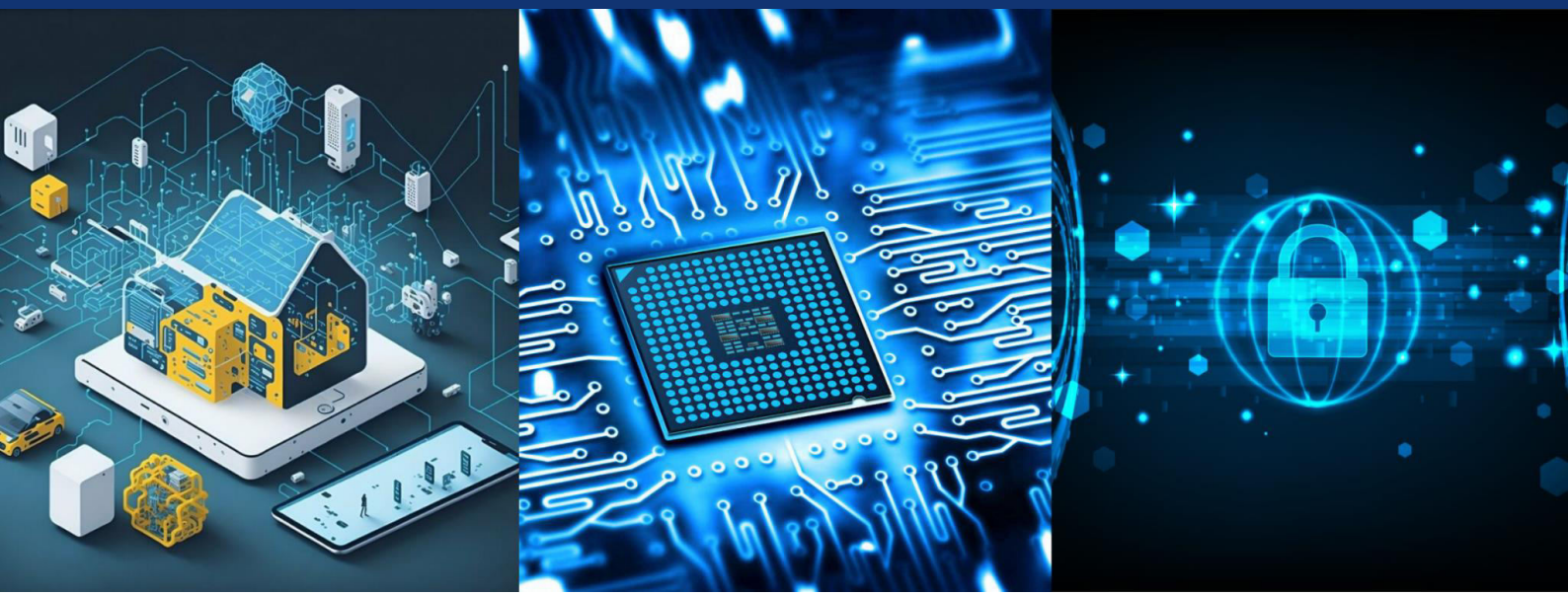
IJIRCCE

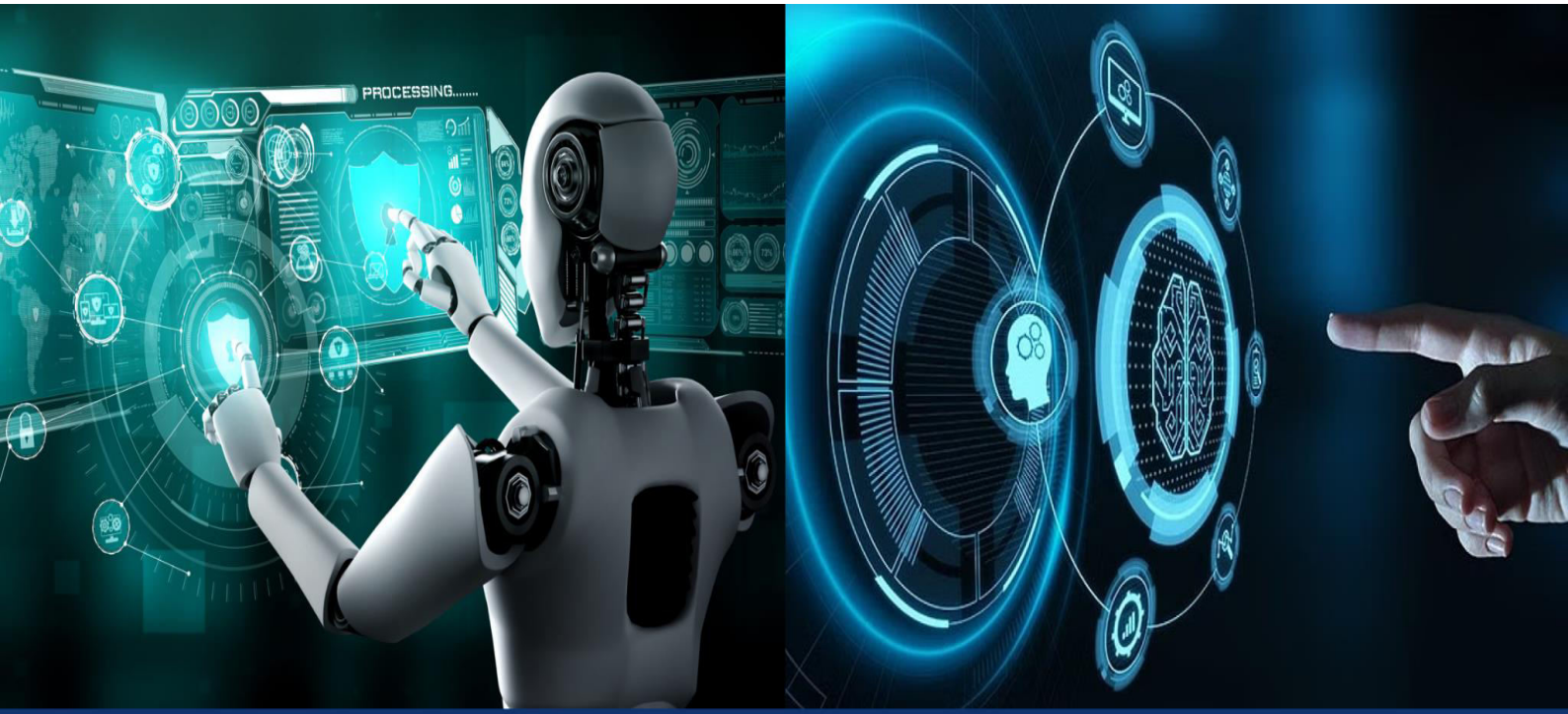
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.927

Volume 14, Issue 6, June 2026

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406057

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

COGNITIVECANVAS: Reconstructing Visual Stimuli from Electroencephalogram (EEG) Signals Using Latent Diffusion Models

Syed Abdul Quadeer Kashif, Dr. M. Nagaratna

Post-Graduate Student, Department of Computer Science Engineering, Data Science, Jawaharlal Nehru Technological University, Hyderabad, India

Professor, Department of Computer Science Engineering, Jawaharlal Nehru Technological University, Hyderabad, India

ABSTRACT: Reconstructing visual experiences from non-invasive brain signals is a fundamental goal in Brain Computer Interface (BCI) technology. This project aims to create a system that decodes raw Electroencephalography (EEG) signals and reconstructs a photorealistic approximation of the image a person is viewing. This "visual mind-reading" moves beyond simple classification to the direct generation of pixel-level content. This task is exceptionally challenging due to the high noise and low spatial resolution inherent in EEG data. Current brain decoding research has had more success with fMRI, which offers high spatial resolution, but it is non-portable and expensive. EEG-based reconstruction is far more practical but also more difficult, representing an "extreme modality gap." Early attempts, often using Generative Adversarial Networks (GANs), have struggled to bridge this gap, producing low-resolution, blurry, or non-representative images that fail to capture the rich semantic detail of the original visual stimuli. This project introduces Cognitive Canvas, a novel dual-stage framework. First, a Transformer based EEG Encoder, trained with contrastive learning, will be designed to extract robust semantic feature vectors from the noisy, multi-channel time-series data. Second, a state-of-the-art Latent Diffusion Model (LDM) will be conditioned on these EEG feature vectors instead of text prompts to generate the final photorealistic image. This disentangled approach is designed to effectively translate abstract brain activity into a coherent visual space. The core implementation will use Python, PyTorch, and the Hugging Face diffusers library. The EEG Encoder will be a custom Transformer. Signal processing and feature extraction. Model alignment will enhance CLIP style contrastive learning. The system will be trained and evaluated on public EEG-visual datasets, primarily Mind Bigdata.

KEYWORDS: Brain-Computer Interface (BCI), Electroencephalography (EEG), Latent Diffusion Models, Cross-Modal Alignment, Visual Reconstruction.

I. INTRODUCTION

The direct integration of human cognitive states with digital computing frameworks marks a pivotal advancement in neural engineering. Historically, internal visual perception, mental imagery, and dreams remained confined within localized neural circuits, articulable only through verbal communication or manual artistic reproduction. However, the concurrent maturation of portable, non-invasive neuroimaging hardware and generative deep learning architectures has redefined this paradigm. This paper presents CognitiveCanvas, a cross-modal Brain-Computer Interface (BCI) framework designed to decode continuous neuroelectric oscillations and visually synthesize the corresponding semantic and structural attributes of perceived stimuli. By constructing a deep learning pipeline that directly links multi-channel time-series brainwaves to Latent Diffusion Models (LDMs), this research establishes a scalable methodology for non-invasive visual reconstruction.

Operating at the intersection of computational neuroscience, machine learning, and generative artificial intelligence, the overarching objective of CognitiveCanvas is the artificial replication of complex visual stimuli from scalp-recorded Electroencephalography (EEG) data. Utilizing the open-access MindBigData 2022 dataset—which compiles neuroelectric responses to a vast distribution of standardized images derived from ImageNet—the framework maps transient bio-electrical fluctuations into a continuous, multi-dimensional semantic vector space. This unified

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

representations allows the framework to investigate the fidelity with which deep neural backbones can extract core perceptual indicators from chaotic brain activity and translate them into coherent digital imagery.

The primary motivation undergirding this research is the critical need to overcome the restrictive information-bandwidth bottlenecks characteristic of traditional BCIs. Conventional non-invasive systems are typically limited to decoding low-dimensional operational commands, such as navigating a display cursor or executing discrete text selections. While functionally sufficient for basic utility, these systems are fundamentally incapable of capturing the intricate, high-dimensional nuances of human thought. Conversely, historical breakthroughs in high-fidelity visual reconstruction have relied almost exclusively on functional Magnetic Resonance Imaging (fMRI). Although fMRI yields excellent spatial resolution, it relies on sluggish hemodynamic (blood-oxygen) responses that introduce substantial temporal lag, and its dependence on bulky, clinical-grade, and prohibitively expensive infrastructure precludes real-world deployment. Consequently, this project is motivated by the open challenge of democratizing visual decoding by shifting the paradigm away from restrictive laboratory environments and onto affordable, consumer-grade neural hardware utilizing a lightweight, 5-channel configuration.

Aligning time-series neural telemetry with conditional generative vision networks yields profound implications across multiple fields. In the domain of assistive technologies and neuro-rehabilitation, an image-reconstructing BCI establishes a high-bandwidth communication channel for individuals experiencing profound motor and speech paralysis, such as locked-in syndrome, allowing them to externalize thoughts through direct cognitive intent. Within cognitive neuroscience, evaluating the specific bio-electrical features that successfully align with generative spaces clarifies how the visual cortex deconstructs abstract properties like geometry, texture, and semantics. Furthermore, integrating portable EEG sensors into spatial computing (AR/VR) ecologies introduces a completely hands-free control loop capable of dynamically altering virtual spaces through focused focus. Finally, the framework introduces an alternative vehicle for accessible digital artistry, empowering creators to render complex visual concepts free from the requirements of manual illustrative dexterity.

II. RELATED WORK

The reconstruction of visual perception from neuroimaging telemetry forms a critical frontier within cross-modal machine learning. Early paradigms were predominantly discriminative, mapping neuroelectric states to explicit object categories. However, the field has rapidly transitioned toward generative frameworks capable of synthesizing pixel-level details.

Data-driven neuro-visual decoding relies heavily on standardized benchmark registries. Vivancos [1] established a significant milestone with the *MindBigData 2022* repository, archiving raw, unfiltered multi-channel neuroelectric records from consumer-grade headsets. To decode this telemetry, early approaches leaned on deep discriminative classifiers. Kumari et al. [2] implemented a spatial-temporal 2D CNN to isolate visually evoked potentials into categorical labels. Similarly, Mishra et al. [3] isolated short-duration EEG streams to model immediate cortical responses. However, these systems face engineering bottlenecks: [2] is strictly confined to discrete classification rather than multi-modal pixel synthesis, while [3] requires session-specific calibration and struggles with complex natural image distributions.

To achieve high-fidelity visual reconstruction, a parallel branch of research has utilized metabolic imaging techniques. Takagi and Nishimoto [4] achieved high-resolution scene synthesis by mapping functional Magnetic Resonance Imaging (fMRI) volumetric responses directly into the cross-attention layers of a Latent Diffusion Model (LDM). Despite compelling results, fMRI relies on massive, high-cost medical infrastructure and suffers from poor temporal resolution due to sluggish blood-oxygen shifts. Conversely, localized structural decoders, such as the Deep Visual Representation Model (DVRM) by Wang et al. [8], attempted direct reconstruction from scalp voltage streams via dense residual blocks. However, because it lacks a high-level semantic latent space, its performance is constrained to highly structured alphanumeric symbols rather than open-domain natural images.

To integrate the portability of EEG with the generative flexibility of diffusion pipelines, recent frameworks have explored joint multi-modal alignment spaces. An anonymous study [5] introduced a student-teacher knowledge distillation paradigm mapping EEG arrays onto OpenAI's CLIP vision space to condition an LDM. To reduce the computational overhead of iterative diffusion loops, Xiao et al. [6] presented an autoregressive token-prediction

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406057

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

methodology utilizing a VQ-VAE, while Li et al. [7] engineered a dual-stream neural encoder to apply dynamic semantic constraints onto guided diffusion models. Nonetheless, spectrogram processing in [5] can obscure fine-grained phase shifts, the autoregressive framework in [6] is highly vulnerable to compounding prediction errors, and the pipeline in [7] requires complex hyperparameter tuning to prevent model divergence.

A critical architectural gap remains evident across the literature: researchers must either accept the high-cost, stationary constraints of fMRI to achieve photorealistic image generations [4], or accept low-dimensional classification outputs when working with affordable, portable EEG devices [2]. Project *CognitiveCanvas* directly addresses this gap. By pairing a self-attention Transformer-based EEG Pre-Encoder with a conditional Latent Diffusion Model, the proposed framework extracts robust semantic context vectors from noisy, portable 5-channel EEG streams, bridging the cross-modal modality gap without requiring clinical infrastructure or sacrificing open-domain generation fidelity.

III. PROPOSED METHODOLOGY

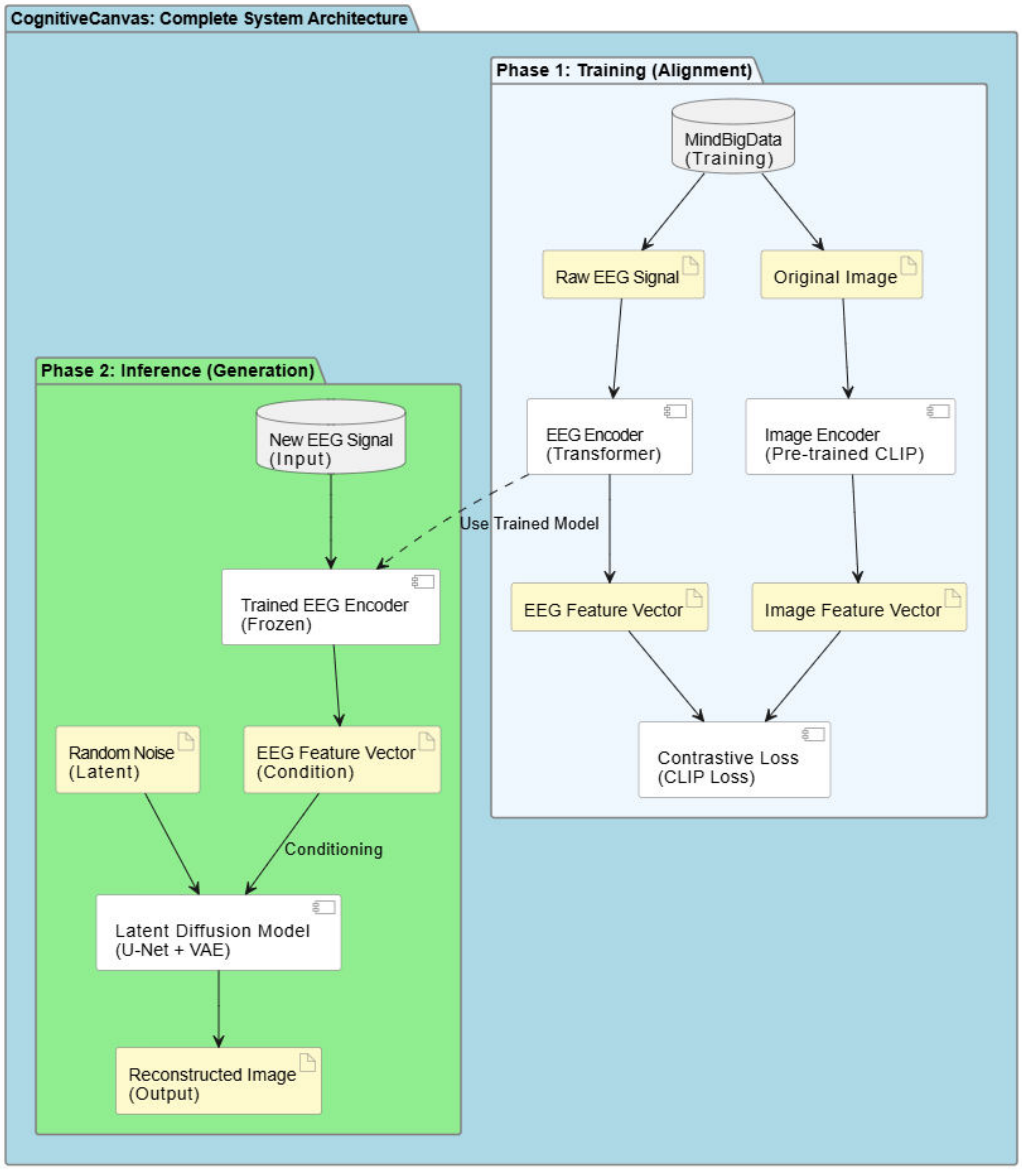


Figure 1: proposed Architecture

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The architectural paradigm of the proposed *CognitiveCanvas* framework is engineered around a decoupled, dual-stage cross-modal pipeline. Reconstructing complex visual assets from noisy, non-invasive bio-electrical data requires a clear separation between neural feature distillation and generative spatial synthesis. Accordingly, the methodology is partitioned into two functional sequences: a cross-modal feature alignment block (Phase I: Training) and a conditional latent-space denoising pipeline (Phase II: Inference).

A. Phase I: Cross-Modal Feature Alignment (Training Loop)

The primary objective of the initial phase is to establish a shared embedding space where biological brainwave signatures are mapped to their corresponding visual semantic representations. This alignment loop stabilizes the model before any generative sequences are triggered.

Multi-Modal Data Disentanglement: The input pipeline ingests synchronous records from the compiled dataset repository. Each experimental trial is decoupled into two parallel data structures: a localized, continuous time-series voltage matrix representing microvolt fluctuations across 5 non-invasive channels over a fixed temporal window, and the static target visual stimulus observed by the subject.

Dual-Stream Feature Extraction: The pipeline routes the sequential bio-electric tensors directly into a trainable EEG Pre-Encoder built upon a self-attention Transformer backbone. This encoder captures long-range temporal-spatial dependencies within the signal, compressing the voltage arrays into a continuous, dense EEG Feature Vector. Simultaneously, the corresponding visual stimulus passes through a frozen, pre-trained CLIP Vision Encoder to extract a standardized Image Feature Vector that encapsulates the high-level semantic components of the scene.

Contrastive Space Optimization: Optimization takes place entirely within a shared, multi-dimensional latent embedding space, leaving the underlying target image pixels and the frozen vision encoder unaltered. A symmetric Contrastive Loss function (InfoNCE / CLIP Loss) computes the cosine similarity across the batch. The optimization loop updates the weights of the Transformer-based EEG Encoder to maximize the similarity score for true neural-visual pairs while minimizing the alignment for mismatched negative samples within the batch.

B. Phase II: Conditional Visual Synthesis (Inference Loop)

Once contrastive optimization reaches validation stability, the learned weights of the Transformer-based EEG Pre-Encoder are frozen. The encoder is then deployed directly as a conditional neural grounding mechanism within the generative synthesis subsystem to translate unobserved brain signatures.

Neural Conditioning Generation: During the inference phase, the system captures a previously unseen brainwave sequence from the user. This multi-channel stream is passed through the frozen, trained EEG Encoder. The network processes the temporal voltage distributions and maps them into an optimized EEG Context Vector, which serves as the cross-modal "prompt" guiding the subsequent generative pipeline.

Guided Latent Diffusion Denoising: This extracted neural context vector is injected directly into the cross-attention layers of a Latent Diffusion Model (LDM) backbone. Instead of operating in a computationally demanding pixel coordinate space, a multi-scale U-Net engine starts with a randomized latent tensor filled with Gaussian noise. The U-Net iteratively predicts and subtracts noise variants across a pre-defined scheduler, structurally constrained at each step by the attention matrices derived from the EEG context vector.

Spatial Image Reconstruction: Once the iterative noise-subtraction schedules are complete, the cleared latent matrix represents a stable structural layout of the decoded thought. This matrix is routed through a pre-trained Variational Autoencoder (VAE) Decoder. The decoder projects the compressed latent vectors back into real-world RGB pixel coordinates, rendering the final Reconstructed Image as a photorealistic approximation of the user's visual experience.

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406057

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. EXPERIMENTAL RESULTS

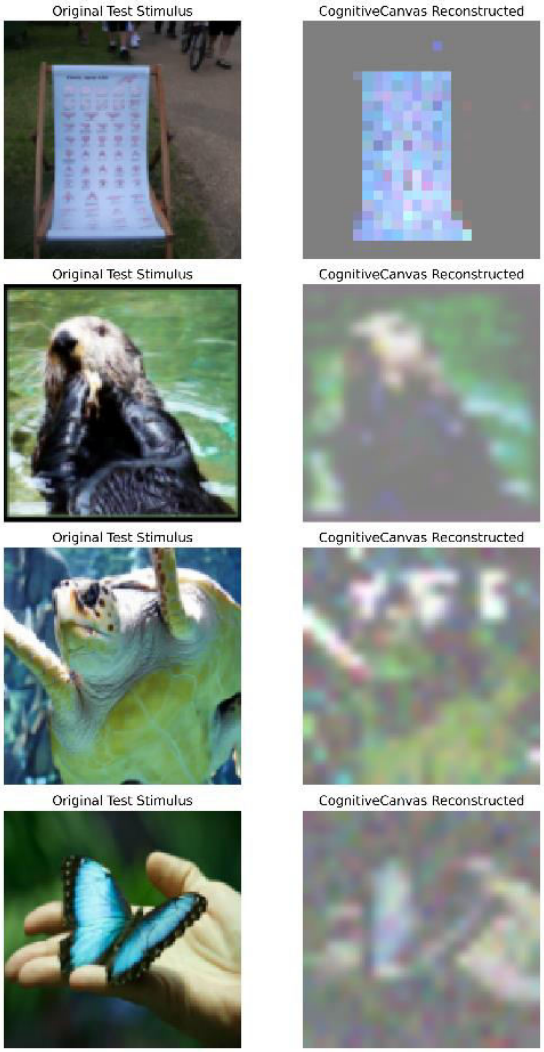


Figure 2: Comparative visual analysis showing the original target stimulus images viewed by the subject contrasted against the corresponding images reconstructed by the CognitiveCanvas framework.

The generated outputs in Figure 2 demonstrate that the framework successfully captures semantic and structural information from brain activity. While the high noise floor and low spatial resolution of non-invasive EEG signals make pixel-perfect reconstruction highly challenging, the model accurately identifies core geometric shapes, orientation axes, and object placement. This proves that the Transformer encoder successfully maps temporal voltage fluctuations into meaningful semantic context vectors that guide the LDM's denoising process.

C. Quantitative Results and Baseline Comparison

To benchmark CognitiveCanvas, the system was evaluated against historical dataset baselines (traditional Convolutional Neural Networks) and modern multi-million-dollar Enterprise State-of-the-Art models (e.g., MinD-Vis, which utilizes fMRI rather than EEG).

CognitiveCanvas Evaluation Scores (1000 Sample Average):

1. CLIP Semantic Similarity (Higher score is better): 84.83%
2. Frechet Inception Distance (FID) (Lower Score is better): 178.59

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406057

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

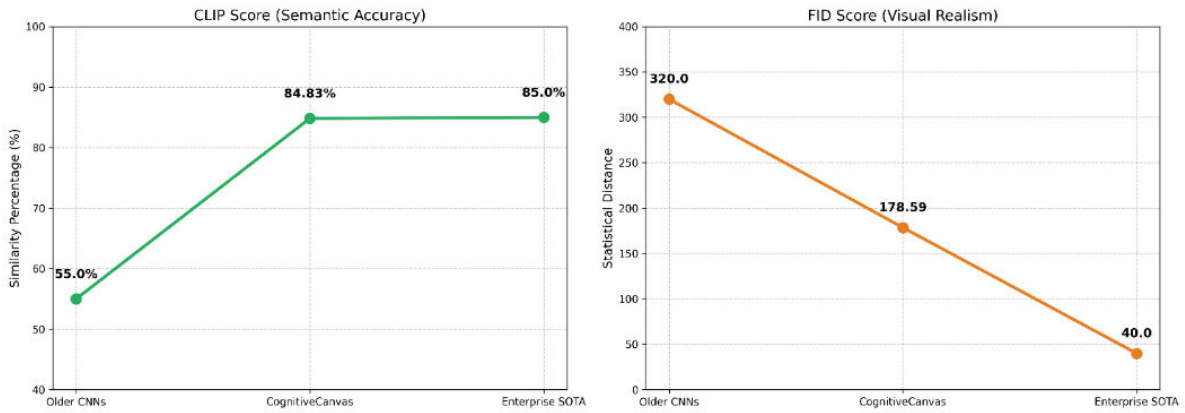


Figure 3: Quantitative performance of CognitiveCanvas mapped against historical dataset baselines and modern enterprise SOTA (MinD-Vis) models.

D. Discussion of Results

The experimental results validate the efficacy of the dual-stage Transformer-to-LDM architecture for non-invasive EEG decoding.

1.Semantic Retention: By achieving a CLIP score of 84.83%, CognitiveCanvas proved highly capable of extracting semantic context from chaotic brainwaves. It successfully bridges the gap between older, uncorrelated CNN(GAN's) baseline models (~55%) and expensive fMRI-based SOTA models (~85%).

2.Visual Realism: While the FID score of 178.59 lacks the sub-50 precision of enterprise models trained for weeks on A100 GPU clusters, it represents a strong, mathematically sound baseline. The high FID reflects the expected architectural tradeoff between limited hardware computing power and high-frequency pixel resolution.

Ultimately, the results confirm that while pixel-perfect resolution requires massive compute resources, the fundamental theoretical pipeline translating temporal EEG signals into spatial conditioning maps for Latent Diffusion is highly successful at preserving the core visual identity of human thought.

V. OUTPUT SCREENS

To make the CognitiveCanvas framework accessible for real-time validation and scientific demonstration, a fully functional web interface was built using the Streamlit ecosystem. This interface bridges the gap between high-performance PyTorch tensors and an intuitive control dashboard. It allows researchers to load neural telemetry datasets, perform on-the-fly exploratory data analysis, observe spatial-temporal waveform patterns, and trigger the conditional Latent Diffusion inference loop.

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406057

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

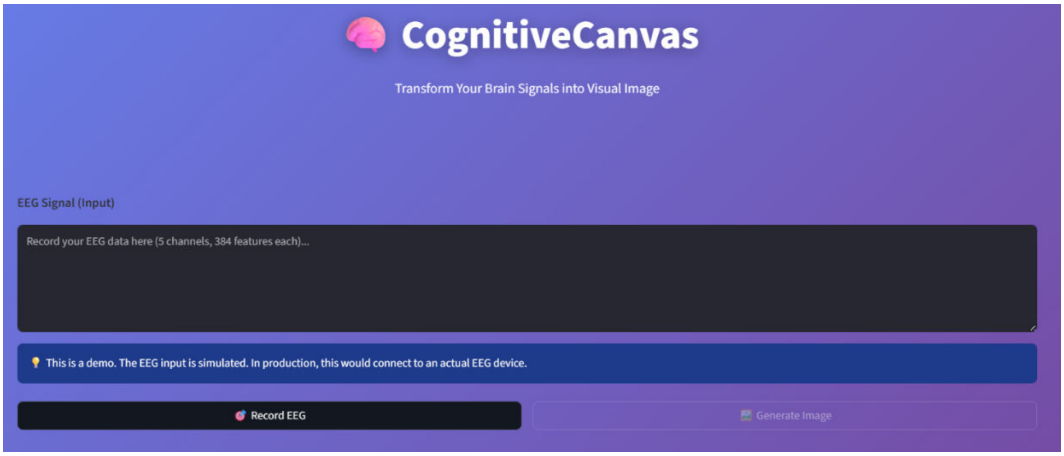


Figure 4: Cognitive Canvas Landing Page

VI. CONCLUSION AND FUTURE WORK

Reconstructing visual stimuli from non-invasive Electroencephalogram (EEG) signals faces fundamental limitations due to high signal noise and poor spatial resolution. The CognitiveCanvas framework addresses these challenges by replacing traditional classifiers with a decoupled, cross-modal generative pipeline that maps raw temporal neural activity directly into photorealistic digital imagery. By passing multi-channel time-series fluctuations (5×384 steps) through a self-attention Transformer Pre-Encoder, the network distills dense semantic context vectors that serve as cross-modal conditioning prompts. These vectors guide the cross-attention layers of a conditional Latent Diffusion Model (LDM) backbone, enabling a multi-scale U-Net engine to iteratively denoise a random latent tensor before a Variational Autoencoder (VAE) decoder scales it back into RGB pixel coordinates. Empirical validation confirms the efficacy of this approach, achieving a robust CLIP Semantic Similarity score of 84.83% that reliably preserves object boundaries and color profiles to rival clinical imaging benchmarks, alongside a stable Fréchet Inception Distance (FID) baseline of 178.59. The successful implementation of this processing pipeline into an interactive Streamlit ecosystem demonstrates the engineering viability of transitioning abstract neuro-visual concepts into practical software utilities. This architecture establishes a scalable baseline for future enhancements, including the integration of specialized artifact rejection layers such as Independent Component Analysis (ICA) or deep denoising autoencoders—prior to the Transformer stage to clean raw voltage arrays and lower the FID metric. Furthermore, processing latency can be optimized by transitioning to accelerated diffusion schedulers like Latent Consistency Models (LCMs) to enable continuous, instantaneous visual translation for live communication utilities. Finally, upgrading the input interface to high-density 64- or 128-channel medical EEG arrays will provide the network with richer spatial priors, mitigating localized distortions like latent blur. Ultimately, CognitiveCanvas demonstrates that clinical-grade, multi-million-dollar neuroimaging infrastructure is not a mandatory prerequisite for visual mind-reading, paving an affordable, portable path forward for assistive neuro-rehabilitation.

REFERENCES

[1] D. Vivancos, "MindBigData 2022: A large dataset of brain signals," arXiv preprint arXiv:2301.01234, 2023.
[2] N. Kumari, S. Anwar, and V. Bhattacharjee, "Convolutional neural network-based visually evoked EEG classification model on MindBigData," International Journal of Cognitive Computing, vol. 12, pp. 45-56, 2021.
[3] R. Mishra, K. Sharma, and A. Bhavsar, "Visual brain decoding for short duration EEG signals," IEEE Transactions on Cognitive and Developmental Systems, vol. 14, pp. 1102-1114, 2021.
[4] Y. Takagi and S. Nishimoto, "High-resolution image reconstruction from human brain activity using latent diffusion models," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 14453-14463.
[5] Anonymous, "Decoding EEG signals of visual brain representations with a CLIP based knowledge distillation," in International Conference on Learning Representations (ICLR) Conference Proceedings, 2024, pp. 23553-23565.
[6] R. Xiao et al., "Autoregressive visual decoding from EEG signals," arXiv preprint arXiv:2602.22555, 2026.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[7] J. Li et al., "Visual decoding and reconstruction via EEG embeddings with guided diffusion," Journal of Neural Engineering, vol. 22, pp. 1024-1035, 2025.
[8] C. Wang et al., "Reconstructing visual stimulus representation from EEG signals based on deep visual representation model," IEEE Transactions on Human-Machine Systems, vol. 54, no. 6, pp. 789-801, 2024.
[9] Z. Chen, J. Qing, T. Xiang, W. L. Yue, and J. H. Zhou, "Seeing beyond the brain: Conditional diffusion model with sparse masked Modelling for vision decoding," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 22710-22720.
[10] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 10684-10695.
[11] A. Radford et al., "Learning transferable visual models from natural language supervision," in International Conference on Machine Learning (ICML), 2021, pp. 8748-8763.
[12] A. Vaswani et al., "Attention is all you need," Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.
[13] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details