducted paired-samples t-tests on the metric values between conditions. This approach directly tests whether the same individuals exhibited systematically greater or smaller variability across conditions.

Table 4 summarizes the results of the paired test t comparing the variability metrics between conditions EEG and NORMAL. On average, the EEG trials produced slightly higher variability (positive mean differences), but none of these differences reached statistical significance.

Table 4: Paired t-test results for within-participant stress variability (EEG − NORMAL). Positive mean differences indicate larger variability in the EEG condition.

Metric	n	Mean Diff	t	df	p
SD	29	+1.00	1.01	28	.321
MAD	29	+0.16	0.20	28	.844
CV	29	+0.026	1.35	28	.189

In general, no significant differences were observed ($p > .05$), suggesting that the variability of stress in the 3 minute sessions was comparable between conditions EEG and NORMAL.

5.2.2 Questionnaire results. Figure 9 compares EEG-adaptive and EEG-no EEG conditions in the text-based brainstorming session. In

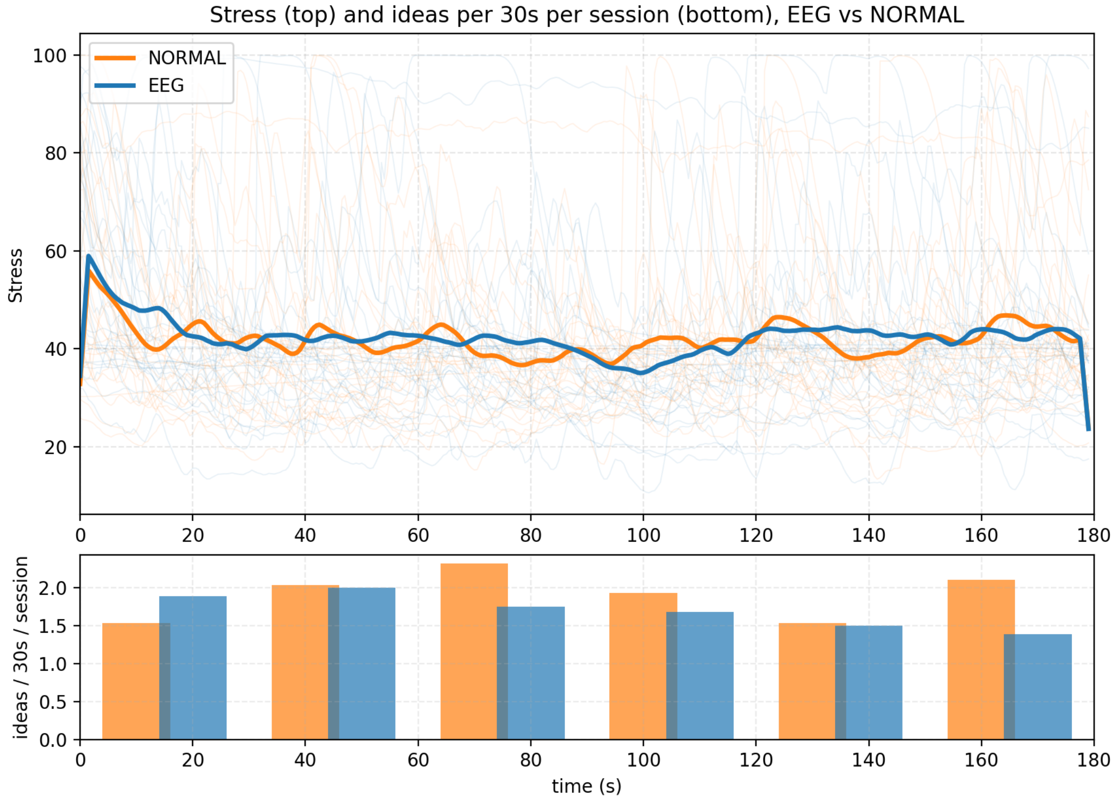


Figure 8: Stress trajectories during the first 3 minutes. Thin translucent lines depict individual participants; thick lines show group trends computed via a 7-point moving average (approx. 3.5 s). Time is sampled at 2 Hz (0.5 s resolution). Orange denotes EEG, blue denotes Normal.

general, the participants reported more positive experiences when the system adapted the commands based on EEG signals.

For the item *"I would like to use this AI brainstorming assistant again in the future"*, the participants expressed a greater willingness to reuse the adaptive EEG system. A similar trend was observed for *"I felt the AI and I worked well together as a team to generate ideas"*, where the scores in the EEG condition were consistently higher.

Regarding *"The AI's prompts were clear and easy to understand"*, both conditions received generally favorable responses. However, the EEG condition showed a more polarized distribution, with relatively higher proportions of strong agreement and strong disagreement. This pattern may reflect individual variability in stress-triggered adaptation, where some participants benefited more directly from adaptive prompting than others.

For *"I was comfortable with the way AI prompts and suggests ideas"*, the EEG-adaptive system again outperformed the control condition. A plausible explanation is that stress-triggered adaptation occasionally led to more concrete and actionable suggestions, whereas the EEG non-convention mainly provided generic or comfort-oriented suggestions.

Finally, participants indicated that *"I was able to generate a large number of ideas during this session"* more strongly under adaptive conditions of EEG than in control. Taken together, these results suggest that EEG-driven adaptation improved both the perceived usefulness and productivity of the text-based brainstorming process, although with some variation between individuals.

5.2.3 *Perceived Stress and Subjective Evaluations (Condition–Centered).*

We analyzed post-task interviews comparing EEG (assisted) and Normal (non-assisted) idea-generation sessions, focusing on five aspects participants raised most often: perceived personalization and idea stimulation, clarity and relevance of prompts, depth and divergence of ideas, emotional reactions under time pressure, and overall satisfaction/preferences. In brief, participants tended to describe EEG interventions as more stimulating yet sometimes abrupt or opaque, whereas Normal prompts felt steadier and easier to parse but less likely to spark new directions.

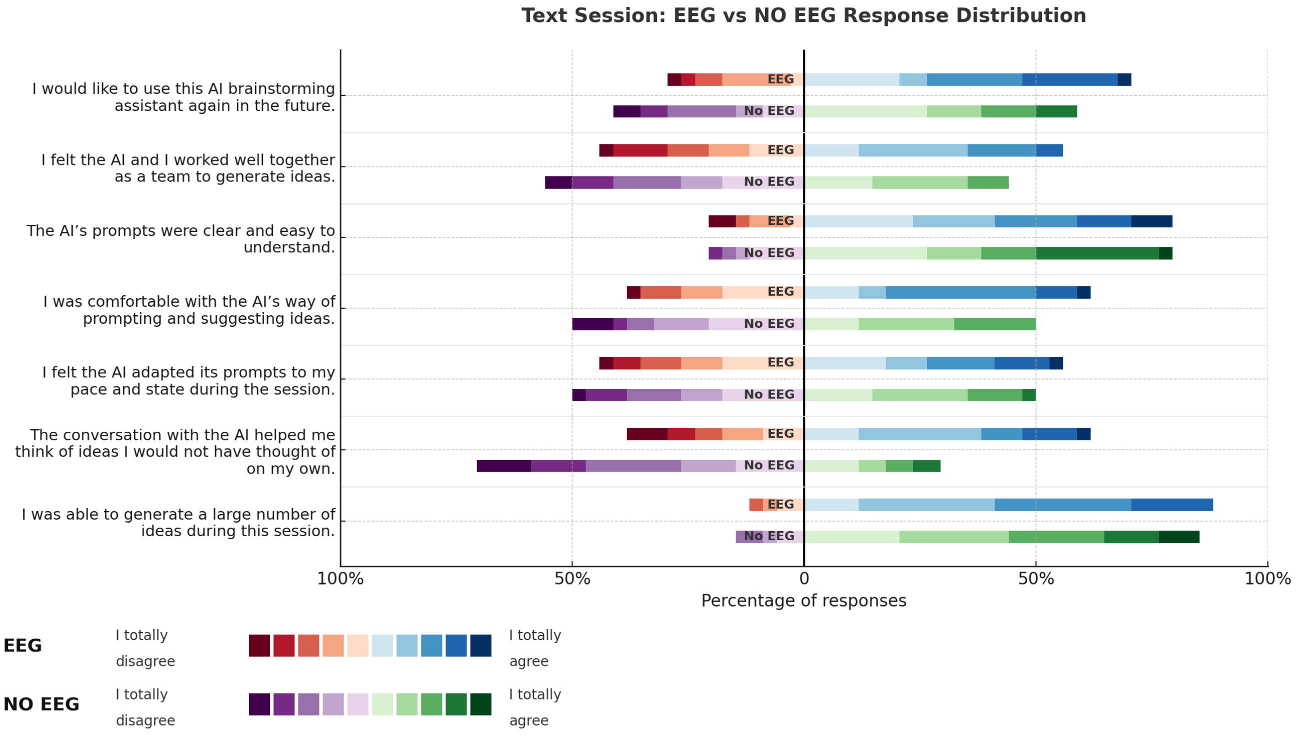


Figure 9: Questionnaire results for Study B (Text ideation) comparing EEG-adaptive and non-adaptive conditions (N=29). Diverging bar chart shows Likert-scale response distributions with red/orange bars representing EEG condition and purple/green bars representing No EEG control. The EEG-adaptive system received more positive ratings across most dimensions, particularly for teamwork perception, idea generation assistance, and future usage intention.

T1: Perceived Personalization and Idea Stimulation. Across sessions, EEG runs were more often described as offering tailored cues that nudged participants past pauses or broadened their angle of attack. Participants pointed to keyword-like hints and "frames" (e.g., role–place–action) as helpful when stuck, while Normal runs were frequently characterized as flat encouragement or trivial restatements. *Trend: EEG ⟶ more "personalized" sparks; Normal ⟶ steadier but less inspiring.*

Several participants credited EEG with concrete lifts in ideation: "In the second EEG session it introduced keywords like dynamics and relaxation—that helped me think of related directions" (S18, translated); "The first EEG was somewhat helpful because it listed more topics; those three themes did have an effect" (S17, translated). Others highlighted specific idea injections: "It suggested adding music to improve the experience, which I hadn't thought of" (S27, translated); "It suddenly told me I could write who is doing what and gave me three words—that led me to generate new ideas" (S34, translated). By contrast, Normal was often described as non-stimulating: "It just kept repeating one sentence… I didn't feel any inspiration" (S02, translated); "The words only varied slightly each time" (S23, translated). Still, EEG's personalization did not land for everyone. Some found the generic "who–where–do-what" schema unhelpful in their context (e.g., "I read it but it didn't really help," S29, translated), or felt EEG merely anticipated what they were already about to say (S25). A few also reported receiving sufficient ideas on their own in Normal, leaving little need for assistance (S39). Mini-synthesis: Personalization helped many—especially when momentarily stuck—but its value depended on fit with the user's task and timing.

T2: Clarity and Relevance of Prompts. Perceived clarity diverged sharply. EEG interventions were sometimes experienced as abrupt or semantically unclear—useful words appearing at the wrong moment, or without enough connective tissue to feel conversational. Normal, while repetitive, was easier to interpret. Trend: EEG ⟶ higher variance in clarity; Normal ⟶ clearer but often generic. Representative accounts: "In the second EEG session something suddenly popped out and I felt puzzled—I didn't really know what it was trying to do in real time" (S06, translated); "Sometimes it didn't feel like a conversation; it was just throwing out words" (S18, translated). Normal, in contrast, was "more content overall… mostly supportive words, but not actually helpful" (S11, translated). Participants also noted presentational frictions: "It was popping out one word after another; the language could be more natural" (S26, translated). At the same time, some EEG scaffolds were praised for being concrete (e.g., the role–place–action structure felt "more specific and useful" than abstract affective labels, S16, translated).

Table 5: Qualitative themes from post-session interviews comparing EEG-adaptive and non-adaptive *text ideation* experiences. EEG provided more personalized sparks and divergent pushes but at times felt opaque or mistimed; Normal was clearer and calmer but less stimulating.

Theme	EEG (Experimental group)	Normal (Control group)
Personalization and Idea Stimulation	Tailored cues/keywords; nudged users past pauses; useful when stuck	Steadier encouragement but generic/repetitive; fewer sparks
Clarity or Relevance of Prompts	Sometimes abrupt/opaque; structured frames clearer	Easier to interpret but generic; supportive yet shallow
Depth and Divergence	Promoted divergence and lateral moves; added detail/switches when relevant	Refined existing ideas; incremental, risk of superficiality
User Experience	Engaging but added time pressure/flow resets; needed gentler, adaptive cadence	Calmer and lower friction, but flatter engagement
Overall Satisfaction and Preference	More novelty/breakthroughs despite occasional irrelevance/opacity; preferred by users seeking stimulus	Safer and steadier; preferred by users seeking clarity/low friction
Trade-offs or Moderators	Benefits shaped by task fit, expectations, and timing	Clarity/calm benefits with less discovery; same moderators apply

Mini-synthesis: Clarity hinged on the form and timing of EEG prompts—structure helped; opaque affective tags did not.

T3: Depth and Divergence of Ideas. Participants frequently described EEG as better at pushing divergence—adding detail, switching angles, or proposing lateral moves—whereas Normal tended to refine what was already on the table. Trend: EEG ⟶ more divergent and elaborative; Normal ⟶ incremental elaboration.

Positive cases centered on concrete pushes: “EEG suggested adding music to improve the experience” (S27); “It told me to write who is doing what and gave three words; that opened new directions” (S34); “EEG made me consider different angles—mood vs. experience—so new ideas came out” (S20; all translated). By contrast, Normal was “just refining my existing idea without helping me think of new ones” (S28, translated). Yet some EEG prompts were experienced as off-context: “It listed who/where/what but that didn’t really fit this task,” (S11, translated), echoing concerns in PS2. *Mini-synthesis: EEG more often catalyzed divergence, conditional on relevance; Normal provided stability but risked superficiality.*

T4: User Experience and Emotional Reactions (Stress, Flow, Timing). EEG’s real-time nature occasionally heightened time pressure or interruption cost. Participants mentioned watching countdowns, rushing to read prompts, or feeling their flow reset by the system’s timing. Normal, while less engaging, was also less tense. *Trend: EEG ⟶ more engagement with potential stress; Normal ⟶ calmer but flatter.*

Illustrative accounts: “I worried I couldn’t finish typing; when I saw the countdown it reset right after giving feedback—that extra stress came from that” (S13, translated). Others asked for a prebrief of the EEG agent’s “role” or a gentler cadence: “It suddenly jumped in… I didn’t know what it was doing; please tell me in advance what its role is” (S06, translated). Participants also proposed adaptive intensity—lighter suggestions when stress is low, stronger guidance when stress is high (S20)—and early “starter” hints if the system detects long input gaps (S07). *Mini-synthesis: Emotional reactions were tightly coupled to* tempo *and* explanations*; pacing and forewarning buffered stress without sacrificing help.*

T5: Overall Satisfaction and Preferences. On balance, participants reported more moments of novelty and breakthrough under EEG, despite occasional irrelevance or confusion; Normal felt safer but unexciting. “The first Normal session was not helpful at all… the second EEG at least gave me some fresh words” (S02, translated). “If it were always Normal, it would be conventional; EEG, although sometimes abrupt, inspired me to think of something new” (S23, translated). Several others echoed “EEG had some help; Normal basically none” (S17, S18; translated). Countervailing voices preferred Normal’s steadiness or criticized EEG’s opacity (S11, S29, S24), underscoring taste and task fit. *Mini-synthesis: Satisfaction split along a control–novelty frontier—those seeking stimulus gravitated to EEG; those seeking crisp, low-friction guidance leaned Normal.*

Summary and Boundaries. Subjective evaluations portray a consistent trade-off: EEG prompts increased perceived personalization and divergence at the cost of occasional opacity and time-pressure; Normal prompts offered clarity and calm but less discovery. These judgments were moderated by prompt–task fit, user expectation (what “help” should look like), and the micro-timing of interventions. In particular, this picture mirrors the patterns observed in the image study: EEG felt more ’continuous’ and recoverable while Normal was more ’jumpy’ or easily derailed, suggesting consistency between tasks in how neuroadaptive scaffolding shapes user sense of control and progress.

6 Discussion

This work asked whether short, event-anchored EEG windows can productively gate *prompt-level* policies in open-ended human–AI collaboration. Across two matched studies, we find a consistent but asymmetric pattern: in image co-creation, EEG-gated branches improved perceived teaming, iteration efficiency, and satisfaction; in text ideation, EEG-gated nudges improved perceived collaboration but produced limited and variable changes in stress dynamics and idea outcomes. Below we consolidate what changed, why effects differed by task, and how to design affect-adaptive prompting that is stable, legible, and respectful of user state.

6.1 Key findings across two creative tasks

Image co-creation (Study A).. Under identical UI and timing, the EEG condition raised user ratings ($\beta = +0.50$, 95% CI [0.10, 0.91], $p = .016$; Table 2) and produced later-phase trajectories with faster recovery and earlier peaks in Excitement (Fig. 6). Qualitatively, participants repeatedly described EEG runs as *more continuous* and easier to "carry on," in contrast to Normal runs that "jumped" or "drifted" between steps. Typical comments include: "in the EEG session it felt coherent and aware of what I'd said before; in Normal it changed too fast" (S06, translated), and "EEG made me want to keep going" (S01, translated). Conversely, a minority preferred Normal's unpredictability for serendipity when exploration was the goal (e.g., S27 found an unexpected base style and stopped there).

Text ideation (Study B).. EEG-triggered nudges (based on 10s post-reply Stress changes) were perceived as more stimulating and personalized by many participants—e.g., "it introduced keywords like *dynamics* and *relaxation* that opened directions" (S18), and "it suggested adding music, which I hadn't thought of" (S27). Yet others found some interventions abrupt or semantically thin—"it suddenly popped something up; I didn't know what it was trying to do" (S06)—and Normal prompts, while repetitive, easier to parse ("it kept repeating one sentence; no inspiration," S02). In aggregate, survey items on teaming and comfort moved positively for EEG, whereas per-participant stress variability did not differ significantly from control (Table 4).

Why adaptation helped more for images. The asymmetry across tasks reflects three intertwined factors. First, artifact immediacy: visual outputs allow EEG signals to deterministically alter what appears next, creating a legible chain from signal to branch to visible change, which reinforced momentum and continuity. Second, policy determinism: image branches prescribed concrete operations (two-stage edit vs. new prompt), while text branches only modulated the style and force of a nudge, leaving more variance to user interpretation and timing. Finally, signal–task alignment: arousal- and interest-driven tags map naturally onto "explore vs. refine" choices in visual iteration, whereas stress-driven nudges in text chat could either unblock or interrupt depending on conversational context, reducing the consistency of benefits.

6.2 Implications for affect-adaptive prompting

Drawing on the above mechanisms and our design principles, we articulate seven actionable guidelines for neuroadaptive LLM interfaces:

- **Anchor to UI events.** Compute short, post-stimulus windows (e.g., 0–5s after image display; 10s after AI reply) to avoid motor contamination and minimize latency drift.
- **Gate with discrete tags.** Convert small relative changes (e.g., ±10% arousal; ±5% stress) into a few auditable policy states instead of injecting raw numbers into prompts.
- **Arbitrate with a second channel.** Use a coarse valence cue (e.g., Interest) to up/down-weight arousal-driven decisions and curb false positives in "refine vs. regenerate."
- **Front-load exploration, then local edits.** Default to exploration early; when arousal/interest rise, pivot to precise edits to reduce churn and strengthen continuity.
- **Pre-brief the agent's role.** Before adaptive nudges appear, prime users that "the assistant may occasionally offer structured hints when you seem stuck," reducing surprise cost.
- **Match nudge force to stress.** Under LOW stress, prefer light, optional cues; under HIGH stress, escalate to concrete scaffolds (e.g., role–place–action frames or idea templates).
- **Explain minimally, log completely.** Keep the UI free of raw physiology to avoid demand characteristics, but fully log windows/tags/branches for audit and debugging.

When adaptation backfires. Two failure modes emerged. *(i) Timing/opacity.* Some EEG interventions arrived mid-typing or felt like "throwing out words," raising pressure under countdowns (S06/S13); pre-briefs and gentler cadence can buffer this. *(ii) Under/over-steer.* Participants wanted micro-hints when they had ideas flowing, and stronger scaffolds only when truly stuck; purely generic encouragement in control felt safe but unhelpful (S11/S23), while overly abstract EEG cues sometimes missed context (S29). These tensions argue for *intensity control* and *context-aware templates* rather than one-size-fits-all nudges.

6.3 Limitations, ethics, and future work

Construct validity. We treat Emotiv Performance Metrics as black-box affect observables; algorithms are proprietary and update at 2 Hz. Our event-anchored, within-session deltas mitigate but do not remove detector latency.

Internal validity. We hid EEG numerics and matched UI/timing across conditions, yet adaptive branches may still have been indirectly noticeable. Subtle expectancy effects cannot be ruled out.

External validity. Tasks (chair/lamp; WFH scenarios) and a campus sample limit generality; Study B sessions were brief (3 min), potentially underestimating long-horizon benefits.

Ethics and privacy. EEG traces can bear biometric signatures; our design avoids persisting raw signals in UI and stores only minimal logs for analysis. Broader deployment should include consent granularity, on-device processing, and retention controls.

Future directions. Two promising paths are: **(1) Personalization kernels** that learn per-user thresholds/utilities while keeping the discrete policy surface; **(2) Policy introspection** that surfaces a short "why this action" trace (without exposing physiology), supporting trust and user correction.

7 Conclusion

We presented a general framework that steers large models via *prompt-level* policies gated by short, event-anchored EEG windows. The framework runs in real time without weight updates and keeps adaptation auditable. In image co-creation, EEG gating improved perceived teaming, iteration efficiency, and satisfaction; in text ideation, it improved the sense of collaboration while effects on stress variability and idea outcomes were limited and individual-dependent. From these results we distill design guidance—anchor to renders, gate with discrete tags, arbitrate with a second channel, and stage exploration before local edits—that transfers to other fast implicit channels beyond EEG. We release implementation details and logging schemas to support replication and sensitivity analyses (artifacts withheld for review), and argue that neuroadaptive *prompt policies* offer a pragmatic route to emotion-aware, stable, and interpretable agents across creative modalities.

Acknowledgments

The authors thank all the participants. This research was approved by the TU Delft HREC (5380).

References

[1] 2017. Evaluation of a Dry EEG System for Application of Passive Brain-Computer Interfaces in Autonomous Driving. *Frontiers in Human Neuroscience* 11 (2017), 00078. doi:10.3389/fnhum.2017.00078

[2] 2017. Online EEG-Based Workload Adaptation of an Arithmetic Learning Environment. *Frontiers in Human Neuroscience* 11 (2017), 243. doi:10.3389/fnhum.2017.00243

[3] 2020. Human Emotion Recognition: Review of Sensors and Methods. *Sensors* 20, 3 (2020), 592. doi:10.3390/s20030592

[4] 2024. A Comprehensive Review of Multimodal Emotion Recognition: Techniques, Challenges, and Future Directions. *Frontiers in Psychology* (2024). https://pmc.ncbi.nlm.nih.gov/articles/PMC12292624/

[5] 2024. Enhancing Robot Behavior with EEG, Reinforcement Learning and Beyond: A Review of Techniques in Collaborative Robotics. *Applied Sciences* 14, 14 (2024), 6345. doi:10.3390/app14146345

[6] 2024. From Neural Networks to Emotional Networks: A Systematic Review of EEG-Based Emotion Recognition in Cognitive Neuroscience and Real-World Applications. *Frontiers in Neuroscience* (2024). https://pmc.ncbi.nlm.nih.gov/articles/PMC11940461/

[7] 2024. NeuroChat — MIT Media Lab. https://www.media.mit.edu/projects/neurochat-ai-bci/overview/

[8] Mohamed Aly and Abdelrahman Aly. 2020. Adaptive user interface design and analysis using emotion recognition through facial expressions and body posture from an RGB-D sensor. *PLOS ONE* 15, 7 (2020), e0235908. doi:10.1371/journal.pone.0235908

[9] Matthijs Baas and Marieke Roskes. 2020. Mood and Risk-Taking as Momentum for Creativity. *Frontiers in Psychology* 11 (2020), 610562. doi:10.3389/fpsyg.2020.610562

[10] Oana Balan, Gabriela Moise, et al. 2019. Automatic Adaptation of Exposure Intensity in VR Acrophobia Therapy. In *Proceedings of the European Conference on Information Systems (ECIS)*. https://aisel.aisnet.org/ecis2019_rp/52/

[11] D. Baradari, N. Kosmyna, O. Petrov, R. Kaplun, and P. Maes. 2025. NeuroChat: A Neuroadaptive AI Chatbot for Customizing Learning Experiences. *arXiv* (2025). arXiv:2503.07599

[12] Dünya Baradari, Nataliya Kosmyna, Oscar Petrov, Rebecah Kaplun, and Pattie Maes. 2025. NeuroChat: A Neuroadaptive AI Chatbot for Customizing Learning Experiences. *arXiv preprint arXiv:2503.07599* (2025).

[13] Maxime Beauchemin et al. 2024. EEG-Based Passive BCI System Adapts Learning Speed to Cognitive Load in Real Time. *Frontiers in Human Neuroscience* 18 (2024).

[14] BrainAccess. 2024. Rethinking Education with EEG: Hyperscanning, AI, and the Future of Learning. BrainAccess Blog. https://www.brainaccess.ai/rethinking-education-with-eeg-hyperscanning-ai-and-the-future-of-learning/

[15] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, et al. 2020. Language Models are Few-Shot Learners. *Advances in neural information processing systems* 33 (2020), 1877–1901. https://arxiv.org/abs/2005.14165

[16] N. Crilly. 2018. Managing the tension between persistence and flexibility in creative tasks. https://www.repository.cam.ac.uk/bitstreams/1a875792-e1bd-4bb4-878d-90ab10a480e1/download

[17] Sinéad Cullen. 2024. Befriending your creative process. Medium. https://medium.com/@Creativityforchange/befriending-your-creative-process-f48cf12da906

[18] Label Your Data. 2025. In-Context Learning: Enhancing Model Performance in 2025. Label Your Data Blog. https://labelyourdata.com/articles/llm-fine-tuning/in-context-learning

[19] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805* (2018). https://arxiv.org/abs/1810.04805

[20] Drew. 2024. Does frustration fuel our creativity? Actually, yes. Rise With Drew Blog. https://risewithdrew.com/does-frustration-fuel-our-creativity-actually-yes/

[21] MDPI Electronics. 2025. Smart Transparency: A User-Centered Approach to Improving. *Electronics* 14, 3 (2025), 420. https://www.mdpi.com/2079-9292/14/3/420

[22] Emotiv. 2017. What are the Performance Metrics Detection Suite? https://www.emotiv.com/tools/knowledge-base/faq/what-are-the-performance-metrics-detection-suite

[23] Emotiv. 2023. How do you measure emotions in the first place to compare the outputs and come up with a number? https://www.emotiv.com/tools/knowledge-base/epoc-getting-started/how-do-you-measure-emotions-in-the-first-place-to-compare-the-outputs-and-come-up-with-a-number Accessed: 2025-06-02.

[24] Emotiv. 2025. *EPOC+ Headset Details*. https://emotiv.gitbook.io/epoc-user-manual/using-headset/epoc+_headset_details

[25] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences. *Behavior Research Methods* 39, 2 (2007), 175–191. doi:10.3758/BF03193146

[26] Giovanni Ferraro. 2025. Neuroadaptive XR Interfaces: A Generative AI Framework for Real-Time Emotion-Aware Personalization in VR/AR. *International Journal of Virtual and Augmented Reality* 3, 1 (2025).

[27] Barbara L. Fredrickson. 2005. Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition & Emotion* 19, 3 (2005), 313–332. doi:10.1080/02699930441000238

[28] Lukas Gehrke, Aleksandrs Koselevs, Marius Klug, and Klaus Gramann. 2025. Neuroadaptive Haptics: Comparing Reinforcement Learning from Explicit Ratings and Neural Signals for Adaptive XR Systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2025).

[29] Open Access Government. 2024. AI breakthrough reveals new ways of observing gravitational waves and extreme events. Online article. https://www.openaccessgovernment.org/ai-breakthrough-reveals-new-ways-of-observing-gravitational-waves-and-extreme-events/191738/

[30] Aakash Gupta. 2024. Context Engineering: The Evolution Beyond Prompt Engineering That's Revolutionizing AI Agent. Medium. https://aakashgupta.medium.com/context-engineering-the-evolution-beyond-prompt-engineering-thats-revolutionizing-ai-agent-0dcd57095c50

[31] IJIRSET. 2025. Adaptive Interface Framework. *International Journal of Innovative Research in Science, Engineering and Technology* (March 2025). https://www.ijirset.com/upload/2025/march/360_Adaptive.pdf

[32] Z. Ivcevic and J. Hoffmann. 2023. Emotions and Creativity: From Process to Person and Product. *ResearchGate* (2023). https://www.researchgate.net/publication/373951088_Emotions_and_Creativity_From_Process_to_Person_and_Product

[33] H. Jiang, Y. Chen, D. Wu, and J. Yan. 2024. EEG-Driven Automatic Generation of Emotive Music Based on Transformer. *Frontiers in Neurorobotics* 18 (2024), 1437737.

[34] Ilkka Kosunen et al. 2016. RelaWorld: Neuroadaptive Relaxation in Virtual Reality. In *IUI*.

[35] Minji Lee, Joon Park, et al. 2024. Exploring the Impact of AI Value Alignment in Collaborative Ideation: Effects on Perception, Ownership, and Output. *arXiv preprint* (2024). https://arxiv.org/html/2402.12814v2

[36] Jesse A. Mark et al. 2022. Neuroadaptive Flight Training via fNIRS. *Frontiers in Neuroergonomics* 3 (2022).

[37] C. Martinho, L. Roque, et al. 2024. Human-AI Co-Creativity: Exploring Synergies Across Levels of Creative Collaboration. https://arxiv.org/html/2411.12527v2 arXiv:2411.12527v2.

[38] Ryan McGuire, Emily Johnson, et al. 2024. Establishing the importance of co-creation and self-efficacy in creative collaboration with AI. *Scientific Reports* 14, 1 (2024), 12345. doi:10.1038/s41598-024-69423-2

[39] Ananya Mishra et al. 2024. Thought2Text: Text Generation from EEG Signals using Large Language Models. In *EMNLP (to appear)*.

[40] Decrypt News. 2024. Google DeepMind AI Cracks Century-Old Fluid Mysteries, Pointing to New Era in Science. Online news. https://decrypt.co/340451/google-deepmind-ai-cracks-fluid-mysteries-new-era-science

[41] Paulo A. Nogueira, Rui Rodrigues, Eduardo Oliveira, and Urbano J. Nunes. 2011. Emotion assessment from physiological signals for adaptation of game difficulty. In *2011 IEEE 1st International Conference on Serious Games and Applications for Health (SeGAH)*. IEEE, 1–6. doi:10.1109/SeGAH.2011.6165464

[42] Luiz Pessoa. 2010. The impact of emotion on perception, attention, memory, and decision-making. *Nature Reviews Neuroscience* 11, 9 (2010), 673–682. doi:10.1038/nrn2901

[43] S. Pillalamarri, J. Kim, and R. Rao. 2025. Multimodal Physiological Sensing for Adaptive Human–AI Interaction: A Review. *IEEE Transactions on Human–Machine Systems* 55, 2 (2025), 1–19.

[44] Mark O. Riedl et al. 2014. EEG-Adaptive Intelligent Tutoring: A Closed-Loop Study. In *ICCE*.

[45] Greg Robison. 2024. Move 37 for AI Design: Why Machines Are Now Discovering Breakthroughs Humans Never Imagined. Medium. https://gregrobison.medium.com/move-37-for-ai-design-why-machines-are-now-discovering-breakthroughs-humans-never-imagined-2c4d0a9d71d4

[46] A. Savić, D. Novak, and A. Beg. 2024. ChatGPT and BCI–VR: A New Integrated Diagnostic and Therapeutic Paradigm. *Frontiers in Digital Health* 6 (2024), 1223456.

[47] Skyhigh Security. 2024. When AI Turns Adversarial: Lessons from the Claude-Powered Cybercrime Spree. Intelligence Digest. https://www.skyhighsecurity.com/about/resources/intelligence-digest/when-ai-turns-adversarial-lessons-from-the-claude-powered-cybercrime-spree.html

[48] Ana M. Serrano-Bedia, María López-Fernández, et al. 2024. Analysis of the influence of emotions on the decision-making of entrepreneurs using neurotechnologies. *International Entrepreneurship and Management Journal* (2024). doi:10.1007/s11365-024-00960-y

[49] S. Shukla, M. Patel, and M. Rahman. 2025. Bridging EEG Signals and Generative AI: A Survey. *arXiv* (2025). arXiv:2502.06543

[50] S. Singh et al. 2024. Artifact Removal from EEG Signals. *ResearchGate* (2024). https://www.researchgate.net/publication/260649732_Artifact_Removal_from_EEG_Signals

[51] J. Smith and A. Johnson. 2023. AI as a Creative Partner: Enhancing Artistic Creation and Acceptance. *ResearchGate* (2023). https://www.researchgate.net/publication/376021812_AI_as_a_Creative_Partner_Enhancing_Artistic_Creation_and_Acceptance

[52] P. Sorino, G. M. Biancofiore, D. Lofù, et al. 2024. ARIEL: Brain–Computer Interfaces Meet Large Language Models for Emotional Support Conversation. In *Adjunct Proceedings of UMAP '24*. ACM, 601–609.

[53] Psychology Town. 2024. Psychophysiological Assessment: Measuring the Body's Response to Behaviour. Psychology Town. https://psychology.town/psychodiagnostics/psychophysiological-assessment-body-response-behaviour/

[54] Liang Wang, Chen Wu, and Li Wang. 2024. BrainDreamer: Reasoning-Coherent and Controllable Image Generation from EEG via Language Guidance. *arXiv preprint arXiv:2409.14021* (2024).

[55] Rene Weber et al. 2024. Neuroadaptive VR Exposure Therapy for Arachnophobia. In *MetroXRAINE*.

[56] J. Wick et al. 2024. A scoping review on the use of consumer-grade EEG devices for research. *PMC* (2024). https://pmc.ncbi.nlm.nih.gov/articles/PMC10917334/

[57] Sandy Wu et al. 2021. AI Creativity: The Human–AI Co-Creation Model. In *Human–Computer Interaction: HCII 2021*. Springer, 215–231. doi:10.1007/978-3-030-78462-1_13

[58] Georgios N. Yannakakis, Antonios Liapis, and Constantine Alexopoulos. 2014. Mixed-Initiative Co-Creativity. In *Proceedings of the 9th International Conference on the Foundations of Digital Games*. ACM, 4–11. doi:10.1145/2556582.2556602

[59] Thorsten O. Zander and Christian Kothe. 2011. Towards Passive Brain–Computer Interfaces: Applying BCI Technology to Human–Machine Systems in General. *Journal of Neural Engineering* 8, 2 (2011), 025005.

[60] Xiaoli Zhou et al. 2024. Evaluating Recent Advances in Affective Intelligent Tutoring Systems: A Scoping Review of Educational Impacts and Future Prospects. *Education Sciences* 14, 8 (2024), 839. doi:10.3390/educsci14080839

[61] Y. Zhu, H. Wang, and P. Liu. 2023. DreamDiffusion: Generating High-Quality Images Directly from EEG. *arXiv* (2023). arXiv:2306.16934

A System Prompts

A.1 Prompt Templates (Study A: Image)

EmotivChat_Image 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.

B Prompt Templates (Study B: Text)

EmotivChat_Image 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.

C Full Questionnaires

C.1 Demographics & Background Survey

(1) Your age (years): {18–24, 25–30, 31–40, >40}
(2) Your academic/professional background: {Design, Science, Engineering, Social Sciences, Humanities, Medical/Health, Others}
(3) Frequency of using generative AI: 1=Never to 5=Everyday
(4) Familiarity with EEG technology: {None, Basic, Moderate, Advanced}
(5) Hours of sleep last night: {≤5, 5–6, 6–7, 7–8, 8–9, ≥9}
(6) Caffeine intake within 4h before experiment: {None, 1 cup, 2 cups, ≥3 cups}

C.2 Post Text-Session User Survey

(1) I feel calm and relaxed after this conversation. (1–5 Likert)
(2) The conversation with the AI helped reduce my level of stress. (Yes/No)
(3) I was comfortable with the AI's way of communicating (tone and wording). (1–5 Likert)
(4) I could easily understand everything the AI said to me. (1–5 Likert)
(5) I felt that the AI understood my feelings and needs during our conversation. (1–5 Likert)
(6) I would be happy to keep talking with this AI for a longer time. (Yes/No)

C.3 Post Image-Session User Survey

(1) I felt excited and energized while creating images with the AI. (1–5 Likert)
(2) I was absorbed in the image creation process with the AI, losing track of time or distractions. (Yes/No)
(3) I am satisfied with the images/results that I got from this co-creative session. (1–5 Likert)
(4) The process of iterating and refining images with the AI felt efficient and productive. (1–5 Likert)
(5) The AI understood what I was looking for in the images. (1–5 Likert)
(6) I would like to continue using this AI for creative design tasks in the future. (Yes/No)

C.4 Interview Questions (Text Session)

(1) During the conversations, was there a moment when you felt especially understood or comforted by the AI?
(2) If you could improve the AI's behavior or abilities for comforting people, what would you change or add first, and why?

C.5 Interview Questions (Image Session)

(1) Which of the images generated inspired you the most, and what about it or the process made it particularly inspiring?
(2) Did the AI's use of your emotional signals (from EEG) help your design process in any way? Can you explain why it did or did not help?