

## Generative AI for Brain-Computer Interfaces Decoding: A Systematic Review

**Authors:** Yi Guo, Shiqiang Ma, Shuqiang Wang, Shuqiang Wang

**Date:** 2025-08-14T00:00:00+00:00

### Abstract

Brain-computer interfaces (BCIs) establish a direct communication channel between the human brain and external devices, not only enabling a novel human-computer interaction paradigm but also providing an important technical foundation for innovative clinical treatments. BCI decoding methods can be divided into two major categories: discriminative and generative. Discriminative decoding models are primarily used for predefined tasks such as emotion recognition and motor intention classification, but exhibit limited performance in reconstructing high-dimensional semantic content such as natural language or visual images. In recent years, generative artificial intelligence (AI)—particularly the emergence of diffusion models and autoregressive Transformers—has rapidly propelled the transformation of BCI decoding paradigms, making generative decoding a research frontier. Generative decoding can directly reconstruct semantically rich content, including natural language and images, from neural activity, thereby overcoming the inherent limitations in expressive capability of traditional methods. Furthermore, generative AI has also played multiple roles in enhancing performance and generalization capabilities through data augmentation and assisting electrode material optimization. Nevertheless, BCI decoding based on generative AI still faces significant challenges, including limited decoding accuracy, insufficient cross-subject generalization capability, and algorithmic fairness issues arising from imbalanced BCI datasets. This paper systematically reviews the latest advances in generative AI-enabled BCI decoding, focusing on innovations in language decoding and visual decoding methods, and explores the diverse roles of generative AI in advancing the BCI decoding ecosystem, finally summarizing current limitations and promising future research directions.

## Full Text

### Preamble

#### Generative AI for Brain-Computer Interfaces Decoding: A Systematic Review

Yi Guo, Shiqiang Ma and Shuqiang Wang

**Abstract—** Brain-computer interfaces (BCIs) establish direct communication pathways between the human brain and external devices, enabling novel modes of human-machine interaction and providing essential technological foundations for innovative clinical therapies. BCI decoding encompasses discriminative and generative approaches. Discriminative decoding models are primarily used for predefined tasks such as emotion recognition and motor intention classification, but they exhibit limited performance in reconstructing high-dimensional semantic content such as natural language or visual images.

In recent years, the rapid advancement of generative artificial intelligence (AI)—particularly the emergence of diffusion models and autoregressive Transformers—has fundamentally reshaped the BCI decoding paradigm, propelling generative decoding to the forefront of research. Generative decoding enables the direct reconstruction of semantically rich content, including natural language and images, from neural activity, thereby overcoming the expressive limitations inherent to traditional approaches. Furthermore, generative AI offers multifaceted support for BCI decoding by enhancing data augmentation and assisting in electrode material optimization, further improving both performance and generalizability. Despite these advances, generative AI-driven BCI decoding still faces substantial challenges, including limited decoding accuracy, poor cross-subject generalization, and algorithmic fairness issues arising from imbalanced BCI datasets. Here, we systematically review recent progress in generative AI-powered BCI decoding, with a particular focus on methodological innovations in language and visual decoding. We further discuss the diverse roles of generative AI in advancing the BCI decoding ecosystem and highlight current limitations as well as promising future research directions.

**Index Terms—** brain-computer interface, generative artificial intelligence, brain decoding, language decoding, visual decoding

## I. Introduction

BCI technology, conceptualized in the 1970s, has always aimed to establish direct communication channels between the human brain and external devices, providing alternative interaction solutions for patients with motor impairments, neurodegenerative diseases, and language disorders. Traditional BCI systems primarily rely on discriminative models such as Support Vector Machines and Linear Discriminant Analysis to map neural signals like electroencephalograms (EEG) and electrocorticograms (ECoG) to limited action commands or prede-

defined categories (e.g., binary choices, motor imagery classification). While these systems have demonstrated some success in laboratory settings, their core limitation lies in their closed task structure and restricted semantic expression. Specifically, they can only recognize predefined discrete intentions and fail to capture the rich cognitive content that naturally emerges in the brain, such as continuous language flow, complex visual scenes, and emotional states. This limitation has resulted in approximately 30% of users being unable to effectively utilize traditional BCI systems due to “BCI illiteracy,” highlighting the significant gap between current neural decoding technology and the complexity of real brain cognition.

Throughout BCI technology development, decoding has remained the core challenge. To provide a clearer conceptual framework, we consider BCI decoding to comprise two fundamental paradigms: discriminative decoding, which maps brain signals to predefined categories, and generative decoding, which reconstructs semantically rich, open-ended content from neural activity. Traditional BCI decoding relies mostly on discriminative models focusing on classification tasks such as image classification and emotion recognition with predefined categories. In contrast, generative BCI decoding models aim to directly generate semantic-level content (such as text, speech, and images) from brain signals, characterized by open-ended task goals and strong expressive power. Generative decoding models can better capture the rich information of brain activity, achieving higher-level semantic expression and cognitive restoration.

This shift from discriminative to generative models represents not only a technological advancement but also a profound transformation of BCI decoding concepts. The emergence of generative decoding models enables BCI systems to move from simple command recognition to more complex semantic understanding and generation, providing the possibility for more natural and efficient human-computer interaction.

Generative AI models such as Variational Autoencoders (VAEs) and Diffusion Models are reshaping the BCI technology ecosystem at multiple levels. Algorithmically, generative models have expanded the boundaries of decoding tasks, enabling the generation of complex content such as natural language, speech, and images from brain signals, and can also be used in reverse to generate stimulating content, achieving bidirectional BCI. This bidirectional interaction capability brings new possibilities to BCI technology, allowing it to move beyond simple command transmission and enable more complex semantic communication. In terms of data, Generative AI has strong data augmentation capabilities, supporting cross-modal generation and virtual sample synthesis, providing data support for training high-performance BCI models. In terms of materials, generative models are being explored to assist in optimizing electrode materials and screen for new types of flexible materials, improving signal acquisition efficiency and user comfort.

These multidimensional applications not only expand the boundaries of BCI technology, making it more flexible and adaptable in practical applications,

but also show great potential in fields such as medical rehabilitation, human-computer collaboration, and neuroscience research.

In language decoding, generative language models have been used to restore sentences that individuals hear, speak, or think, breaking through the limitations of traditional command-based BCI in terms of vocabulary scale and expressive freedom. Application scenarios include assisting communication for people with language disorders and building silent communication systems. Research cases include speech synthesis driven by ECoG or EEG signals and neural translation from brain signals to text. These studies demonstrate the strong capabilities of Generative AI in BCI language decoding and provide new ideas for the rehabilitation of language disorder patients and the optimization of human-computer interaction. Through Generative AI, BCI systems can directly convert brain signals into natural language, enabling users to express their thoughts more naturally.

In visual decoding, BCI tasks model brain cortical visual signals (such as fMRI, EEG, MEG) and combine them with generative models to reconstruct images, videos, and even dreams and visual imagination. Related research has shown that generative models can effectively improve the spatial resolution and semantic consistency of decoded images, providing important technical support for cognitive neuroscience, visual prosthetics, and psychiatry.

Generative AI is not only one of the core models for BCI decoding but also provides systematic support for its development from peripheral dimensions. In terms of data augmentation and generalization capability improvement, given the scarcity of brain signal samples and strong heterogeneity among individuals, generative models support strategies such as cross-modal data generation, virtual brain signal synthesis, and brain signal super-resolution reconstruction. They play a key role in enhancing the diversity of training data and the generalization capability of models, especially showing strong adaptability in cross-task scenarios and multi-subject settings. This capability provides important support for the practical application of BCI systems, enabling them to achieve better performance with limited data and making them more feasible in real-world applications. Additionally, Generative AI can assist in BCI electrode design optimization, for example by simulating the electrical conductivity of electrode materials in different physiological environments and achieving low-invasive, high-density flexible electrode array design. Virtual screening and reverse design can significantly accelerate the conversion cycle of BCI hardware from prototype to final product, supporting the development of high-performance, user-friendly BCI systems.

Despite the broad application prospects of Generative AI in the BCI field, its development still faces many challenges. First, although invasive BCI has the advantage of high-quality signals, it is difficult to obtain samples and has high ethical costs. Non-invasive BCI, while easy to obtain data, has low signal-to-noise ratios and limited precision. Generative AI still needs to make further breakthroughs in improving data quality and representation capabilities, espe-

cially in characterizing complex brain states. Moreover, Generative AI-driven BCI models show significant performance differences across groups based on gender, age, and race, potentially harboring algorithmic bias risks. Current BCI datasets have structural deficiencies in individual and group representation, necessitating the building of diverse data resource libraries and introduction of fairness metrics and debiasing mechanisms to ensure ethical use of the technology. At the same time, the complexity and high computational resource demands of Generative AI models may limit their deployment in practical applications. These problems not only affect the widespread application of Generative AI in the BCI field but also pose challenges to the sustainable development of the technology.

[Figure 1: see original paper]

## II. Theoretical Foundations of Generative AI for BCI Decoding

Generative AI fundamentally differs from traditional discriminative models in its approach to modeling and synthesis. Rather than classifying neural signals into a set of predefined categories, generative models learn the underlying data distribution and enable open-ended generation of complex, high-dimensional content. This capability naturally aligns with the inherent structure of neural signals in the brain, which reflect distributed, continuous, and semantically rich cognitive states. Theoretically, this synergy arises because both brain representations and generative models—such as diffusion models, autoregressive transformers, VAEs, and generative adversarial networks (GANs)—excel at modeling structured, multi-modal, and non-linear manifolds. Such alignment moves beyond the closed-world assumption of traditional BCI decoding, paving the way for a more flexible and expressive interface between neural activity and external systems.

The integration of generative AI into BCI decoding establishes a paradigm shift from “decoding as classification” to “decoding as open-ended generation.” Within this paradigm, brain signals are treated not merely as feature vectors to be mapped to labels, but as samples from a high-dimensional latent space capable of supporting natural language, speech, and visual representations. Generative AI models enable this transformation via three inter-related mechanisms: (1) direct semantic reconstruction—mapping neural activity to continuous, interpretable outputs (e.g., language, images); (2) data augmentation and domain adaptation—synthesizing diverse and realistic neural data to mitigate training data scarcity and inter-individual variability; and (3) cross-modal neural modeling—translating across distinct brain signal modalities, such as EEG and fMRI, by learning shared latent representations that unify heterogeneous neural data. Together, these mechanisms offer a unified theoretical framework that supports more adaptive, scalable, and semantically meaningful BCIs.

In summary, the foundations of generative AI-BCI integration rest on a unique

theoretical compatibility: both domains exploit high-capacity, generative models to represent and reconstruct the richness of human cognition. This chapter introduces the essential theoretical constructs underlying this synergy, which will be further elaborated in subsequent sections—detailing recent algorithmic advances, application domains, and enabling technologies. This layered approach not only clarifies the mechanisms driving current progress but also sets the stage for addressing key challenges and realizing the full potential of generative AI-driven BCI systems.

### III. Generative AI for BCI Decoding Methods

#### A. From Discriminative Decoding to Generative Decoding for BCI

Traditional BCI decoding tasks have predominantly relied on discriminative models, which map neural signals to a fixed set of predefined categories via supervised classification frameworks. Representative applications include motor imagery recognition and emotional state classification, typically based on EEG or ECoG signals. Although such approaches deliver high accuracy for narrowly defined tasks, they inherently restrict the expressivity and generalization of BCIs to closed, limited output spaces.

In contrast, generative decoding establishes a fundamentally different paradigm: instead of classifying neural patterns, generative models reconstruct semantically rich and open-ended content—such as sentences, continuous speech, or images—directly from brain activity. Leveraging advances in generative artificial intelligence, these models are capable of synthesizing high-dimensional, contextually meaningful outputs that align more closely with the distributed and dynamic nature of human cognition. Recent studies have demonstrated the superiority of generative models over discriminative ones in reconstructing natural language, continuous speech, and visual imagery from neural data.

Overall, the shift from discriminative to generative decoding not only enhances the expressive range of BCI systems but also moves the field toward authentic semantic reconstruction and cognitive restoration.

#### B. BCI Language Decoding with Generative AI

Language decoding in BCI refers to the process of translating neural activity into interpretable linguistic outputs, typically in the form of natural language text or synthesized speech (Figure 2 [Figure 2: see original paper]). This capability is fundamental for restoring communication in individuals with severe paralysis or speech impairments. Traditional approaches have been constrained by predefined vocabularies or fixed linguistic categories, which limit expressive capacity and generalizability. Recent advances in generative AI, particularly the development of large language models, have fundamentally transformed the field by enabling direct mapping of complex brain signals to open-ended linguistic outputs, including continuous text generation and speech synthesis.

**1) BCI Text Decoding** Text decoding has emerged as a fundamental research direction in BCI language decoding, aiming to reconstruct natural language content directly from neural activity. Current studies span both invasive and non-invasive signal acquisition modalities.

Invasive BCIs typically leverage ECoG or microelectrode arrays (MEA) to record cortical neural activity. Early efforts focused predominantly on discriminative classification tasks, such as intent-driven spelling devices and word-pair classification. With advances in deep learning and the introduction of generative AI models, invasive approaches have gradually evolved from simple classification to open-vocabulary text reconstruction. Notably, several studies employing high-density ECoG arrays have achieved continuous speech-to-text reconstruction, significantly reducing word error rates and demonstrating real-time decoding of natural language. The adoption of representation learning and transformer-based generative models has further enabled large-vocabulary and sentence-level neural text decoding, substantially advancing the prospects of neural speech prostheses. Moreover, the transfer of motor-related decoding paradigms to language applications has yielded important breakthroughs, including real-time decoding of imagined handwriting signals, highlighting the stability of neural representations and the adaptive capabilities of current models. Advanced neural prosthetic systems now enable multimodal synchronous control of text, speech audio, and even virtual avatar facial animations.

In recent years, non-invasive BCI text decoding has also witnessed rapid progress, primarily utilizing fMRI, MEG, and EEG. Early research in this domain mainly addressed the discriminative classification of basic semantic categories, analyzing neural responses to different linguistic stimuli and revealing the distributed nature of semantic representations in the brain. Subsequently, decoding efforts expanded from coarse semantic classes to individual words and phrases, marking initial steps toward reconstructing continuous natural language semantics.

With the advent of generative AI—especially large language models (LLMs) such as transformers, BART, and GPT—non-invasive text decoding capabilities have achieved unprecedented improvements, including the first demonstrations of open-vocabulary, continuous text generation from brain signals. To address challenges posed by the low signal-to-noise ratio and substantial inter-subject variability of EEG data, recent work has explored multimodal fusion, cross-subject representation alignment, and self-supervised contrastive learning, yielding notable gains in open-vocabulary EEG text decoding. In parallel, strategies involving cross-modal semantic encoding and joint training with language models have effectively bridged representation gaps, enhancing both the semantic consistency and generalizability of decoded text. For fMRI and MEG, end-to-end natural text generation frameworks have enabled direct mapping from brain signals to text embedding spaces, resulting in marked improvements in decoding accuracy and semantic relevance. In particular, the combination of MEG's high temporal and spatial resolution with transformer models has achieved real-

time, sentence-level continuous decoding, highlighting the considerable practical potential of non-invasive BCIs.

More recently, pioneering studies have explored multilingual decoding as well as reconstructing semantically relevant textual descriptions of images from neural responses to visual stimuli, expanding the application scenarios and task boundaries of BCI text decoding. End-to-end LLM-based frameworks demonstrate superior performance in multimodal representation fusion and generalization tasks, advancing BCI technology toward more natural and universal communication interfaces. Additionally, integration of audiovisual semantic fusion and multi-task joint decoding strategies has broadened the applicability of BCI text decoding to a wider range of real-world scenarios.

**2) BCI Audio Decoding** In recent years, audio decoding via BCIs has become a key technology for restoring communication in individuals with neurological impairments. Compared to text decoding, audio decoding can directly generate continuous, natural speech signals, offering greater expressive richness and enhanced potential for real-time interaction. Research in this domain spans both invasive and non-invasive recording modalities, leveraging advanced generative AI techniques to reconstruct clear and expressive speech content from neural activity.

Invasive approaches have achieved substantial progress in audio decoding. Early studies used neural electrode arrays to record cortical activity and employed linear regression or classical algorithms to decode and synthesize vowels and basic speech units. With the introduction of deep learning, including convolutional and recurrent neural networks, speech quality has markedly improved. The integration of transformer-based encoders with neural vocoders has enabled accurate reconstruction of natural spoken sentences directly from ECoG signals, and transformer models have demonstrated potential for transfer from overt to imagined speech decoding tasks.

Generative AI-based end-to-end audio decoding has recently demonstrated even greater performance, particularly excelling in real-time generation and naturalness of speech. Studies employing transformer and recurrent models have enabled continuous, real-time synthesis of speech that preserves speaker-specific vocal features. Recent neural prosthetic systems based on stereo-electroencephalography (SEEG) or ECoG have demonstrated real-time speech synthesis for restoring vocal communication, with additional capabilities including prosodic modulation and—most notably—the restoration of singing in a pioneering study. These approaches have shown clinical stability and practicality in long-term implant trials. Moreover, subcortical structures such as the thalamus have been found to contain abundant language-related information, offering prospects for further improvements in decoding performance and underscoring the importance of both cortical and subcortical regions in audio generation.

In the non-invasive setting, research has focused on extracting neural information from scalp EEG to reconstruct speech signals. Initial work centered on reconstructing speech envelopes and basic acoustic features, but the application of deep models such as convolutional and transformer architectures has greatly enhanced speech quality and intelligibility. Nonlinear modeling methods have shown clear advantages over traditional linear approaches and demonstrated strong generalization capacity. Recent advances in multiscale fusion networks and state space models have enabled efficient processing of long EEG sequences for the reconstruction of continuous speech spectrograms.

The latest studies have further introduced generative models for non-invasive EEG-based audio reconstruction, achieving superior alignment between EEG and audio representations with pre-trained text-to-speech models and generating high-quality, natural speech. In addition, convolutional network architectures have enabled real-time online reconstruction of continuous, natural sentences. These advances collectively demonstrate that non-invasive EEG signals can support the generation of complex, high-quality speech, providing a robust foundation for practical, portable, and cost-effective BCIs.

Overall, the rapid development of audio decoding in BCIs is driven by deep learning, and especially by the introduction of generative AI. Both invasive and non-invasive research lines have demonstrated significant clinical promise, substantially broadening communication options for patients with neurological injuries and advancing BCI technology toward more natural, real-time, and emotionally expressive human-machine interaction.

### C. BCI Visual Decoding with Generative AI

Recent advances in visual decoding leverage generative AI techniques to substantially enhance reconstruction fidelity and semantic consistency. This progress spans multiple domains, including static image decoding, dynamic video decoding, and multimodal visual decoding, collectively facilitating richer, more accurate visual content reconstruction from brain signals (Figure 3 [Figure 3: see original paper]).

**1) BCI Image Decoding** The main goal of image decoding is to reconstruct static image content from brain signals, representing an individual's visual experience at a particular moment. This typically involves the reconstruction of simple or complex static images. Image decoding can help visually impaired individuals reconstruct static images such as objects or scenes that they imagine, and can also be employed to study human visual perception and cognitive processes, particularly the brain activity patterns when processing static visual information. In addition, it can generate high-resolution medical images to assist doctors in diagnosis and help artists create imaginative images through brain signals, thereby enhancing their creative efficiency and outcomes.

Image decoding is a significant research direction in the field of BCI, with its

development history tracing back to early attempts at processing visual signals. Early attempts included reconstructing simple visual images from brain activity signals (e.g., fMRI, EEG, MEG). These studies primarily focused on decoding simple visual stimuli such as black-and-white patterns, handwritten digits, and human faces, laying the groundwork for subsequent decoding of more complex visual content. By analyzing patterns of brain activity, it was possible to identify the general shapes and positions of the simple patterns being viewed.

With the rapid development of deep learning technologies, the field of image decoding has witnessed remarkable progress. Deep learning-based methods possess a more powerful capacity for feature capturing and integration. They can learn to map neural activities in the visual cortex to the visual stimuli observed during the perceptual process through joint training of generation and adversarial learning. By training a deep generative adversarial network and combining it with a perceptual loss term, the model is able to reconstruct images that are highly similar to the original stimulus images. Due to the difficulty in obtaining paired brain signal and visual stimulus image training samples, many methods apply semi-supervised, self-supervised, and unsupervised techniques to the pre-training process of generative models to enhance their visual representation capabilities.

In recent years, deep generative models, particularly Generative Adversarial Networks (GANs) and diffusion models, have emerged as key technologies for decoding brain activity and reconstructing perceived images. The primary goal of these methods is to transform complex brain signals, such as EEG and fMRI, into understandable visual outputs, thereby revealing the mechanisms by which the brain encodes cognitive content. GANs have demonstrated their ability to generate corresponding images from brain signals (e.g., thoughts of numbers, characters, or objects), providing strong evidence that brain signals encode visual cognitive content. For EEG signals, researchers have developed more advanced GAN frameworks. For instance, combining stacked LSTM to learn compact and noise-free representations of EEG data has been used to generate specific visual stimuli. Specialized architectures, such as the Dual-Condition and Lateralization-Supported GAN (DCLS-GAN), have further enhanced the ability to visualize image-evoked thoughts from EEG signals. In the fMRI domain, GANs have been successfully applied to address inter-subject heterogeneity by reconstructing visual images from fMRI data acquired while participants perceive images, significantly improving accuracy. By integrating techniques such as Variational Autoencoders (VAEs), GANs have achieved robust decoding from fMRI data and even accomplished tasks like gender classification. GANs that introduce semantic feature similarity as a condition (e.g., Similarity-Conditioned GAN - SC-GAN) have significantly increased the similarity between reconstructed images and original stimuli. Collectively, these works highlight the substantial potential of GANs in enhancing the precision and interpretability of brain decoding.

Diffusion models, with their powerful generative capabilities and flexibility, have

shown remarkable performance in brain decoding, especially in fMRI image reconstruction. Some methods have designed multi-stage processes, such as first aligning semantics and then generating details, or combining multiple generative models (e.g., masked autoencoders + latent diffusion models, or VDVAE + BLIP + universal diffusion models) to hierarchically recover visual and semantic details. Other approaches have employed data-driven methods to synthesize images that activate specific brain regions; improved classification and retrieval performance by linearly mapping to semantic features; and achieved state-of-the-art performance in image retrieval and reconstruction tasks by integrating contrastive learning and diffusion priors. In addition, to alleviate the problem of insufficient training data, some methods have adopted self-supervised learning (especially masked modeling) to learn robust and information-rich fMRI representations. Diffusion models have also been applied to EEG signal decoding. By leveraging the strong transferability of pre-trained text-to-image diffusion models and combining techniques such as temporal masked signal modeling to address signal noise and individual variability, high-quality image generation from EEG has been achieved. Other diffusion-based frameworks have further improved reconstruction quality by optimizing internal mechanisms and introducing edge estimation strategies.

Furthermore, the further development of multimodal data fusion technology is also an important direction for the future. By combining more modalities of data (such as EEG, MEG, etc.), the precision and robustness of image decoding can be further improved. At the same time, refining cross-subject decoding technology is also a focus of future research. Developing cross-subject decoding models with strong generalizability will help promote the widespread application of image decoding technology. Multimodal data joint decoding refers to the integration of data from different modalities (such as brain signals, images, text, and speech) to enhance the precision and robustness of decoding. This approach aims to provide a more comprehensive understanding of brain activity and cognitive processes by leveraging complementary information from multiple data sources. For instance, combining fMRI signals with image data can reconstruct visual content; integrating EEG signals with text descriptions can decode emotional states; and merging MEG signals with speech data can help study language processing. The combination of these modalities provides a richer set of information, thereby enhancing the accuracy and reliability of decoding. By combining brain signals with other modalities (e.g., text descriptions, clinical information), richer information can be provided for visual decoding, which better constrains the generated visual content and thereby helps the generative models reconstruct high-quality images from brain activity. Feature disentanglement and semantic representation are crucial for interpreting neural responses and have played an important role in shaping the high-level neural representations underlying visual perception. In the field of emotion decoding, multi-view multi-label models predict multiple emotional states through fine-grained decoding, breaking through the previous limitation of decoding only single emotional categories.

Finally, with the continuous development of BCI technology, image decoding is expected to play a greater role in fields such as medical rehabilitation and human-computer collaboration. For instance, BCI image decoding technology can help visually impaired individuals reconstruct images or scenes from their imagination and improve their quality of life. Image decoding technology can also be applied in virtual reality and augmented reality fields to provide users with more immersive experiences.

**2) BCI Video Decoding** Brain-evoked signals encode not only static images but also temporally continuous dynamic scenes. Effectively extracting and decoding these temporal representations promises new insights into the dynamic mechanisms of the human visual system and will advance the next generation of brain-computer video interfaces (BCVIs). One of the most challenging tasks in this domain is video information decoding, whose goal is to classify perceived dynamic stimuli or to reconstruct individual frames/short clips from fMRI-measured brain activity.

In recent years, researchers have made significant progress in this field, proposing a variety of innovative methods to improve the quality and efficiency of video reconstruction. For example, NeuroCine introduced a new dual-phase framework that enhances fMRI representation through spatial masking and temporal interpolation, and uses a diffusion model augmented by correlated prior noise for video generation. NeuralFlix employs spatial and temporal enhancements to learn fMRI representations and utilizes a diffusion model enhanced by prior noise to generate videos. Additionally, Brain Netflix accurately generates 2- and 3-second video clips from brain activities of different participants and datasets through multi-dataset and multi-subject training. These studies demonstrate the potential for reconstructing dynamic visual experiences from fMRI signals and provide an important technical foundation for the development of future BCVIs.

Recently, three-dimensional convolutional networks (3D-CNNs) and two-stream spatiotemporal networks have been shown to generate hierarchical spatiotemporal features that bridge dynamic visual stimuli and brain signals. To mitigate the scarcity of paired training data, researchers first pre-train a 3D-CNN on large-scale video datasets such as Kinetics-400 and Moments-in-Time, and then use the resulting spatiotemporal features as the decoding target for reconstructing natural video clips.

In recent years, the rapid advancement of artificial intelligence has spurred explosive growth in methods that decode brain signals to generate videos. Diffusion models, owing to their powerful generative capabilities, have been widely adopted for brain-based video decoding tasks. One line of work employs a cascaded diffusion backbone conditioned on text prompts to synthesize high-frame-rate, high-fidelity videos from fMRI, adding motion-consistency regularizers that markedly reduce temporal jitter. NeuroClips inserts spatiotemporal Transformer blocks into the diffusion trunk, using local-global attention

for smooth frame transitions and enabling minute-long sequence reconstruction. Others couple diffusion priors with adversarial discriminators and cross-frame contrastive losses to refine dynamic details and achieve the first faithful recovery of natural motion scenes.

To harness richer multimodal cues, certain methods introduce text-image-video tri-modal alignment, allowing users to steer reconstruction via language instructions. To mitigate label scarcity, self-supervised diffusion pipelines have been proposed that reconstruct long videos from fMRI without paired annotations. EEG2Video pioneers joint training of noisy EEG signals with text-conditioned diffusion, achieving  $512 \times 512$  dynamic-scene decoding through cross-modal alignment and laying the groundwork for portable brain-video interfaces.

Compared to images, videos contain richer information and more complex scene transitions, which pose significant challenges for decoding extensive visual stimuli from brain signals. Some studies have attempted to reconstruct high-quality video content from brain activity by integrating brain signals with other semantic information. Multimodal data joint decoding offers significant advantages for brain decoding research. First, by integrating data from different modalities, it provides a more comprehensive view of brain activity and enables more accurate capture of complex brain activity patterns. Second, multimodal decoding enhances the robustness and accuracy of decoding, with the complementary nature of different modality data compensating for the shortcomings of single-modality data. Moreover, multimodal decoding can reveal the intrinsic connections between different modality data, helping researchers better understand how the brain processes and integrates information from different senses. This integrative approach offers new perspectives and tools for in-depth research into the neural mechanisms of the brain. These studies have further optimized the performance of multimodal decoding and provided technical support for real-time applications of video decoding.

Despite its great potential, multimodal data joint decoding for video decoding tasks still faces many challenges. Firstly, feature extraction and alignment of data from different modalities is a complex issue, necessitating the development of more advanced algorithms for effective data integration. Secondly, the complexity of multimodal decoding models is relatively high, with increased computational costs and data requirements, which limit their widespread application. Moreover, real-time capability is a key issue, as current decoding methods mostly rely on offline processing, failing to meet the needs of real-time interaction.

#### IV. Generative AI Empowers BCI Decoding Through Data Augmentation and Sensor Optimization

Beyond its direct use in BCI decoding algorithms, generative AI is also enhancing BCI decoding by improving upstream components, including data augmentation and support for sensor optimization (Figure 4 [Figure 4: see original

paper]). The following sections detail these two directions.

### **A. Generative AI-Driven Data Augmentation for BCI Decoding**

As summarized in Table I, each neural recording modality used in BCI systems presents specific strengths and limitations that directly influence decoding performance. Non-invasive techniques such as EEG and fNIRS are portable and cost-effective but have limited spatial resolution and are sensitive to artifacts. MEG offers high temporal precision yet requires a magnetically shielded environment and is non-portable, while fMRI provides excellent spatial detail but has low temporal resolution and limited applicability in real-time settings. Invasive methods including ECoG, SEEG, and MEA achieve high signal quality and access to deep brain structures but involve substantial surgical risks, high procedural costs, and restricted data availability. These trade-offs constrain the scalability, robustness, and generalizability of BCI decoding.

Generative modeling techniques are increasingly applied to address these constraints by enhancing data quality and diversity. Neural signal super-resolution methods reconstruct high-resolution signals from lower-resolution recordings, improving both spatial and temporal fidelity in modalities such as EEG and MEG. Cross-modal generation enables translation between modalities, for example between EEG, fMRI, and SEEG, supporting multimodal integration and improving the generalizability of hybrid systems. Synthetic data generation expands available datasets by creating realistic neural activity patterns, increasing variability across subjects and tasks, and reducing overfitting in data-limited scenarios. Together, these approaches provide complementary means to overcome modality-specific limitations and support more accurate and generalizable BCI decoding.

### **B. Generative AI-Driven Sensor Optimization for BCI Decoding**

Beyond data-centric strategies, generative models are also being applied to the design and optimization of neural interface hardware. By accelerating the in silico discovery of candidate materials and electrode architectures, these approaches support the development of next-generation sensors with improved biocompatibility, conductivity, and signal fidelity. Integration of computational design with experimental validation is enabling faster iteration cycles in BCI hardware development, potentially narrowing the performance gap between invasive and non-invasive systems.

## **V. Challenges and Future Directions in Generative BCI Decoding**

Despite the transformative potential of generative AI in expanding the capabilities of BCI decoding, several critical challenges remain unresolved in practice. Chief among these are limited decoding accuracy, poor cross-subject

generalizability of BCI models, and fairness concerns arising from the under-representation of diverse populations in training data. This section examines each challenge and outlines potential directions for future research. Figure 5 [Figure 5: see original paper] conceptually summarizes these core challenges and generative AI-based directions.

### A. BCI Decoding Accuracy Constraints

BCI decoding accuracy is fundamentally constrained by the signal acquisition modality. Invasive BCIs offer high signal fidelity and enhanced decoding precision, yet they are limited by ethical concerns and inherent risks of surgical intervention. In contrast, non-invasive BCIs, which are more scalable and ethically acceptable, face significant challenges in signal quality due to low signal-to-noise ratios inherent in external neural recordings. To address these challenges, recent advances in generative AI show promise in improving the quality of non-invasive signals, with the potential to synthesize high-fidelity neural representations. These innovations aim to reduce the performance disparity between invasive and non-invasive systems, broadening the clinical applicability of BCIs while avoiding the ethical and practical limitations of surgical procedures.

To overcome the limitations of non-invasive BCI systems, generative AI presents a promising solution. By integrating multimodal data and utilizing cross-modal generation techniques, generative models can enhance the quality of non-invasive signals, even synthesizing high-fidelity neural representations akin to those obtained from invasive methods. This approach holds the potential to bridge the performance gap between invasive and non-invasive systems, enabling effective BCI use without invasive procedures.

However, current generative methods still face challenges in modeling the complex neural states associated with high-level decoding tasks, such as language and image reconstruction. To achieve more accurate and versatile decoding, the development of more advanced generative architectures is essential—architectures that can capture the high-dimensional and intricate dynamics of brain activity. Overcoming these challenges is crucial for expanding the scope of cognitive functions that BCIs can decode, thereby enhancing their practical applicability in real-world scenarios.

### B. BCI Cross-Subject Generalization Challenges

Limited generalizability across individuals remains a major bottleneck for widespread BCI deployment. Inter-subject variability in anatomical structures and neural activation patterns often leads to substantial performance drops when decoding models are applied to new users or across different experimental settings. Generative AI holds considerable promise in addressing this challenge by learning to model latent data distributions and extracting invariant representations that generalize across subjects.

In practical terms, generative models offer the ability to learn abstract representations of neural data that emphasize patterns consistent across individuals while reducing the impact of subject-specific variability. By capturing the underlying structure of brain signals, these models can help filter out irrelevant differences—such as signal drift or individual physiological traits—and retain information more directly related to the decoding task. This approach is increasingly recognized as a promising direction for improving cross-subject generalization of BCI systems.

Nevertheless, despite progress enabled by generative AI, current models still fall short of achieving robust cross-subject generalization at scale. Most approaches remain limited in their adaptability to heterogeneous user populations and often require extensive subject-specific calibration. To bridge this gap, future research should prioritize the development of generative frameworks that integrate few-shot learning, domain adaptation, and personalized pretraining strategies. Such advancements are essential for building scalable and reliable BCI systems capable of operating effectively across diverse individuals and real-world scenarios.

### C. Fairness Challenges in BCI Decoding

Extensive studies in medical AI have highlighted that imbalances in demographic representation, inconsistencies in data collection protocols, and systematic underrepresentation of certain groups can lead to uneven model performance across populations. Similar concerns are increasingly evident in BCI decoding. Since neural data collection relies on specific participants and experimental paradigms, structural differences in neural signal expression across gender, age, and ethnicity—combined with demographic biases in data acquisition—can result in performance disparities. These disparities are particularly pronounced in underrepresented groups, where decoding accuracy often drops, raising critical concerns about the fairness and inclusiveness of BCI systems. As such, data imbalance has become a key barrier to the scalable and equitable deployment of BCI technologies.

Looking ahead, generative AI offers a novel avenue for mitigating fairness-related challenges in BCI. By modeling latent distributions from limited data and generating diverse, high-quality synthetic neural signals, generative models can help fill gaps in data coverage for marginalized populations without imposing additional burdens on neural data acquisition. This may contribute to more demographically balanced training datasets and, in turn, more equitable decoding performance across user groups. Such approaches have been shown to improve model fairness by enhancing data representation for underrepresented groups. Nevertheless, the safety, controllability, and fidelity of generated data remain active areas of concern.

Future efforts should promote multi-center collaboration and explore paradigms such as federated learning combined with generative modeling to foster a more inclusive, reliable, and socially responsible BCI ecosystem—one that ensures eq-

uitable access to intelligent interaction technologies across diverse populations.

## VI. Conclusions

The rapid development of Generative AI is reshaping the core capabilities of BCI systems. By integrating deeply with BCI technology, generative models have made significant progress in decoding paradigms, data ecosystems, and material hardware. They have enhanced BCI systems' cognitive restoration capabilities, generalization, and usability. However, real-world deployment faces challenges in data acquisition, generalization, and fairness. Future development of generative BCI requires multidisciplinary collaboration to achieve widespread application in medical rehabilitation, human-computer collaboration, and neuroscience. Experts from computer science, neuroscience, and materials science must work together to optimize algorithms, understand brain signals, and innovate hardware. Policymakers, ethicists, and sociologists should also be involved to ensure sustainable development and social acceptance. Only through joint efforts can generative BCI technology be sustainably developed and widely accepted, contributing more to human welfare.

## References

- [1] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Brain-computer interfaces for communication and rehabilitation," *Nature Reviews Neurology*, vol. 12, no. 9, pp. 513–525, 2016.
- [2] M. A. Schwemmer, N. D. Skomrock, P. B. Sederberg, J. E. Ting, G. Sharma, M. A. Bockbrader, and D. A. Friedenberg, "Meeting brain-computer interface user performance expectations using a deep neural network decoding framework," *Nature medicine*, vol. 24, no. 11, pp. 1669–1676, 2018.
- [3] A. D. Degenhart, W. E. Bishop, E. R. Oby, E. C. Tyler-Kabara, S. M. Chase, A. P. Batista, and B. M. Yu, "Stabilization of a brain-computer interface via the alignment of low-dimensional spaces of neural activity," *Nature biomedical engineering*, vol. 4, no. 7, pp. 672–685, 2020.
- [4] Y. Ding, C. Udompanyawit, Y. Zhang, and B. He, "EEG-based brain-computer interface enables real-time robotic hand control at individual finger level," *Nature communications*, vol. 16, no. 1, pp. 1–20, 2025.
- [5] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system," *IEEE Transactions on biomedical engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.
- [6] K. T. Littlejohn, C. J. Cho, J. R. Liu, A. B. Silva, B. Yu, V. R. Anderson, C. M. Kurtz-Miott, S. Brosler, A. P. Kashyap, and I. P. Hallinan, "A streaming brain-to-voice neuroprosthesis to restore naturalistic communication," *Nature neuroscience*, vol. 28, no. 4, pp. 902–912, 2025.

- [7] G. Santhanam, S. I. Ryu, B. M. Yu, A. Afshar, and K. V. Shenoy, "A high-performance brain-computer interface," *Nature*, vol. 442, no. 7099, pp. 195–198, 2006.
- [8] M. Wairagkar, N. S. Card, T. Singer-Clark, X. Hou, C. Iacobacci, L. M. Miller, L. R. Hochberg, D. M. Brandman, and S. D. Stavisky, "An instantaneous voice-synthesis neuroprosthesis," *Nature*, p. 1–8, 2025.
- [9] D. Huang, K. Qian, D.-Y. Fei, W. Jia, X. Chen, and O. Bai, "Electroencephalography (EEG)-based brain-computer interface (BCI): A 2-D virtual wheelchair control based on event-related desynchronization/synchronization and state control," *IEEE transactions on Neural Systems and Rehabilitation engineering*, vol. 20, no. 3, pp. 379–388, 2012.
- [10] D. B. Silversmith, R. Abiri, N. F. Hardy, N. Natraj, A. Tu-Chan, E. F. Chang, and K. Ganguly, "Plug-and-play control of a brain-computer interface through neural map stabilization," *Nature biotechnology*, vol. 39, no. 3, pp. 326–335, 2021.
- [11] Y. Zhong, Y. Wang, D. Farina, and L. Yao, "A closed-loop tactile stimulation training protocol for motor imagery-based BCI: Boosting BCI performance for BCI-deficiency users," *IEEE Transactions on Biomedical Engineering*, pp. 1–11, 2025.
- [12] T. Horikawa and Y. Kamitani, "Generic decoding of seen and imagined objects using hierarchical visual features," *Nature communications*, vol. 8, no. 1, p. 15037, 2017.
- [13] A. Fares, S.-h. Zhong, and J. Jiang, "EEG-based image classification via a region-level stacked bi-directional deep learning framework," *BMC medical informatics and decision making*, vol. 19, no. Suppl 6, p. 268, 2019.
- [14] P. Mukherjee, A. Das, A. K. Bhunia, and P. P. Roy, "Cogni-net: Cognitive feature learning through deep visual perception," in *2019 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2019, p. 4539–4543.
- [15] Y. Song, Y. Wang, H. He, and X. Gao, "Recognizing natural images from EEG with language-guided contrastive learning," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.
- [16] P. Patel, S. B, and R. N. Annavarapu, "Application of supervised machine learning models in human emotion classification using Tsallis entropy as a feature," *Journal of Big Data*, vol. 12, no. 1, p. 126, 2025.
- [17] K. Fu, C. Du, S. Wang, and H. He, "Multi-view multi-label fine-grained emotion decoding from human brain activity," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 7, p. 9026–9040, 2022.
- [18] Z. Huang, C. Du, C. Li, K. Fu, and H. He, "Identifying the hierarchical emotional areas in the human brain through information fusion," *Information Fusion*, vol. 113, p. 102613, 2024.

- [19] M. Jin, C. Du, H. He, T. Cai, and J. Li, "PGCN: Pyramidal graph convolutional network for EEG emotion recognition," *IEEE Transactions on Multimedia*, vol. 26, p. 9070–9082, 2024.
- [20] S. Huang, Y. Wang, and H. Luo, "CCSUMSP: A cross-subject Chinese speech decoding framework with unified topology and multi-modal semantic pre-training," *Information Fusion*, vol. 119, p. 103022, 2025.
- [21] X. Zhao, J. Sun, S. Wang, J. Ye, X. Zhang, and C. Zong, "Mapguide: A simple yet effective method to reconstruct continuous language from brain activities," in *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2024*, 2024, p. 3822–3832.
- [22] C. de la Torre-Ortiz, M. M. Spapé, N. Ravaja, and T. Ruotsalo, "Cross-subject EEG feedback for implicit image generation," *IEEE transactions on cybernetics*, vol. 54, no. 10, p. 6105–6117, 2024.
- [23] L. Meng and C. Yang, "Dual-guided brain diffusion model: image reconstruction from human visual stimulus fMRI," *Bioengineering*, vol. 10, no. 10, p. 1117, 2023.
- [24] O. Ozdenizci, Y. Wang, T. Koike-Akino, and D. Erdogmus, "Transfer learning in brain-computer interfaces with adversarial variational autoencoders," in *9th International IEEE EMBS Conference on Neural Engineering, NER 2019*, March 20–23, 2019, IEEE Computer Society, 2019, p. 207–210.
- [25] P. Scotti, A. Banerjee, J. Goode, S. Shabalin, A. Nguyen, A. Dempster, N. Verlinde, E. Yundler, D. Weisberg, and K. Norman, "Reconstructing the mind's eye: fMRI-to-image with contrastive learning and diffusion priors," *Advances in Neural Information Processing Systems*, vol. 36, p. 24705–24728, 2023.
- [26] J.-H. Jeong, J.-H. Cho, B.-H. Lee, and S.-W. Lee, "Real-time deep neurolinguistic learning enhances noninvasive neural language decoding for brain-machine interaction," *IEEE Transactions on Cybernetics*, vol. 53, no. 12, p. 7469–7482, 2023.
- [27] R. Antonello, A. Vaidya, and A. Huth, "Scaling laws for language encoding models in fMRI," *Advances in Neural Information Processing Systems*, vol. 36, p. 21895–21907, 2023.
- [28] Z. Ye, Q. Ai, Y. Liu, M. de Rijke, M. Zhang, C. Lioma, and T. Ruotsalo, "Generative language reconstruction from brain recordings," *Communications Biology*, vol. 8, no. 1, p. 346, 2025.
- [29] S. L. Metzger, K. T. Littlejohn, A. B. Silva, D. A. Moses, M. P. Seaton, R. Wang, M. E. Dougherty, J. R. Liu, P. Wu, and M. A. Berger, "A high-performance neuroprosthesis for speech decoding and avatar control," *Nature*, vol. 620, no. 7976, p. 1037–1046, 2023.

- [30] Y. Liu, Z. Zhao, M. Xu, H. Yu, Y. Zhu, J. Zhang, L. Bu, X. Zhang, J. Lu, and Y. Li, "Decoding and synthesizing tonal language speech from brain activity," *Science Advances*, vol. 9, no. 23, p. eadh0478, 2023.
- [31] Z. Tian, R. Quan, F. Ma, K. Zhan, and Y. Yang, "BRAIN-GUARD: Privacy-preserving multisubject image reconstructions from brain activities," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, 2025, p. 14414-14422.
- [32] Y. Song, B. Liu, X. Li, N. Shi, Y. Wang, and X. Gao, "Decoding natural images from EEG for object recognition," in *12th International Conference on Learning Representations, ICLR 2024*, 2024.
- [33] W. Xia, R. De Charette, C. Oztireli, and J.-H. Xue, "Dream: Visual decoding from reversing human visual system," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, p. 8226-8235.
- [34] S. Kumar, T. R. Sumers, T. Yamakoshi, A. Goldstein, U. Hasson, K. A. Norman, T. L. Griffiths, R. D. Hawkins, and S. A. Nastase, "Shared functional specialization in transformer-based language models and the human brain," *Nature communications*, vol. 15, no. 1, p. 5523, 2024.
- [35] M. Ferrante, T. Boccato, F. Ozelik, R. VanRullen, and N. Toschi, "Multimodal decoding of human brain activity into images and text," in *Proceedings of Machine Learning Research*, vol. 243, 2023, pp. 11-26.
- [36] Q. Liu, H. Zhu, N. Chen, B. Huang, W. Lu, and Y. Wang, "Mind-bridge: reconstructing visual images based on diffusion model from human brain activity," *Signal, Image and Video Processing*, vol. 18, no. Suppl 1, p. 953-963, 2024.
- [37] W. Ko, E. Jeon, J. S. Yoon, and H.-I. Suk, "Semi-supervised generative and discriminative adversarial learning for motor imagery-based brain-computer interface," *Scientific reports*, vol. 12, no. 1, p. 4587, 2022.
- [38] S. Wang, T. Zhou, Y. Shen, Y. Li, G. Huang, and Y. Hu, "Generative ai enables EEG super-resolution via spatio-temporal adaptive diffusion learning," *IEEE Transactions on Consumer Electronics*, 2025.
- [39] J. Vetter, J. H. Macke, and R. Gao, "Generating realistic neurophysiological time series with denoising diffusion probabilistic models," *Patterns*, vol. 5, no. 9, 2024.
- [40] J. Kwon and C.-H. Im, "Novel signal-to-signal translation method based on StarGAN to generate artificial EEG for SSVEP-based brain-computer interfaces," *Expert Systems with Applications*, vol. 203, p. 117574, 2022.
- [41] W. Yao, Z. Lyu, M. Mahmud, N. Zhong, B. Lei, and S. Wang, "CATD: Unified representation learning for EEG-to-fMRI cross-modal generation," *IEEE Transactions on Medical Imaging*, 2025.

- [42] Y. Li, Y. Wang, B. Lei, and S. Wang, "SCDM: Unified representation learning for EEG-to-fNIRS cross-modal generation in MI-BCIs," *IEEE Transactions on Medical Imaging*, 2025.
- [43] M. Hu, J. Chen, S. Jiang, W. Ji, S. Mei, L. Chen, and X. Wang, "E2SGAN: EEG-to-SEEG translation with generative adversarial networks," *Frontiers in Neuroscience*, vol. 16, p. 971829, 2022.
- [44] A. Antoniadou, L. Spyrou, D. Martin-Lopez, A. Valentin, G. Alarcon, S. Sanei, and C. C. Took, "Deep neural architectures for mapping scalp to intracranial EEG," *International journal of neural systems*, vol. 28, no. 08, p. 1850009, 2018.
- [45] R. Zhang, Y. Zeng, L. Tong, J. Shu, R. Lu, K. Yang, Z. Li, and B. Yan, "ERP-WGAN: A data augmentation method for EEG single-trial detection," *Journal of Neuroscience Methods*, vol. 376, p. 109621, 2022.
- [46] F. Fahimi, S. Dosen, K. K. Ang, N. Mrachacz-Kersting, and C. Guan, "Generative adversarial networks-based data augmentation for brain-computer interface," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 9, p. 4039–4051, 2020.
- [47] Y. Tang, D. Chen, H. Liu, C. Cai, and X. Li, "Deep EEG super-resolution via correlating brain structural and functional connectivities," *IEEE Transactions on Cybernetics*, vol. 53, no. 7, p. 4410–4422, 2022.
- [48] K. G. Hartmann, R. T. Schirrmeyer, and T. Ball, "EEG-GAN: Generative adversarial networks for electroencephalographic (EEG) brain signals," Preprint at arXiv, 2018, <https://doi.org/10.48550/arXiv.1806.01875>.
- [49] A. Gayon-Lombardo, L. Mosser, N. P. Brandon, and S. J. Cooper, "Pores for thought: generative adversarial networks for stochastic reconstruction of 3D multi-phase electrode microstructures with periodic boundaries," *npj Computational Materials*, vol. 6, no. 1, p. 82, 2020.
- [50] B. Sanchez-Lengeling and A. Aspuru-Guzik, "Inverse molecular design using machine learning: Generative models for matter engineering," *Science*, vol. 361, no. 6400, p. 360–365, 2018.
- [51] M. Manica, J. Born, J. Cadow, D. Christofidellis, A. Dave, D. Clarke, Y. G. N. Teukam, G. Giannone, S. C. Hoffman, and M. Buchan, "Accelerating material design with the generative toolkit for scientific discovery," *npj Computational Materials*, vol. 9, no. 1, p. 69, 2023.
- [52] Y. Dan, Y. Zhao, X. Li, S. Li, M. Hu, and J. Hu, "Generative adversarial networks (GAN) based efficient sampling of chemical composition space for inverse design of inorganic materials," *npj Computational Materials*, vol. 6, no. 1, p. 84, 2020.
- [53] I. Cajigas, K. C. Davis, B. Meschede-Krasa, N. W. Prins, S. Gallo, J. A. Naeem, A. Palermo, A. Wilson, S. Guerra, B. A. Parks et al., "Implantable

brain-computer interface for neuroprosthetic-enabled volitional hand grasp restoration in spinal cord injury,” *Brain communications*, vol. 3, no. 4, p. fcab248, 2021.

[54] S. Wang, W. Yu, Z. Chen et al., “Smart diagnosis assistance method to solve results of inaccurate classification of image, and terminal based on medical images,” Mar. 18 2025, US Patent 12,254,684.

[55] A. Zulauf-Czaja, M. K. Al-Taleb, M. Purcell, N. Petric-Gray, J. Cloughley, and A. Vuckovic, “On the way home: a BCI-FES hand therapy self-managed by sub-acute SCI participants and their caregivers: a usability study,” *Journal of NeuroEngineering and Rehabilitation*, vol. 18, no. 1, p. 44, 2021.

[56] Z. Jiao, H. You, F. Yang, X. Li, H. Zhang, and D. Shen, “Decoding EEG by visual-guided deep neural networks,” in *IJCAI International Joint Conference on Artificial Intelligence*, vol. 28, 2019, p. 1387–1393.

[57] K. Fang, Z. Wang, Y. Tang, X. Guo, X. Li, W. Wang, B. Liu, and Z. Dai, “Dynamically controlled flight altitudes in robo-pigeons via locus coeruleus neurostimulation,” *Research*, vol. 8, p. 0632, 2025.

[58] Y. Liu, M. Wang, S. Hou, X. Wang, and B. Shi, “Deep learning-based markerless hand tracking for freely moving non-human primates in brain-machine interface applications,” *Electronics*, vol. 14, no. 5, p. 920, 2025.

[59] F. H. Guenther, J. S. Brumberg, E. J. Wright, A. Nieto-Castanon, J. A. Tourville, M. Panko, R. Law, S. A. Siebert, J. L. Bartels, and D. S. Andreasen, “A wireless brain-machine interface for real-time speech synthesis,” *PloS one*, vol. 4, no. 12, p. e8218, 2009.

[60] J. Tang and A. G. Huth, “Semantic language decoding across participants and stimulus modalities,” *Current Biology*, vol. 35, no. 5, p. 1023–1032.e6, 2025.

[61] S. Komeiji, T. Mitsuhashi, Y. Iimura, H. Suzuki, H. Sugano, K. Shinoda, and T. Tanaka, “Feasibility of decoding covert speech in ECoG with a transformer trained on overt speech,” *Scientific Reports*, vol. 14, no. 1, p. 11491, 2024.

[62] S. L. Metzger, J. R. Liu, D. A. Moses, M. E. Dougherty, M. P. Seaton, K. T. Littlejohn, J. Chartier, G. K. Anumanchipalli, A. Tu-Chan, and K. Ganguly, “Generalizable spelling using a speech neuroprosthesis in an individual with severe limb and vocal paralysis,” *Nature communications*, vol. 13, no. 1, p. 6510, 2022.

[63] T. Proix, J. Delgado Saa, A. Christen, S. Martin, B. N. Pasley, R. T. Knight, X. Tian, D. Poeppel, W. K. Doyle, and O. Devinsky, “Imagined speech can be decoded from low-and cross-frequency intracranial EEG features,” *Nature communications*, vol. 13, no. 1, p. 48, 2022.

[64] C. Fan, N. Hahn, F. Kamdar, D. Avansino, G. Wilson, L. Hochberg, K. V. Shenoy, J. Henderson, and F. Willett, “Plug-and-play stability for intracortical

brain-computer interfaces: a one-year demonstration of seamless brain-to-text communication,” *Advances in neural information processing systems*, vol. 36, p. 42258–42270, 2023.

[65] Y. Qi, X. Zhu, X. Xiong, X. Yang, N. Ding, H. Wu, K. Xu, J. Zhu, J. Zhang, and Y. Wang, “Human motor cortex encodes complex handwriting through a sequence of stable neural states,” *Nature Human Behaviour*, vol. 9, p. 1260–1271, 2025.

[66] E. Miliotou, P. Kyriakis, J. D. Hinman, A. Irimia, and P. Bogdan, “Generative decoding of visual stimuli,” in *International Conference on Machine Learning*, PMLR, 2023, p. 24775–24790.

[67] A. Liu, H. Jing, Y. Liu, Y. Ma, and N. Zheng, “Hidden states in LLMs improve EEG representation learning and visual decoding,” in *Frontiers in Artificial Intelligence and Applications*, vol. 392, 2024, p. 2130–2137.

[68] S. Huang, L. Sun, M. Yousefnezhad, M. Wang, and D. Zhang, “Functional alignment-auxiliary generative adversarial network-based visual stimuli reconstruction via multi-subject fMRI,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, p. 2715–2725, 2023.

[69] S. Shimizu, A. Ota, and A. Nakane, “Reviving intentional facial expressions: an interface for ALS patients using brain decoding and image-generative AI,” in *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 2025, p. 1–10.

[70] G. Mahajan, R. Jeevan, L. Divija, P. D. Kumari, and S. Narayan, “Deciphering EEG waves for the generation of images,” in *2024 12th International Winter Conference on Brain-Computer Interface (BCI)*, 2024, p. 1–6.

[71] J. Yeung, A. F. Luo, G. Sarch, M. M. Henderson, D. Ramanan, and M. J. Tarr, “Reanimating images using neural representations of dynamic stimuli,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, p. 5331–5340.

[72] Y. Fu, J. Gao, B. Yang, and J. Feng, “Making your dreams a reality: Decoding the dreams into a coherent video story from fMRI signals,” Preprint at arXiv, 2025, <https://doi.org/10.48550/arXiv.2501.09350>.

[73] T. Horikawa, M. Tamaki, Y. Miyawaki, and Y. Kamitani, “Neural decoding of visual imagery during sleep,” *Science*, vol. 340, no. 6132, pp. 639–642, 2013.

[74] J. L. Breedlove, G. St-Yves, C. A. Olman, and T. Naselaris, “Generative feedback explains distinct brain activity codes for seen and mental images,” *Current Biology*, vol. 30, no. 12, p. 2211–2224.e6, 2020.

[75] J. B. Ritchie, D. M. Kaplan, and C. Klein, “Decoding the brain: Neural representation and the limits of multivariate pattern analysis in cognitive neuroscience,” *The British journal for the philosophy of science*, 2019.

- [76] S. Wang, C. Jiang, Y. Yu, Z. Zhang, R. Quhe, R. Yang, Y. Tian, X. Chen, W. Fan, and Y. Niu, "Tellurium nanowire retinal nanoprostheses improves vision in models of blindness," *Science*, vol. 388, no. 6751, p. eadu2987, 2025.
- [77] D. J. Kravitz, K. S. Saleem, C. I. Baker, and M. Mishkin, "A new neural framework for visuospatial processing," *Nature Reviews Neuroscience*, vol. 12, no. 4, pp. 217–230, 2011.
- [78] J. Zhu and Y. Yu, "Measuring brain structure-function coupling: A promising approach to decode psychiatric neuropathology," *Biological Psychiatry*, vol. 97, no. 3, pp. 212–214, 2025.
- [79] S. Wang, W. Yu, X. Chenchen, and H. Shengye, "Visualization method for evaluating brain addiction traits, apparatus, and medium," Sep. 17 2024, US Patent 12,093,833.
- [80] M. Kwon, S. Han, K. Kim, and S. C. Jun, "Super-resolution for improving EEG spatial resolution using deep convolutional neural network—feasibility study," *Sensors*, vol. 19, no. 23, p. 5317, 2019.
- [81] G. Bao, Q. Zhang, Z. Gong, J. Zhou, W. Fan, K. Yi, U. Naseem, L. Hu, and D. Miao, "Wills Aligner: Multi-subject collaborative brain visual decoding," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, 2025, pp. 14194–14202.
- [82] D. Li, C. Du, S. Wang, H. Wang, and H. He, "Multi-subject data augmentation for target subject semantic decoding with deep multi-view adversarial learning," *Information Sciences*, vol. 547, p. 1025–1044, 2021.
- [83] G. Gaziv, R. Beliy, N. Granot, A. Hoogi, F. Strappini, T. Golan, and M. Irani, "Self-supervised natural image reconstruction and large-scale semantic classification from brain activity," *NeuroImage*, vol. 254, p. 119121, 2022.
- [84] Y. Akamatsu, R. Harakawa, T. Ogawa, and M. Haseyama, "Brain decoding of viewed image categories via semi-supervised multi-view bayesian generative model," *IEEE Transactions on Signal Processing*, vol. 68, p. 5769–5781, 2020.
- [85] D. Li, C. Du, and H. He, "Semi-supervised cross-modal image generation with generative adversarial networks," *Pattern Recognition*, vol. 100, p. 107085, 2020.
- [86] E. Dhamala, B. T. Yeo, and A. J. Holmes, "One size does not fit all: methodological considerations for brain-based predictive modeling in psychiatry," *Biological Psychiatry*, vol. 93, no. 8, pp. 717–728, 2023.
- [87] F. Ozcelik and R. VanRullen, "Natural scene reconstruction from fMRI signals using generative latent diffusion," *Scientific Reports*, vol. 13, no. 1, p. 15666, 2023.
- [88] Y. Takagi and S. Nishimoto, "High-resolution image reconstruction with latent diffusion models from human brain activity," in *Proceedings of*

the *IEEE/CVF conference on computer vision and pattern recognition*, 2023, p. 14453–14463.

[89] J. Tang, A. LeBel, S. Jain, and A. G. Huth, “Semantic reconstruction of continuous language from non-invasive brain recordings,” *Nature Neuroscience*, vol. 26, no. 5, p. 858–866, 2023.

[90] C. Du, K. Fu, J. Li, and H. He, “Decoding visual neural representations by multimodal learning of brain-visual-linguistic features,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 9, p. 10760–10777, 2023.

[91] J. Guo, C. Yi, F. Li, P. Xu, and Y. Tian, “MindLDM: Reconstruct visual stimuli from fMRI using latent diffusion model,” in *2024 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, IEEE, 2024, p. 1–6.

[92] W. Yao, X. Chen, and S. Wang, “Empowering functional neuroimaging: A pre-trained generative framework for unified representation of neural signals,” Preprint at arXiv, 2025, <https://doi.org/10.48550/arXiv.2506.02433>.

[93] G. Shen, D. Zhao, X. He, L. Feng, Y. Dong, J. Wang, Q. Zhang, and Y. Zeng, “Neuro-vision to language: Enhancing brain recording-based visual reconstruction and language interaction,” in *Advances in Neural Information Processing Systems*, vol. 37, 2024, p. 98083–98110.

[94] A. Zhang, B. Wang, X. Wu, and J. Chen, “A novel multimodal method for decoding speech perception from brain activities,” in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASP)*, IEEE, 2025, p. 1–5.

[95] S. Luo, M. Angrick, C. Coogan, D. N. Candrea, K. Wyse-Sookoo, S. Shah, Q. Rabbani, G. W. Milsap, A. R. Weiss, and W. S. Anderson, “Stable decoding from a speech BCI enables control for an individual with ALS without recalibration for 3 months,” *Advanced Science*, vol. 10, no. 35, p. 2304853, 2023.

[96] N. Xi, S. Zhao, H. Wang, C. Liu, B. Qin, and T. Liu, “Unicorn: Unified cognitive signal reconstruction bridging cognitive signals and human language,” in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2023, p. 13277–13291.

[97] G. H. Wilson, S. D. Stavisky, F. R. Willett, D. T. Avansino, J. N. Kelemen, L. R. Hochberg, J. M. Henderson, S. Druckmann, and K. V. Shenoy, “Decoding spoken english from intracortical electrode arrays in dorsal precentral gyrus,” *Journal of neural engineering*, vol. 17, no. 6, p. 066007, 2020.

[98] M. J. Vansteensel, E. G. Pels, M. G. Bleichner, M. P. Branco, T. Denison, Z. V. Freudenburg, P. Gosselaar, S. Leinders, T. H. Ottens, and M. A. Van Den Boom, “Fully implanted brain-computer interface in a locked-in patient with ALS,” *New England Journal of Medicine*, vol. 375, no. 21, p. 2060–2066, 2016.

- [99] F. R. Willett, D. T. Avansino, L. R. Hochberg, J. M. Henderson, and K. V. Shenoy, "High-performance brain-to-text communication via handwriting," *Nature*, vol. 593, no. 7858, p. 249–254, 2021.
- [100] S. Martin, P. Brunner, I. Iturrate, J. d. R. Millán, G. Schalk, R. T. Knight, and B. N. Pasley, "Word pair classification during imagined speech using direct brain recordings," *Scientific reports*, vol. 6, no. 1, p. 25803, 2016.
- [101] C. Herff, D. Heger, A. De Pesters, D. Telaar, P. Brunner, G. Schalk, and T. Schultz, "Brain-to-text: decoding spoken phrases from phone representations in the brain," *Frontiers in neuroscience*, vol. 8, p. 141498, 2015.
- [102] D. A. Moses, M. K. Leonard, J. G. Makin, and E. F. Chang, "Real-time decoding of question-and-answer speech dialogue using human cortical activity," *Nature communications*, vol. 10, no. 1, p. 3096, 2019.
- [103] D. A. Moses, S. L. Metzger, J. R. Liu, G. K. Anumanchipalli, J. G. Makin, P. F. Sun, J. Chartier, M. E. Dougherty, P. M. Liu, and G. M. Abrams, "Neuroprosthesis for decoding speech in a paralyzed person with anarthria," *New England Journal of Medicine*, vol. 385, no. 3, p. 217–227, 2021.
- [104] S. Duraivel, S. Rahimpour, C.-H. Chiang, M. Trumpis, C. Wang, K. Barth, S. C. Harward, S. P. Lad, A. H. Friedman, and D. G. Southwell, "High-resolution neural recordings improve the accuracy of speech decoding," *Nature communications*, vol. 14, no. 1, p. 6938, 2023.
- [105] J. G. Makin, D. A. Moses, and E. F. Chang, "Machine translation of cortical activity to text with an encoder-decoder framework," *Nature neuroscience*, vol. 23, no. 4, p. 575–582, 2020.
- [106] P. Sun, G. K. Anumanchipalli, and E. F. Chang, "Brain2Char: a deep architecture for decoding text from brain recordings," *Journal of neural engineering*, vol. 17, no. 6, p. 066015, 2020.
- [107] F. R. Willett, E. M. Kunz, C. Fan, D. T. Avansino, G. H. Wilson, E. Y. Choi, F. Kamdar, M. F. Glasser, L. R. Hochberg, and S. Druckmann, "A high-performance speech neuroprosthesis," *Nature*, vol. 620, no. 7976, p. 1031–1036, 2023.
- [108] I. Simanova, P. Hagoort, R. Oostenveld, and M. A. Van Gerven, "Modality-independent decoding of semantic information from the human brain," *Cerebral cortex*, vol. 24, no. 2, p. 426–434, 2014.
- [109] A. Défossez, C. Caucheteux, J. Rapin, O. Kabeli, and J.-R. King, "Decoding speech perception from non-invasive brain recordings," *Nature Machine Intelligence*, vol. 5, no. 10, p. 1097–1107, 2023.
- [110] Q. Chen, Y. Wang, F. Wang, D. Sun, and Q. Li, "Decoding text from electroencephalography signals: a novel hierarchical gated recurrent unit with masked residual attention mechanism," *Engineering Applications of Artificial Intelligence*, vol. 139, p. 109615, 2025.

- [111] Z. Wang and H. Ji, “Open vocabulary electroencephalography-to-text decoding and zero-shot sentiment classification,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, 2022, p. 5350–5358.
- [112] Y. Duan, J. Zhou, Z. Wang, Y.-K. Wang, and C.-t. Lin, “Dewave: Discrete encoding of EEG waves for EEG to text translation,” *Advances in Neural Information Processing Systems*, vol. 36, p. 9907–9918, 2023.
- [113] X. Chen, C. Du, C. Liu, Y. Wang, and H. He, “BP-GPT: Auditory neural decoding using fnri-prompted llm,” in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2025, p. 1–5.
- [114] H. Liu, D. Hajialigol, B. Antony, A. Han, and X. Wang, “EEG2Text: Open vocabulary eeg-to-text translation with multi-view transformer,” in *2024 IEEE International Conference on Big Data (BigData)*, IEEE, 2024, p. 1824–1833.
- [115] J. Zhou, Y. Duan, Y.-C. Chang, Y.-K. Wang, and C.-T. Lin, “BELT: bootstrapped EEG-to-language training by natural language supervision,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.
- [116] H. Amrani, D. Micucci, and P. Napoletano, “Deep representation learning for open vocabulary electroencephalography-to-text decoding,” *IEEE Journal of Biomedical and Health Informatics*, p. 1–12, 2024.
- [117] Y. Tao, Y. Liang, L. Wang, Y. Li, Q. Yang, and H. Zhang, “See: Semantically aligned EEG-to-text translation,” in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2025, p. 1–5.
- [118] J. Wang, Z. Song, Z. Ma, X. Qiu, M. Zhang, and Z. Zhang, “Enhancing EEG-to-text decoding through transferable representations from pre-trained contrastive eeg-text masked autoencoder,” in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2024, p. 7278–7289.
- [119] K. Luo, “Real-time open-vocabulary sentence decoding from MEG signals using transformers,” in *2024 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*, IEEE, 2024, p. 460–463.
- [120] W. Qiu, Z. Huang, H. Hu, A. Feng, Y. Yan, and R. Ying, “MindLLM: A subject-agnostic and versatile model for fMRI-to-text decoding,” Preprint at arXiv, 2025, <https://doi.org/10.48550/arXiv.2502.15786>.
- [121] M. Boyko, P. Druzhinina, G. Kormakov, A. Beliaeva, and M. Sharaev, “Megformer: enhancing speech decoding from brain activity through extended semantic representations,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2024, p. 281–290.
- [122] B. Wang, X. Xu, L. Zhang, B. Xiao, X. Wu, and J. Chen, “Semantic reconstruction of continuous language from meg signals,” in *ICASSP 2024-2024 IEEE*

*International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2024, p. 2190-2194.

[123] Y. Guo, Y. Dong, M. K.-P. Ng, and S. Wang, “A pre-trained framework for multilingual brain decoding using non-invasive recordings,” Preprint at arXiv, 2025, <https://doi.org/10.48550/arXiv.2506.03214>.

[124] A. B. Silva, J. R. Liu, S. L. Metzger, I. Bhaya-Grossman, M. E. Dougherty, M. P. Seaton, K. T. Littlejohn, A. Tu-Chan, K. Ganguly, and D. A. Moses, “A bilingual speech neuroprosthesis driven by cortical articulatory representations shared between languages,” *Nature Biomedical Engineering*, vol. 8, no. 8, p. 977-991, 2024.

[125] D. Zhang, Z. Wang, Y. Qian, Z. Zhao, Y. Liu, X. Hao, W. Li, S. Lu, H. Zhu, and L. Chen, “A brain-to-text framework for decoding natural tonal sentences,” *Cell Reports*, vol. 43, no. 11, 2024.

[126] W. Huang, P. Yang, Y. Tang, F. Qin, H. Li, D. Wu, W. Ren, S. Wang, J. Li, and Y. Zhu, “From sight to insight: A multi-task approach with the visual language decoding model,” *Information Fusion*, vol. 112, p. 102573, 2024.

[127] J. Han, K. Gong, Y. Zhang, J. Wang, K. Zhang, D. Lin, Y. Qiao, P. Gao, and X. Yue, “Onellm: One framework to align all modalities with language,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, p. 26584-26595.

[128] J. Chen, Y. Qi, Y. Wang, and G. Pan, “Mindgpt: Interpreting what you see with non-invasive brain recordings,” *IEEE Transactions on Image Processing*, 2025.

[129] S. Feng, H. Liu, Y. Wang, and Y. Wang, “Towards an end-to-end framework for invasive brain signal decoding with large language models,” in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2024, p. 1495-1499.

[130] A. Sato and I. Kobayashi, “Decoding semantic representations in the brain under language stimuli with large language models,” in *Proceedings of the First Workshop on Writing Aids at the Crossroads of AI, Cognitive Science and NLP (WRAICOGS 2025)*, 2025, p. 53-67.

[131] W. Xia, R. de Charette, C. Oztireli, and J.-H. Xue, “Umbræ: Unified multimodal brain decoding,” in *European Conference on Computer Vision*, Springer, 2024, p. 242-259.

[132] S. Wang, S. Yanyan, and W. Zhang, “Enhanced generative adversarial network and target sample recognition method,” Nov. 26 2024, US Patent 12,154,036.

[133] J. Tang, Y. Yang, Q. Zhao, Y. Ding, J. Zhang, Y. Song, and W. Kong, “Visual guided dual-spatial interaction network for fine-grained brain semantic decoding,” *IEEE Transactions on Instrumentation and Measurement*, 2024.

- [134] Y. Ikegawa, R. Fukuma, H. Sugano, S. Oshino, N. Tani, K. Tamura, Y. Iimura, H. Suzuki, S. Yamamoto, and Y. Fujita, "Text and image generation from intracranial electroencephalography using an embedding space for text and images," *Journal of Neural Engineering*, vol. 21, no. 3, p. 036019, 2024.
- [135] N. F. Ramsey, E. Salari, E. J. Aarnoutse, M. J. Vansteensel, M. G. Bleichner, and Z. Freudenburg, "Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids," *Neuroimage*, vol. 180, p. 301-311, 2018.
- [136] M. Angrick, C. Herff, E. Mugler, M. C. Tate, M. W. Slutzky, D. J. Krusienski, and T. Schultz, "Speech synthesis from ECoG using densely connected 3d convolutional neural networks," *Journal of neural engineering*, vol. 16, no. 3, p. 036019, 2019.
- [137] C. Herff, L. Diener, M. Angrick, E. Mugler, M. C. Tate, M. A. Goldrick, D. J. Krusienski, M. W. Slutzky, and T. Schultz, "Generating natural, intelligible speech from brain activity in motor, premotor, and inferior frontal cortices," *Frontiers in neuroscience*, vol. 13, p. 1267, 2019.
- [138] J. Berezutskaya, Z. V. Freudenburg, M. J. Vansteensel, E. J. Aarnoutse, N. F. Ramsey, and M. A. van Gerven, "Direct speech reconstruction from sensorimotor brain activity with optimized deep learning models," *Journal of neural engineering*, vol. 20, no. 5, p. 056010, 2023.
- [139] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, no. 7753, p. 493-498, 2019.
- [140] K. Shigemi, S. Komeiji, T. Mitsuhashi, Y. Iimura, H. Suzuki, H. Sugano, K. Shinoda, K. Yatabe, and T. Tanaka, "Synthesizing speech from ECoG with a combination of transformer-based encoder and neural vocoder," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2023, p. 1-5.
- [141] M. B. B. Ticha, X. Ran, P. Roussel, F. Bocquelet, G. Le Godais, M. Aubert, T. Costecalde, L. Struber, S. Zhang, and G. Charvet, "A vision transformer architecture for overt speech decoding from ECoG data," in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2024, p. 1-4.
- [142] X. Chen, R. Wang, A. Khalilian-Gourtani, L. Yu, P. Dugan, D. Friedman, W. Doyle, O. Devinsky, Y. Wang, and A. Flinker, "A neural speech decoding framework leveraging deep learning and speech synthesis," *Nature Machine Intelligence*, vol. 6, no. 4, p. 467-480, 2024.
- [143] K. Meng, F. Goodarzy, E. Kim, Y. J. Park, J. S. Kim, M. J. Cook, C. K. Chung, and D. B. Grayden, "Continuous synthesis of artificial speech sounds from human cortical surface recordings during silent speech production," *Journal of Neural Engineering*, vol. 20, no. 4, p. 046019, 2023.

- [144] X. Wu, S. Wellington, Z. Fu, and D. Zhang, "Speech decoding from stereo-electroencephalography (sEEG) signals using advanced deep learning methods," *Journal of Neural Engineering*, vol. 21, no. 3, p. 036055, 2024.
- [145] H. Zheng, H. Wang, W. Jiang, Z. Chen, L. He, P. Lin, P. Wei, G. Zhao, and Y. Liu, "Du-IN: Discrete units-guided mask modeling for decoding speech from intracranial neural signals," *Advances in Neural Information Processing Systems*, vol. 37, p. 79996–80033, 2024.
- [146] M. Angrick, M. C. Ottenhoff, L. Diener, D. Ivucic, G. Ivucic, S. Goulis, J. Saal, A. J. Colon, L. Wagner, and D. J. Krusienski, "Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity," *Communications biology*, vol. 4, no. 1, p. 1055, 2021.
- [147] M. Angrick, S. Luo, Q. Rabbani, D. N. Candrea, S. Shah, G. W. Milsap, W. S. Anderson, C. R. Gordon, K. R. Rosenblatt, and L. Clawson, "Online speech synthesis using a chronically implanted brain-computer interface in an individual with ALS," *Scientific reports*, vol. 14, no. 1, p. 9617, 2024.
- [148] N. S. Card, M. Wairagkar, C. Iacobacci, X. Hou, T. Singer-Clark, F. R. Willett, E. M. Kunz, C. Fan, M. Vahdati Nia, and D. R. Deo, "An accurate and rapidly calibrating speech neuroprosthesis," *New England Journal of Medicine*, vol. 391, no. 7, p. 609–618, 2024.
- [149] H. Wu, C. Cai, W. Ming, W. Chen, Z. Zhu, C. Feng, H. Jiang, Z. Zheng, M. Sawan, and T. Wang, "Speech decoding using cortical and subcortical electrophysiological signals," *Frontiers in Neuroscience*, vol. 18, p. 1345308, 2024.
- [150] C. Feng, L. Cao, D. Wu, E. Zhang, T. Wang, X. Jiang, J. Chen, H. Wu, S. Lin, and Q. Hou, "Acoustic inspired brain-to-sentence decoder for logossyllabic language," *Cyborg and Bionic Systems*, vol. 6, p. 0257, 2025.
- [151] B. Accou, J. Vanthornhout, H. V. hamme, and T. Francart, "Decoding of the speech envelope from EEG using the VLAAl deep neural network," *Scientific Reports*, vol. 13, no. 1, p. 812, 2023.
- [152] B. Van Dyck, L. Yang, and M. M. Van Hulle, "Decoding auditory EEG responses using an adapted wavenet," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–2.
- [153] C. Fan, S. Zhang, J. Zhang, E. Liu, X. Li, M. Zhao, and Z. Lv, "DMF2Mel: A dynamic multiscale fusion network for EEG-driven Mel spectrogram reconstruction," Preprint at arXiv, 2025, <https://doi.org/10.48550/arXiv.2507.07526>.
- [154] C. Fan, S. Zhang, J. Zhang, Z. Pan, and Z. Lv, "SSM2Mel: State space model to reconstruct Mel spectrogram from the EEG," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5.

- [155] D. Qi, L. Kong, L. Yang, and C. Li, “Audiiodiffusion: Generating high-quality audios from eeg signals: Reconstructing audio from EEG signals,” in *2023 4th International Symposium on Computer Engineering and Intelligent Communications (ISCEIC)*, IEEE, 2023, p. 344–348.
- [156] W. Xiong, L. Ma, and H. Li, “Synthesizing intelligible utterances from EEG of imagined speech,” *Frontiers in Neuroscience*, vol. 19, p. 1565848, 2025.
- [157] J.-W. Lee, S.-H. Lee, Y.-E. Lee, S. Kim, and S.-W. Lee, “Sentence reconstruction leveraging contextual meaning from speech-related brain signals,” in *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, 2023, p. 3721–3726.
- [158] A. Craik, H. Dial, and J. L. Contreras-Vidal, “Continuous and discrete decoding of overt speech with scalp electroencephalography (EEG),” *Journal of Neural Engineering*, vol. 22, no. 2, p. 026017, 2025.
- [159] R. Xia, C. Yin, and P. Li, “Decoding the echoes of vision from fmri: Memory disentangling for past semantic information,” in *EMNLP 2024 - 2024 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2024, p. 2040–2052.
- [160] L. Meng and C. Yang, “Semantics-guided hierarchical feature encoding generative adversarial network for visual image reconstruction from brain activity,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, p. 1267–1283, 2024.
- [161] T. Dado, Y. Güçlütürk, L. Ambrogioni, G. Ras, S. Bosch, M. van Gerven, and U. Güçlü, “Hyperrealistic neural decoding for reconstructing faces from fmri activations via the GAN latent space,” *Scientific reports*, vol. 12, no. 1, p. 141, 2022.
- [162] Y. Bai, X. Wang, Y.-p. Cao, Y. Ge, C. Yuan, and Y. Shan, “Dreamdiffusion: Generating high-quality images from brain EEG signals,” Preprint at arXiv, 2023, <https://doi.org/10.48550/arXiv.2306.16934>.
- [163] H. Wen, J. Shi, Y. Zhang, K.-H. Lu, J. Cao, and Z. Liu, “Neural encoding and decoding with deep learning for dynamic natural vision,” *Cerebral cortex*, vol. 28, no. 12, p. 4136–4160, 2018.
- [164] R. Quan, W. Wang, Z. Tian, F. Ma, and Y. Yang, “Psychometry: An omnifit model for image reconstruction from human brain activity,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025, p. 233–242.
- [165] H. Wu, Q. Li, C. Zhang, Z. He, and X. Ying, “Bridging the vision-brain gap with an uncertainty-aware blur prior,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, p. 2246–2257.
- [166] T. Naselaris, R. J. Prenger, K. N. Kay, M. Oliver, and J. L. Gallant, “Bayesian reconstruction of natural images from human brain activity,” *Neuron*,

vol. 63, no. 6, p. 902-915, 2009.

[167] C. Du, C. Du, L. Huang, and H. He, "Conditional generative neural decoding with structured cnn feature prediction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, p. 2629-2636.

[168] Y. Miyawaki, H. Uchida, O. Yamashita, M.-a. Sato, Y. Morito, H. C. Tanabe, N. Sadato, and Y. Kamitani, "Visual image reconstruction from human brain activity using a combination of multiscale local image decoders," *Neuron*, vol. 60, no. 5, p. 915-929, 2008.

[169] C. Du, C. Du, L. Huang, and H. He, "Reconstructing perceived images from human brain activities with bayesian deep multiview learning," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 8, p. 2310-2323, 2018.

[170] Y. Güçlütürk, U. Güçlü, K. Seeliger, S. Bosch, R. van Lier, and M. A. van Gerven, "Reconstructing perceived faces from brain activations with deep adversarial neural decoding," *Advances in neural information processing systems*, vol. 30, 2017.

[171] R. VanRullen and L. Reddy, "Reconstructing faces from fMRI patterns using deep generative neural networks," *Communications biology*, vol. 2, no. 1, p. 193, 2019.

[172] R. Beliy, G. Gaziv, A. Hoogi, F. Strappini, T. Golan, and M. Irani, "From voxels to pixels and back: Self-supervision in natural-image reconstruction from fMRI," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[173] G. Shen, T. Horikawa, K. Majima, and Y. Kamitani, "Deep image reconstruction from human brain activity," *PLoS computational biology*, vol. 15, no. 1, p. e1006633, 2019.

[174] J. Luo, W. Cui, J. Liu, Y. Li, Y. Guo, S. Xu, and L. Wang, "Visual image decoding of brain activities using a dual attention hierarchical latent generative network with multiscale feature fusion," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 15, no. 2, p. 761-773, 2022.

[175] G. Kupersmidt, R. Beliy, G. Gaziv, and M. Irani, "A penny for your (visual) thoughts: Self-supervised reconstruction of natural movies from brain activity," Preprint at arXiv, 2022, <https://doi.org/10.48550/arXiv.2206.03544>.

[176] Q. Zhou, C. Du, D. Li, B. Wen, L. Chang, and H. He, "Interpretable visual neural decoding with unsupervised semantic disentanglement," *Machine Intelligence Research*, p. 1-18, 2025.

[177] P. Tirupattur, Y. S. Rawat, C. Spampinato, and M. Shah, "Thoughtviz: Visualizing human thoughts using generative adversarial network," in *Proceedings of the 26th ACM international conference on Multimedia*, 2018, p. 950-958.

[178] S. Palazzo, C. Spampinato, I. Kavasidis, D. Giordano, and M. Shah, "Generative adversarial networks conditioned by brain signals," in *Proceedings of the*

*IEEE international conference on computer vision*, 2017, p. 3410–3418.

[179] I. Kavasidis, S. Palazzo, C. Spampinato, D. Giordano, and M. Shah, “Brain2image: Converting brain signals into images,” in *Proceedings of the 25th ACM international conference on Multimedia*, 2017, p. 1809–1817.

[180] H. Ahmadi, F. Gasse, and M. H. Moradi, “Visual image reconstruction based on EEG signals using a generative adversarial and deep fuzzy neural network,” *Biomedical Signal Processing and Control*, vol. 87, p. 105497, 2024.

[181] A. Fares, S.-h. Zhong, and J. Jiang, “Brain-media: A dual conditioned and lateralization supported GAN (DCLS-GAN) towards visualization of image-evoked brain activities,” in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, p. 1764–1772.

[182] W. Huang, H. Yan, C. Wang, J. Li, Z. Zuo, J. Zhang, Z. Shen, and H. Chen, “Perception-to-image: Reconstructing natural images from the brain activity of visual perception,” *Annals of Biomedical Engineering*, vol. 48, no. 9, p. 2323–2332, 2020.

[183] Y. Lu, C. Du, Q. Zhou, D. Wang, and H. He, “Minddiffuser: Controlled image reconstruction from human brain activity with semantic and structural diffusion,” in *Proceedings of the 31st ACM international conference on multimedia*, 2023, p. 7202–7211.

[184] Y. Lu, C. Du, C. Wang, X. Zhu, L. Jiang, X. Li, and H. He, “Animate your thoughts: Reconstruction of dynamic natural vision from human brain activity,” in *13th International Conference on Learning Representations*, 2025, pp. 23255–23275.

[185] R. Child, “Very deep vae generalize autoregressive models and can outperform them on images,” Preprint at arXiv, 2020, <https://doi.org/10.48550/arXiv.2011.10650>.

[186] J. Li, D. Li, C. Xiong, and S. Hoi, “Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation,” in *International conference on machine learning*, 2022, pp. 12888–12900.

[187] A. Luo, M. Henderson, L. Wehbe, and M. Tarr, “Brain diffusion for visual exploration: Cortical discovery using large scale generative models,” in *Advances in Neural Information Processing Systems*, vol. 36, 2023, pp. 75740–75781.

[188] M. Ferrante, T. Boccato, L. Passamonti, and N. Toschi, “Retrieving and reconstructing conceptually similar images from fmri with latent diffusion models and a neuro-inspired brain decoding model,” *Journal of Neural Engineering*, vol. 21, no. 4, p. 046001, 2024.

[189] P. S. Scotti, M. Tripathy, C. K. T. Villanueva, R. Kneeland, T. Chen, A. Narang, C. Santhirasegaran, J. Xu, T. Naselaris, K. A. Norman, and T. M. Abraham, “MindEye2: Shared-subject models enable fMRI-to-image with 1 hour of data,” in *Proceedings of Machine Learning Research*, vol. 235, Vienna, Austria, 2024, pp. 44038–44059.

- [190] Z. Chen, J. Qing, T. Xiang, W. L. Yue, and J. H. Zhou, “Seeing beyond the brain: Conditional diffusion model with sparse masked modeling for vision decoding,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, p. 22710–22720.
- [191] Q. Zhou, C. Du, S. Wang, and H. He, “CLIP-MUSED: CLIP-guided multi-subject visual neural information semantic decoding,” in *12th International Conference on Learning Representations, ICLR 2024*, 2024.
- [192] Y. Ma, Y. Liu, L. Chen, G. Zhu, B. Chen, and N. Zheng, “BrainCLIP: Brain representation via CLIP for generic natural visual stimulus decoding,” *IEEE Transactions on Medical Imaging*, 2025.
- [193] S. V. Bhalerao and R. B. Pachori, “Automated classification of cognitive visual objects using multivariate swarm sparse decomposition from multichannel EEG-MEG signals,” *IEEE Transactions on Human-Machine Systems*, vol. 54, no. 4, pp. 455–464, 2024.
- [194] S. Wang, S. Liu, Z. Tan, and X. Wang, “Mindbridge: A cross-subject brain decoding framework,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, p. 11333–11342.
- [195] Z. Gong, Q. Zhang, G. Bao, L. Zhu, R. Xu, K. Liu, L. Hu, and D. Miao, “Mindtuner: Cross-subject visual decoding with visual fingerprint and semantic correction,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, 2025, p. 14247–14255.
- [196] A. F. Luo, M. M. Henderson, M. J. Tarr, and L. Wehbe, “Brainscuba: Fine-grained natural language captions of visual cortex selectivity,” in *12th International Conference on Learning Representations, ICLR 2024*, 2024.
- [197] Y. Zhao, G. Dong, L. Zhu, and X. Ying, “Memory recall: Retrieval-augmented mind reconstruction for brain decoding,” *Information Fusion*, p. 103280, 2025.
- [198] J. Huo, Y. Wang, Y. Wang, X. Qian, C. Li, Y. Fu, and J. Feng, “Neuro-ictor: Refining fMRI-to-image reconstruction via multi-individual pretraining and multi-level modulation,” in *European Conference on Computer Vision*, 2024, pp. 56–73.
- [199] P. Sorino, G. M. Biancofiore, D. Lofù, T. Colafoglio, A. Lombardi, F. Narducci, and T. Di Noia, “Ariel: Brain-computer interfaces meet large language models for emotional support conversation,” in *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*, 2024, pp. 601–609.
- [200] T. Dado, P. Papale, A. Lozano, L. Le, F. Wang, M. van Gerven, P. Roelfsema, Y. Güçlütürk, and U. Güçlü, “Brain2GAN: Feature-disentangled neural encoding and decoding of visual perception in the primate brain,” *PLoS computational biology*, vol. 20, no. 5, p. e1012058, 2024.

- [201] M. Liu and I. Kobayashi, “Do feature representations from different language models affect accuracy of brain encoding models’ predictions?” in *2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, 2024, p. 1-6.
- [202] A. K. Singh, Y.-K. Wang, J.-T. King, and C.-T. Lin, “Extended interaction with a BCI video game changes resting-state brain activity,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 12, no. 4, pp. 809-823, 2020.
- [203] S. Nishimoto, A. T. Vu, T. Naselaris, Y. Benjamini, B. Yu, and J. L. Gallant, “Reconstructing visual experiences from brain activity evoked by natural movies,” *Current biology*, vol. 21, no. 19, p. 1641-1646, 2011.
- [204] M. Hanke, F. J. Baumgartner, P. Ibe, F. R. Kaule, S. Pollmann, O. Speck, W. Zinke, and J. Stadler, “A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie,” *Scientific data*, vol. 1, no. 1, p. 14003, 2014.
- [205] J. Sun, M. Z. Chen, and M.-F. Moens, “Neurocine: Decoding human brain activities,” Preprint at arXiv, 2024, <https://doi.org/10.48550/arXiv.2402.01590>.
- [206] J. Sun, M. Li, and M.-F. Moens, “Neuralflix: A simple while effective framework for semantic decoding of videos from non-invasive brain recordings,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, 2025, pp. 7096-7104.
- [207] C. Fosco, B. Lahner, B. Pan, A. Andonian, E. Josephs, A. Lascelles, and A. Oliva, “Brain netflix: Scaling data to reconstruct videos from brain signals,” in *European Conference on Computer Vision*, 2024, pp. 457-474.
- [208] K. Simonyan and A. Zisserman, “Two-stream convolutional networks for action recognition in videos,” *Advances in neural information processing systems*, vol. 27, 2014.
- [209] J. Carreira and A. Zisserman, “Quo vadis, action recognition? a new model and the kinetics dataset,” in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6299-6309.
- [210] Z. Chen, J. Qing, and J. H. Zhou, “Cinematic mindscapes: High-quality video reconstruction from brain activity,” *Advances in Neural Information Processing Systems*, vol. 36, p. 24841-24858, 2023.
- [211] Z. Gong, G. Bao, Q. Zhang, Z. Wan, D. Miao, S. Wang, L. Zhu, C. Wang, R. Xu, and L. Hu, “NeuroClips: Towards high-fidelity and smooth fMRI-to-video reconstruction,” *Advances in Neural Information Processing Systems*, vol. 37, p. 51655-51683, 2024.
- [212] X.-H. Liu, Y.-K. Liu, Y. Wang, K. Ren, H. Shi, Z. Wang, D. Li, B.-L. Lu, and W.-L. Zheng, “EEG2video: Towards decoding dynamic visual perception from EEG signals,” *Advances in Neural Information Processing Systems*, vol. 37, p. 72245-72273, 2024.

- [213] G. Shen, D. Zhao, X. He, L. Feng, Y. Dong, J. Wang, Q. Zhang, and Y. Zeng, “Neuro-vision to language: Enhancing brain recording-based visual reconstruction and language interaction,” *Advances in Neural Information Processing Systems*, vol. 37, pp. 98083–98110, 2024.
- [214] J. Lin, H. Chen, Y. Fan, Y. Fan, X. Jin, H. Su, J. Fu, and X. Shen, “Multi-layer visual feature fusion in multimodal llms: Methods, analysis, and best practices,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 4156–4166.
- [215] B. Zhang, Y. Fang, T. Ren, and G. Wu, “Multimodal analysis for deep video understanding with video language transformer,” in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 7165–7169.
- [216] T.-j. Luo, Y. Fan, L. Chen, G. Guo, and C. Zhou, “EEG signal reconstruction using a generative adversarial network with wasserstein distance and temporal-spatial-frequency loss,” *Frontiers in neuroinformatics*, vol. 14, p. 15, 2020.
- [217] N. K. N. Aznan, A. Atapour-Abarghouei, S. Bonner, J. D. Connolly, N. Al Moubayed, and T. P. Breckon, “Simulating brain signals: Creating synthetic EEG data via neural-based generative models for improved ssvep classification,” in *2019 International joint conference on neural networks (IJCNN)*, IEEE, 2019, p. 1–8.
- [218] J. Xie, S. Chen, Y. Zhang, D. Gao, and T. Liu, “Combining generative adversarial networks and multi-output CNN for motor imagery classification,” *Journal of neural engineering*, vol. 18, no. 4, p. 046026, 2021.
- [219] R. Yuste, S. Goering, B. A. Y. Arcas, G. Bi, J. M. Carmena, A. Carter, J. J. Fins, P. Friesen, J. Gallant, J. E. Huggins et al., “Four ethical priorities for neurotechnologies and AI,” *Nature*, vol. 551, no. 7679, pp. 159–163, 2017.
- [220] A. J. Anderson, K. McDermott, B. Rooks, K. L. Heffner, D. Dodell-Feder, and F. V. Lin, “Decoding individual identity from brain activity elicited in imagining common experiences,” *Nature communications*, vol. 11, no. 1, p. 5916, 2020.
- [221] C. Li, X. Qian, Y. Wang, J. Huo, X. Xue, Y. Fu, and J. Feng, “Enhancing cross-subject fmri-to-video decoding with global-local functional alignment,” in *European Conference on Computer Vision*, Springer, 2024, p. 353–369.
- [222] S. Wen, A. Yin, T. Furlanello, M. G. Perich, L. E. Miller, and L. Itti, “Rapid adaptation of brain-computer interfaces to new neuronal ensembles or participants via generative modelling,” *Nature biomedical engineering*, vol. 7, no. 4, pp. 546–558, 2023.
- [223] R. J. Chen, J. J. Wang, D. F. Williamson, T. Y. Chen, J. Lipkova, M. Y. Lu, S. Sahai, and F. Mahmood, “Algorithmic fairness in artificial intelligence for medicine and healthcare,” *Nature biomedical engineering*, vol. 7, no. 6, pp. 719–742, 2023.

[224] Y. Yang, H. Zhang, J. W. Gichoya, D. Katabi, and M. Ghassemi, “The limits of fair medical imaging AI in real-world generalization,” *Nature Medicine*, vol. 30, no. 10, pp. 2838–2848, 2024.

[225] R. Liu, Y. Chen, A. Li, Y. Ding, H. Yu, and C. Guan, “Aggregating intrinsic information to enhance BCI performance through federated learning,” *Neural Networks*, vol. 172, p. 106100, 2024.

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv –Machine translation. Verify with original.*