Decoding Speech from Neural Signals: A Comprehensive Analysis of AI-Powered Electroencephalography (EEG) Interpretation

Introduction

The ability to communicate is a cornerstone of human experience, yet millions of individuals are deprived of this capacity due to severe motor and speech impairments resulting from conditions such as amyotrophic lateral sclerosis (ALS), stroke, brain injury, or locked-in syndrome.[1] For these individuals, brain-computer interfaces (BCIs) represent a technological frontier of profound hope, offering a potential pathway to restore communication by directly translating neural activity into commands, text, or synthesized speech.[1] By bypassing the body's damaged neuromuscular pathways, BCIs aim to create a direct link between thought and the external world.

Among the various neuroimaging modalities available for BCI development, electroencephalography (EEG) has emerged as a focal point of research and practical application. As a non-invasive technique that measures electrical activity from the scalp, EEG offers an unparalleled combination of excellent temporal resolution, safety, portability, and relatively low cost, making it the de-facto standard for developing scalable BCI systems intended for widespread use.[9] Its ability to capture brain dynamics on a millisecond timescale is theoretically ideal for tracking the rapid and complex processes underlying speech.[4]

However, the promise of EEG is tempered by a formidable challenge: decoding intelligible words from the inherent noise of the signal. The electrical potentials recorded at the scalp are the faint, summated whisper of millions of neurons, and this delicate signal is profoundly corrupted by both background brain activity and a host of biological and environmental artifacts.[4] The core scientific problem, therefore, is one of signal extraction and interpretation—a task for which artificial intelligence (AI), and particularly deep learning, has become an indispensable tool. Advanced AI models provide the computational power necessary to navigate this noisy landscape, automatically learning to filter, enhance, and ultimately decode the intricate patterns of neural activity associated with speech.[4] The field has thus progressed from classifying simple, isolated commands to the ambitious goal of synthesizing continuous, open-vocabulary speech directly from brainwaves.

This report provides a comprehensive and exhaustive analysis of the current state of AI-powered EEG speech decoding. It begins by establishing the fundamental neurophysiological context, detailing the properties of EEG signals and comparing them to other neuroimaging modalities. It then delves into the critical data processing pipeline required to transform raw neural data into a format suitable for machine learning. The core of the report examines the evolution of AI architectures—from classical machine learning to sophisticated deep learning models like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers—and their application to speech decoding. A critical evaluation of system performance, key breakthroughs, and persistent bottlenecks follows, supported by concrete benchmarks and an analysis of the data limitations that hinder progress. Finally, the report looks to the future, exploring the roadmap for real-world applications, the ongoing technical and hardware challenges, and the profound ethical considerations that must be addressed as this transformative technology moves from the laboratory to society.

Section 1: The Neurophysiological Landscape of Speech and EEG

The endeavor to decode speech from EEG signals is fundamentally constrained by the biophysical properties of the brain and the technologies used to measure its activity. Understanding the advantages and limitations of EEG, how it compares to other neuroimaging modalities, and its relationship with the neural processes of speech is essential for appreciating the complexity of the task and the rationale behind the AI-driven solutions being developed.

1.1 The Nature of EEG Signals: A Double-Edged Sword

EEG measures the voltage fluctuations on the scalp that result from the summation of synchronous postsynaptic potentials generated by millions of large pyramidal neurons in the cerebral cortex.[13] This method of capturing brain activity presents a dichotomy of powerful advantages and significant drawbacks, making it a double-edged sword for BCI development.

Key Advantages

The primary advantage of EEG lies in its exceptional **temporal resolution**. It can capture neural dynamics at the millisecond level, a timescale that is crucial for tracking the rapidly unfolding processes of speech perception and production.[4] This high temporal fidelity gives researchers immediate information about brain activity as it happens, which is a prerequisite for any real-time communication system.

Furthermore, EEG is valued for its **non-invasiveness, accessibility, and portability**. Unlike techniques that require surgical implantation, EEG electrodes are placed on the scalp, posing no risk to the user. The equipment is relatively inexpensive and increasingly portable, with modern systems becoming wireless and wearable.[9] This combination of safety and low cost makes EEG the most viable and widely used modality for developing scalable BCI applications that can move beyond the laboratory and into real-world clinical and consumer settings.

Inherent Disadvantages

Despite its advantages, EEG is plagued by several inherent limitations that form the central challenge for speech decoding. The most significant of these is an intrinsically **low signal-to-noise ratio (SNR)**. The neural signals of interest are incredibly faint (on the order of microvolts) and are easily obscured by the brain's own background electrical activity, as well as by a host of non-neural artifacts.[1] The skull, dura, and scalp act as a volume conductor, effectively smearing the electrical signals before they reach the electrodes.

This volume conduction effect also leads to EEG's second major drawback: **poor spatial resolution**. It is difficult to precisely pinpoint the anatomical source of a recorded signal, as each electrode picks up a mixture of activity from a wide area of the underlying cortex.[4] This limitation contrasts sharply with modalities like fMRI or ECoG, which can offer much greater spatial specificity.

Finally, EEG signals are highly **susceptible to artifacts**. These are electrical signals not of cerebral origin that can be many times larger in amplitude than the target neural signals. Key sources of artifacts include:

- **Electromyographic (EMG) artifacts:** Generated by muscle contractions, particularly from the face, jaw, neck, and scalp. These are especially problematic during overt or even attempted speech tasks.[4]
- **Electrooculographic (EOG) artifacts:** Caused by eye movements and blinks.[9]
- **Electrocardiographic (ECG) artifacts:** Resulting from the electrical activity of the heart.
- **External interference:** Noise from nearby electrical equipment and power lines (e.g., 50/60 Hz hum).[9]

The challenge of EEG-based speech decoding is therefore less about deciphering a clean neural code and more about extracting a faint, meaningful signal from a profoundly noisy and complex background. This reality dictates that any successful decoding pipeline must incorporate sophisticated signal processing and AI-driven techniques capable of acting as powerful denoisers and feature extractors.

1.2 A Comparative Analysis of Neuroimaging Modalities

To fully appreciate the role of EEG, it is useful to compare it with other neuroimaging techniques used in BCI research. The choice of modality establishes a fundamental trade-off between signal quality, invasiveness, and practicality, which in turn dictates the entire subsequent AI and signal processing strategy. Invasive methods provide cleaner signals but are limited to a few clinical patients, while non-invasive methods are scalable but demand more advanced computational solutions to overcome signal quality issues.

Invasive vs. Non-Invasive Approaches: The most fundamental distinction is between invasive and non-invasive methods. Invasive techniques like **Electrocorticography (ECoG)**, where electrodes are placed directly on the surface of the brain, and **Stereoelectroencephalography (sEEG)**, where depth electrodes are inserted into the brain, bypass the distorting effects of the skull and scalp. This provides signals with a much higher SNR, superior spatial resolution, and access to high-frequency brain activity that is attenuated at the scalp.[1] Consequently, invasive BCIs have consistently demonstrated higher decoding accuracies, with some systems achieving remarkable performance in speech-to-text translation.[1] However, their use is restricted to patients undergoing neurosurgery for clinical reasons (e.g., epilepsy monitoring), making them unsuitable for widespread application.[7] Non-invasive methods, led by EEG, are the only viable path toward broadly accessible BCI technology, despite their significant signal processing challenges.[10]

Comparison of Non-Invasive Modalities:

- **EEG vs. Magnetoencephalography (MEG):** MEG measures the magnetic fields produced by the brain's electrical currents. These magnetic fields are not distorted by the skull, giving MEG better spatial resolution than EEG.[10] However, MEG systems are extremely sensitive to movement artifacts, require large, non-portable, and expensive magnetically shielded rooms, and have shown only marginal improvements in decoding accuracy over EEG in some comparative studies.[22] Recent work by Meta AI did show that MEG outperformed EEG in a character decoding task during typing, achieving up to 80% accuracy, but the practical barriers to widespread MEG use remain substantial.[25]
- **EEG vs. fMRI and fNIRS:** Functional Magnetic Resonance Imaging (fMRI) and Functional Near-Infrared Spectroscopy (fNIRS) do not measure neural activity directly. Instead, they measure the slower hemodynamic response—changes in blood flow and oxygenation associated with neural activation. This provides excellent spatial resolution, allowing for precise localization of brain function. However, their temporal resolution is on the order of seconds, making them far too slow to track the millisecond-level dynamics of continuous speech.[9] They are valuable tools for studying the broader neural networks involved in language but are not suitable for real-time speech decoding BCIs.

The following table summarizes the key characteristics of these modalities in the context of speech BCI.

Modality	Invasiveness	Temporal Resolution	Spatial Resolution	Portability / Cost	Key Advantages for Speech BCI	Key Disadvantages for Speech BCI
EEG	Non-invasive	Excellent (ms)	Poor (cm)	High / Low	High temporal resolution,	Low SNR, poor spatial

					portable, low cost, safe, scalable. [9]	resolution, high susceptibility to artifacts. [9]
MEG	Non-invasive	Excellent (ms)	Good (mm)	Low / Very High	High temporal resolution, less signal distortion by skull than EEG. [21]	Requires magnetically shielded room, very expensive, sensitive to movement artifacts. [22]
fNIRS	Non-invasive	Poor (s)	Moderate (mm)	High / Moderate	Portable, less sensitive to movement than EEG/MEG. [23]	Poor temporal resolution, limited to cortical surface. [26]
fMRI	Non-invasive	Poor (s)	Excellent (mm)	None / Very High	Excellent spatial resolution, whole-brain coverage. [10]	Very poor temporal resolution, requires subject to be immobile in a scanner. [10]
ECoG	Invasive	Excellent (ms)	Very Good (mm)	Low (implant)	High SNR, high spatial and temporal resolution, access to high-gamma band. [1]	Requires neurosurgery, limited to clinical patients, risk of infection. [5]

sEEG	Invasive	Excellent (ms)	Very Good (mm)	Low (implant)	High SNR, high resolution, can record from deep brain structures. [5]	Requires neurosurgery, sparser cortical coverage than ECoG. [5]

Table 1: Comparative Analysis of Neuroimaging Modalities for Speech BCI. This table synthesizes data from multiple sources to provide a comparative overview of the primary technologies used in speech decoding research, highlighting the trade-offs that make EEG a focal point despite its limitations.[7]

1.3 Neural Correlates of Speech

Successful decoding relies on identifying consistent patterns of neural activity—or neural correlates—associated with specific speech events. Research has focused on identifying these correlates in both the frequency and spatial domains.

Brain Waves and Their Roles: Different frequency bands of the EEG signal have been linked to distinct aspects of language processing [9]:

- **Delta (δ) band (0.5–4 Hz):** This low-frequency band is associated with the perception of speech rhythm, intonation, and prosodic phrasing.
- **Theta (θ) band (4–8 Hz):** Theta waves are active during tasks that involve piecing words together, such as phonemic restoration and processing co-articulation cues between sounds.
- **Alpha (α) band (8–13 Hz):** Alpha activity is involved in auditory feedback and speech perception. Notably, alpha power is often weaker during imagined speech compared to overt speech.
- **Beta (β) band (13–30 Hz):** Beta waves are often linked to motor processes and feedback, making them relevant during both auditory tasks and the motor planning aspects of speech production.
- **Gamma (γ) band (30-150 Hz):** Changes in high-gamma frequencies are strongly correlated with both overt (spoken) and covert (imagined) speech production, with activity observed in key language and motor areas of the brain.

Cortical Regions: Decoding efforts often target specific brain regions known to be central to language. These include the primary motor cortex, which controls the articulators (tongue, lips, jaw), as well as the classical language centers: **Broca's area** (involved in speech production) and **Wernicke's area** (involved in language comprehension).[9] Identifying which EEG channels overly these regions is a common strategy for improving decoding performance.[10]

Overt vs. Imagined Speech: A critical distinction in BCI research is between overt and imagined speech. Overt speech involves actual vocalization and movement of the articulators, while **imagined speech** (also called covert or inner speech) is the internal monologue or rehearsal of speech without any physical movement.[7] Because it does not require motor control, imagined speech is a prime paradigm for BCIs aimed at helping paralyzed individuals. However, the neural signatures associated with imagined speech are generally weaker and more difficult to distinguish than those of overt speech, posing a greater challenge for decoding algorithms.[9]

Section 2: The Signal Processing Gauntlet: From Raw EEG to Actionable Features

Before an AI model can attempt to decode words from EEG, the raw, noisy signal must undergo a rigorous pipeline of processing and transformation. This "gauntlet" is designed to clean the data, reduce its complexity, and extract the most informative features related to the speech task. The evolution of these techniques reflects a broader conceptual shift in the field, moving from treating EEG as a collection of independent signals to viewing it as a unified, spatiotemporal data structure, a perspective that has unlocked the application of more powerful AI models.

2.1 Pre-processing: The First Line of Defense Against Noise

The primary objective of pre-processing is to enhance the relevant neural information within the EEG signal by increasing the SNR and removing or mitigating the effects of artifacts.[9] This step is crucial for improving the efficiency and accuracy of the subsequent classification or decoding models.

Filtering Techniques: Filtering is one of the most fundamental pre-processing steps.

- **Band-pass Filtering:** This is almost universally applied to isolate the frequency bands of interest. Researchers typically apply a band-pass filter to retain frequencies where speech-related neural information is believed to reside (e.g., between 1 Hz and 45 Hz or 0.3 Hz and 60 Hz) while removing very low-frequency drift and high-frequency noise.[9] A
  notch filter is also commonly used to specifically remove power line interference at 50 Hz or 60 Hz.[29]
- **Spatial Filtering:** These methods leverage the multi-channel nature of EEG to reduce noise that is common across many electrodes. **Common Average Reference (CAR)**, for example, improves SNR by subtracting the average signal across all electrodes from each individual electrode's signal, thereby removing widespread, non-specific activity.[9]
  Laplacian filters are another spatial technique, though they are used less frequently due to the risk of losing valuable information.[9]

Artifact Removal: Given the high amplitude of artifacts compared to neural signals, their removal is critical.

- **Independent Component Analysis (ICA):** This is a powerful and widely used blind source separation technique. ICA decomposes the multi-channel EEG signal into a set of statistically independent components. Components that have the characteristic signatures of stereotyped artifacts, such as eye blinks or muscle activity, can be identified (often through visual inspection or automated classifiers) and removed from the data before reconstructing the cleaned EEG signal.[9]
- **Regression-Based Methods:** In some experimental setups, additional electrodes are placed to specifically record artifactual signals (e.g., EOG channels near the eyes, EMG channels on facial muscles). The signals from these channels can then be used in a regression model to predict and subtract their influence from the EEG channels.

Data Normalization and Segmentation: Finally, the continuous EEG data is prepared for model input. This often involves **downsampling** the signal (e.g., from 1000 Hz to 256 Hz) to reduce the computational complexity without losing critical information in the target frequency bands.[9] The continuous recording is then segmented into

epochs, which are short time windows (e.g., 2-5 seconds) that are time-locked to specific events, such as the presentation of a word cue for the participant to imagine.[9]

2.2 Feature Engineering: Extracting the Essence of Speech

Once the signal is cleaned, the next step in traditional machine learning pipelines is feature engineering: the process of explicitly calculating descriptive characteristics of the signal that a classifier can use to distinguish between different mental states. This process can be performed in three primary domains.[9]

- **Time Domain:** These features are calculated directly from the signal's amplitude over time. Common time-domain features include statistical measures like the **Root Mean Square (RMS)**, **variance**, **standard deviation**, **mean**, and **Hjorth parameters** (which measure signal activity, mobility, and complexity).[9]
- **Frequency Domain:** These features describe the distribution of power across different frequency bands. The most common methods involve transforms that convert the time-domain signal into the frequency domain, such as the **Fast Fourier Transform (FFT)** or the **Short-Time Fourier Transform (STFT)**, which analyzes frequency content in short, overlapping windows. **Wavelet Transforms (WT, DWT, CWT)** are also popular as they provide a time-frequency representation, showing how the frequency content of the signal changes over time.[9] Another powerful feature set borrowed from audio processing is **Mel Frequency Cepstral Coefficients (MFCCs)**, which represent the short-term power spectrum of a sound on a nonlinear mel scale of pitch.[9]
- **Spatial Domain:** These features leverage the spatial arrangement of the EEG electrodes. The most prominent method is **Common Spatial Patterns (CSP)**, an algorithm that designs spatial filters to find projections of the data that maximize the variance for one class while minimizing it for another, making it highly effective for discriminating between two conditions.[9]

The progression from analyzing individual channels to incorporating spatial relationships marks a significant step forward. Methods like **Channel Cross-Covariance (CCV)** matrices capture the statistical interrelationships between signals from different electrodes, providing a more holistic view of brain network activity.[9] This approach acknowledges that brain function is inherently distributed and networked, and that the relationships between brain areas contain valuable information.

A more recent and powerful feature representation technique is the creation of **Topographic Brain Maps**. This method transforms the 1D time-series data from all electrodes into a 2D image (or a sequence of 2D images to form a 3D volume) that represents the spatial distribution of electrical potential across the scalp at a given moment.[15] This is a pivotal innovation because it reframes the EEG decoding problem as an image classification problem. This allows the direct application of highly successful and powerful deep learning architectures from the field of computer vision, such as Convolutional Neural Networks (CNNs), which are expertly designed to learn hierarchical spatial and temporal features from image-like data. This shift from treating EEG as a collection of separate time-series to a unified, dynamic spatiotemporal field has been a key enabler of recent progress.[4]

2.3 The End-to-End Paradigm Shift

The rise of deep learning has ushered in a paradigm shift away from the traditional, multi-step pipeline towards an **end-to-end** approach.[9] In this paradigm, a single, deep neural network is trained to learn the entire mapping from raw (or minimally processed) EEG signals directly to the final output, such as a word label or a synthesized speech waveform.

This approach has several key advantages. It bypasses the need for manual, domain-expert-driven feature engineering, which can be time-consuming, subjective, and may inadvertently discard useful information.[15] Instead, the deep learning model is trusted to automatically learn the optimal hierarchical features from the data itself. Models like CNNs and Transformers are particularly well-suited for this, as their layered structure naturally allows them to learn increasingly abstract and relevant features from the input signal.[4] Recent frameworks like

FESDE (Fully-End-to-end Speech Decoding) epitomize this philosophy, aiming to directly reconstruct audible speech waveforms from EEG signals without any intermediate acoustic feature representation steps, such as conversion to mel-spectrograms.[19] This end-to-end approach represents the ultimate expression of leveraging AI to handle the full complexity of the EEG decoding problem.

Section 3: The Algorithmic Frontier: AI Architectures for Speech

Decoding

The core of any brain-to-text system is the algorithm that translates neural patterns into language. The field has seen a rapid evolution in these algorithms, moving from classical machine learning models that established initial feasibility to increasingly sophisticated deep learning architectures capable of handling the immense complexity of EEG data. This progression is not arbitrary; the architectural choices made by researchers are tailored solutions designed to specifically address the spatial, temporal, and sequential nature of speech-related brain signals.

3.1 Classical Machine Learning: The Foundation

Before the widespread adoption of deep learning, classical machine learning (ML) algorithms were instrumental in demonstrating that decoding information from EEG was possible. While now often used as benchmarks rather than state-of-the-art methods for complex tasks, they laid the essential groundwork for the field.[9]

- **Linear Discriminant Analysis (LDA)** and **Support Vector Machines (SVM)** were common choices for early classification experiments, particularly for tasks with a small number of distinct classes (e.g., discriminating between two or three imagined words or movements).[9]
- **Random Forests (RF)**, an ensemble method based on decision trees, has shown surprisingly robust performance in some studies. One review highlights an RF model achieving a remarkable 94.6% accuracy in a pair-wise classification task, demonstrating its effectiveness for simpler, well-defined problems.[4] Another study used RF for a two-level classification framework, first distinguishing between broad categories (e.g., images vs. characters) with 85.2% accuracy, and then performing finer classification within categories with 67.03% accuracy.[4]

While effective for limited-vocabulary classification, these classical models typically rely on extensive, handcrafted feature engineering and struggle to generalize to the complexities of continuous, open-vocabulary speech decoding.

3.2 The Deep Learning Revolution: Capturing Complexity

The advent of deep learning revolutionized the field by introducing models capable of automatically learning hierarchical feature representations directly from data, thereby overcoming many limitations of classical ML.[4] Two architectures, in particular, have become cornerstones of modern EEG decoding: CNNs and RNNs.

Convolutional Neural Networks (CNNs)

Primarily known for their dominance in computer vision, CNNs are exceptionally good at extracting spatial hierarchies of features. Their application to EEG became particularly effective with the innovation of representing EEG data as image-like topographic maps.[4]

- **1D CNNs** can be applied directly to the raw time-series data from each EEG channel to learn temporal patterns.[21]
- **2D CNNs** are used to process static topographic maps, capturing the spatial distribution of brain activity at a single moment.
- **3D CNNs** represent a significant advance, as they process sequences of 2D topographic maps (i.e., a 3D data volume). This allows the model to learn both spatial features within each map and temporal features across the sequence of maps simultaneously, capturing the dynamic evolution of brain activity.[4]

Recurrent Neural Networks (RNNs)

RNNs are specifically designed to process sequential data, making them a natural fit for modeling both the time-course of EEG signals and the sequential nature of language itself.[4]

- **Long Short-Term Memory (LSTM)** networks are the most common type of RNN used. Their internal "gating" mechanism allows them to selectively remember or forget information over long sequences, effectively addressing the vanishing gradient problem that plagues simple RNNs and enabling the learning of long-term dependencies.[4]
- **Bidirectional LSTMs (BiLSTMs)** improve upon standard LSTMs by processing

the input sequence in both the forward and backward directions. This provides the network with context from both past and future time steps at every point in the sequence, often leading to a richer representation and better performance.[4]
- **Stacked LSTMs**, which consist of multiple hidden LSTM layers, can learn more complex and abstract temporal hierarchies from the data.[4]

Hybrid Models: The Best of Both Worlds

The most powerful architectures often combine the strengths of CNNs and RNNs. In a typical **hybrid CNN-RNN model**, a CNN front-end first extracts robust spatial or spatiotemporal features from the input (e.g., from topographic maps). The output of the CNN, which is a sequence of compact feature vectors, is then fed into an RNN back-end to model the temporal dynamics and dependencies within that sequence.[4] This synergistic approach has proven highly effective. For example, a study using a

3DCNN-BiLSTM model reported a 77.8% accuracy for word-pair classification, demonstrating the power of this combined spatiotemporal and sequential modeling strategy.[4]

3.3 The Transformer Ascendancy: A New Paradigm

More recently, the Transformer architecture, originally developed for natural language processing, has begun to revolutionize EEG analysis.[6] The core innovation of the Transformer is the

self-attention mechanism. Unlike RNNs, which process sequences step-by-step, the attention mechanism allows the model to weigh the influence of all other elements in a sequence simultaneously when processing a given element. This enables the parallel processing of sequences and a more effective way of capturing long-range dependencies, making Transformers highly efficient and powerful.[1]

While some studies report that Transformers train faster and outperform other deep learning models in EEG tasks [36], their superiority is not universal. One study on continuous speech recognition found that while a Transformer-based model trained faster, an RNN-based model achieved a lower Word Error Rate (WER) on test sets with larger vocabularies, indicating that the optimal architecture remains task-dependent.[39]

A truly transformative application of this architecture is its integration with pre-trained **Large Language Models (LLMs)** like GPT and BART.[6] In this paradigm, the goal of the EEG-specific model is no longer to directly classify words, but to translate the neural signals into a meaningful embedding (a vector representation) that an LLM can understand. The LLM then leverages its immense, pre-existing knowledge of language structure, grammar, and semantics to generate coherent, contextually appropriate, and open-vocabulary text. A landmark example,

BrainLLM, used a "brain adapter" network to map fMRI signals to an LLM's input space, enabling the generation of continuous text from brain activity rather than just selecting from a small, predefined set of words.[28] While demonstrated on fMRI, this principle is actively being extended to the more practical EEG modality.[28]

3.4 Generative AI: Synthesizing Reality from Brain Signals

The frontier of the field is now moving beyond classification and discriminative models toward **Generative Artificial Intelligence (GenAI)**, which focuses on creating new data.[6] In the context of BCI, this means generating rich outputs like images, synthesized speech, or text directly from brain signals.

Several classes of generative models are being employed:

- **Generative Adversarial Networks (GANs)** and **Variational Autoencoders (VAEs)** are primarily used for **data augmentation**. Given the chronic problem of small datasets in EEG research, GANs and VAEs can be trained to generate realistic, synthetic EEG data. This augmented data can then be used to train more robust and generalizable classification models.[1]
- **Diffusion Models** are another powerful, more recent class of generative models being explored for high-fidelity EEG data synthesis.[6]
- **Contrastive Learning** frameworks, such as **CLIP (Contrastive Language-Image Pre-training)**, are used to align the representations of data from different modalities. For instance, a model can be trained to learn a shared embedding space where the EEG signal recorded while a person views an image is located close to the text description of that image. This alignment facilitates

cross-modal generation, such as producing text or images from EEG inputs.[6]

The table below provides a summary of the evolution of these AI architectures and their roles in EEG speech decoding.

Model Class	Key Models	Primary Function / Strength	Key Limitations
Classical ML	SVM, LDA, Random Forest	Foundational; good for simple, low-vocabulary classification tasks. [9]	Requires extensive manual feature engineering; poor performance on complex, continuous tasks. [9]
CNN	1D-CNN, 2D-CNN, 3D-CNN	Automatic extraction of spatial and spatiotemporal features, especially from topographic maps. [4]	Struggles to capture long-range sequential dependencies on its own. [36]
RNN	LSTM, BiLSTM, Stacked LSTM	Models temporal and sequential dependencies effectively. [4]	Sequential processing can be slow; can struggle with very long-range dependencies. [31]
Hybrid CNN-RNN	3DCNN-BiLSTM, C-RNN	Combines spatial feature extraction (CNN) with sequential modeling (RNN) for robust performance. [4]	Can be complex to design and train; inherits limitations from both parent architectures.
Transformer	Transformer Encoder, BrainLLM	Captures long-range dependencies effectively via self-attention; enables parallel processing; integrates with LLMs for open-vocabulary generation. [6]	Can be data-hungry; performance on large-vocabulary EEG tasks vs. RNNs is still debated. [39]

Generative AI	GAN, VAE, Diffusion Models	Data augmentation to address small datasets; synthesis of multimodal outputs (speech, text, images). [1]	Training can be unstable; primarily used for data augmentation rather than direct decoding in most current studies.

Table 2: Evolution of AI Models in EEG Speech Decoding. This table outlines the progression of AI architectures, highlighting their specific strengths and weaknesses in the context of translating complex EEG signals into language.[1]

Section 4: State of the Art: Performance, Breakthroughs, and Bottlenecks

Evaluating the progress in EEG speech decoding requires a nuanced look at performance metrics, which have evolved alongside the complexity of the tasks. While early successes in simple classification were promising, the field's ambition has shifted toward continuous speech synthesis and open-vocabulary text generation, revealing both remarkable breakthroughs and persistent, fundamental bottlenecks. A critical analysis shows a clear tension between achieving high accuracy in controlled settings and developing models that are robust and generalizable enough for real-world use.

4.1 From Classification to Generation: The Evolving Goalposts

The history of EEG-BCI performance is one of steadily increasing ambition. Initial research focused on proving the basic feasibility of decoding through simple **classification tasks**. In these paradigms, models were trained to distinguish between a very small set of words or commands, typically between 2 and 5 items. For these strictly delimited tasks, reported accuracies were often high, ranging from 70% to over 90%.[13] For instance, studies have cited a Random Forest model achieving 94.6% accuracy on a pair-wise (two-word) classification task and an Artificial Neural

Network (ANN) reaching 66.92% on a multi-class problem.[4]

However, this performance is highly fragile and degrades rapidly as the complexity of the task increases. When the vocabulary size expands beyond a handful of words, accuracy plummets. For more realistic classification tasks, performance typically falls into the 20-50% range.[20] One study attempting to classify 50 different phrases reported an accuracy of only 5%, which, while better than chance, is far from practical.[29]

This limitation has driven a paradigm shift away from mere classification towards **generation**. The goal is no longer for a user to select from a predefined list but to freely express novel thoughts through continuously generated text or synthesized speech.[4] This represents a monumental leap in complexity, requiring models that can not only discriminate between a few known patterns but also synthesize novel, coherent outputs from the continuous stream of neural data.

4.2 Performance Benchmarking: A Sobering Look at the Numbers

Assessing the state of the art requires examining concrete performance metrics from recent landmark studies, keeping in mind that direct comparisons are often difficult due to variations in tasks, datasets, and methodologies.[4]

Speech Synthesis: The ultimate goal of a speech BCI is to produce audible, intelligible speech.

- The benchmark for this task was largely set by research using invasive **ECoG**, where Anumanchipalli et al. achieved a very low **Word Error Rate (WER) of 3%**, demonstrating that high-fidelity speech synthesis from neural signals is possible with high-quality input data.[18]
- A significant breakthrough in the non-invasive domain was an **EEG-to-Speech (ETS) system** that synthesized intelligible Chinese words from imagined speech EEG. This system achieved an average **word recognition accuracy of 91.23%** by human listeners and a **Mean Opinion Score (MOS) of 3.50** (out of 5), indicating good clarity and intelligibility. This was a landmark result, proving that audible and understandable speech could be generated directly from non-invasive EEG.[18]

Brain-to-Text Generation: Translating neural signals directly into written text is another major research thrust.

- The **DeWave model**, developed at the University of Technology Sydney, represents a state-of-the-art non-invasive system. It achieved a **BLEU-1 score of approximately 40%** when translating silently read text from EEG signals. The BLEU score measures the similarity between the model-generated text and a human reference, and while 40% is far from perfect, it marks a significant advance for open-vocabulary generation from noisy EEG.[6]
- Research from **Meta AI** using the higher-fidelity **MEG** modality demonstrated the ability to decode up to **80% of characters** correctly while participants were typing. This result, while not from EEG, highlights the potential of non-invasive signals when SNR is improved.[25]
- Another recent study, using a CLIP-based approach to align EEG and text representations, achieved a **top-1 accuracy of 48% on a 512-phrase open-vocabulary classification task**. This is an unprecedented result for EEG but comes with a major caveat: the data was collected from a single participant over an extensive 175-hour period, highlighting the trade-off between performance and data requirements.[43]

These results reveal a crucial pattern: the most impressive performance metrics are often achieved under highly specific and constrained conditions—such as using invasive signals, testing on a single subject with massive amounts of data, or simplifying the task to a small vocabulary. This underscores the challenge of **generalizability**, which remains the field's primary obstacle to real-world deployment.

4.3 Landmark Systems and Frameworks

Several recently developed systems exemplify the cutting edge of AI-driven EEG decoding:

- **FESDE (Fully-End-to-end Speech Decoding):** This framework is notable for its architectural philosophy. By aiming to directly reconstruct speech waveforms from EEG signals, it bypasses intermediate steps like converting the signal to a mel-spectrogram. Its architecture, comprising a dedicated EEG module, a speech module built on TTS technology, and a "connector" to bridge them, allows for a simpler and more efficient single-step inference process.[19]
- **ClinClip:** This model demonstrates the power of multimodality. By integrating EEG signals with audio data using a Transformer architecture, ClinClip can improve the

accuracy of speech transcription in noisy environments. It dynamically adjusts to the listener's cognitive state (as measured by EEG) to achieve a lower Word Error Rate (WER) than audio-only systems, showcasing a practical application in complex settings like medical listening assessments.[49]

- **Hybrid 3DCNN-RNN Frameworks:** This class of models highlights the effectiveness of treating EEG as image sequences. By converting multi-channel EEG data into topographic brain maps and feeding them into a 3D-CNN for spatiotemporal feature extraction, followed by an RNN (like BiLSTM) for sequence modeling, these systems have achieved high accuracies (e.g., 77.8% for word-pair classification) on imagined speech tasks.[4]

4.4 The Data Bottleneck: The Field's Achilles' Heel

Across nearly all studies, the most consistently cited limitation is the lack of large, diverse, and standardized datasets. This "data bottleneck" is the single greatest impediment to progress and manifests in several ways:

- **Small and Heterogeneous Datasets:** The vast majority of studies are conducted on a very small number of participants, often fewer than 20, and sometimes as few as four.[4] This makes it difficult to train deep learning models that can generalize beyond the specific individuals in the training set.[1]
- **Inter-Subject Variability:** Every individual's brain is unique, and the neural patterns associated with speech vary significantly from person to person. A model trained on one subject's data often performs at or near chance level on another's data. This necessitates lengthy, subject-specific calibration sessions, which is a major barrier to practical, "out-of-the-box" BCI systems.[1]
- **Inconsistent Preprocessing and Benchmarking:** The lack of standardized preprocessing pipelines and common benchmark datasets makes it extremely difficult to compare the performance of different models and algorithms across studies. This fragmentation slows down collective progress, as it is often unclear whether a reported improvement is due to a superior model architecture or a difference in data handling.[4]

Recognizing this critical need, the research community has begun to create and share public datasets to facilitate more robust and reproducible research. The table below lists some of the key publicly available datasets used for speech decoding tasks.

Dataset Name/Identifier	Task Type	# of Subjects	# of Channels	Key Stimuli	Source/Link
BCI2020 dataset [15]	Imagined Speech	15	64	5 English words/phrases ("Hello", "Help me", etc.)	[15]
OpenNeuro ds006104	Listened Speech (Phoneme Discrimination)	24	64	Single phonemes, CV/VC pairs, words, pseudowords	[51]
Kumar's EEG Imagined Speech	Imagined Speech	23	14	10 characters, 10 digits, 10 object images	[38]
SparrKULee	Listened Speech (Continuous)	85	64	90–150 min of natural speech per participant	[52]
EEG Speech-Robot Interaction	Overt & Imagined Speech	15	64	5 command words ("Left", "Right", "Pick", etc.)	[54]
Zenodo Auditory Attention	Listened Speech (Attentional Decoding)	18	64	Competing continuous speech streams	[56]

Table 3: Publicly Available Datasets for EEG Speech Decoding. This table provides a resource for researchers by summarizing key public datasets, which are crucial for benchmarking new models and addressing the challenge of generalizability.[6]

Section 5: The Path Forward: Applications, Challenges, and the

Future of Neural Speech Interfaces

As the field of EEG-based speech decoding matures, its trajectory is shaped by the immense promise of its applications, the formidable technical and practical challenges that remain, and the profound ethical questions it raises. The ultimate success of this technology will depend not on a single breakthrough, but on the synergistic convergence of advancements in artificial intelligence, sensor hardware, and large-scale data initiatives.

5.1 Clinical and Commercial Applications: The Ultimate Goal

The driving force behind much of this research is the potential to revolutionize assistive technology and human-computer interaction.

- **Assistive Communication:** The primary and most compelling application is restoring the ability to communicate for individuals with severe speech and motor impairments, such as those with ALS, cerebral palsy, stroke, or locked-in syndrome.[1] The goal is to evolve beyond slow, cumbersome speller-based interfaces—which rely on selecting letters one by one—to systems that can generate natural, fluid speech in real-time, thereby restoring a fundamental aspect of human connection.[17]
- **Augmented and Virtual Reality (AR/VR):** As AR and VR technologies become more mainstream, BCIs offer a new frontier for interaction. Instead of relying on manual controllers, users could navigate virtual worlds, manipulate objects, and interact with digital content simply by thinking, enabling a truly seamless and intuitive form of human-machine symbiosis.[6]
- **Silent Communication:** The ability to decode imagined speech opens up applications in environments where audible communication is impractical or unsafe. This includes high-noise settings, such as a factory floor or a military combat zone, where clear vocal communication is difficult, or scenarios requiring privacy and discretion.[1]
- **Broader Applications:** Beyond these core areas, EEG-based decoding has potential applications in gaming (thought-controlled characters), education (monitoring cognitive load and engagement), and mental health (as a diagnostic tool for assessing cognitive states).[3]

5.2 Overcoming the Hurdles: The Roadmap to Real-World Viability

For these applications to become a reality, several significant technical and practical challenges must be overcome.

- **Improving Generalizability and Reducing Calibration:** This is arguably the most critical hurdle. Current high-performing models are often subject-specific, requiring extensive and tedious calibration for each new user. The development of robust, **subject-independent models** is essential for practical deployment. This will require a concerted effort to build larger and more diverse public datasets, along with the development of advanced **transfer learning** and **self-supervised learning** techniques that can adapt a pre-trained model to a new user with minimal data.[4]
- **Achieving Real-Time Performance:** For naturalistic communication, the delay between a thought and its decoded output (latency) must be minimized. While some systems have demonstrated impressive latency reduction—for example, one study reduced an 8-second delay for a full sentence to under 1 second for the first sound—maintaining this speed while increasing accuracy and vocabulary size remains a major engineering challenge. This necessitates the development of computationally efficient AI models and optimized streaming data processing pipelines.[17]
- **Advancements in EEG Hardware:** The future of BCI is wearable and unobtrusive. The bulky, gel-based electrode caps used in laboratory settings are impractical for daily use. The path forward lies in the development of next-generation sensor technology. This includes **dry and semi-dry electrodes** made from novel materials like conductive polymers, flexible silicone, or graphene, which do not require conductive gel and offer greater comfort and ease of use.[10] The ultimate goal is to seamlessly integrate these high-fidelity sensors into everyday objects like headphones, earbuds, glasses, or hats, making continuous brain monitoring a practical reality.[57]
- **Enhancing Expressivity:** Current decoding efforts are overwhelmingly focused on the linguistic content of speech. A crucial next step for restoring truly natural communication is the ability to decode **paralinguistic features**. This includes the prosody of speech—the tone, pitch, loudness, and emotion—that conveys a huge amount of meaning beyond the words themselves.[17]

5.3 Ethical and Privacy Imperatives: The Unseen Challenges

As BCI technology advances from decoding simple motor commands to interpreting complex thoughts, it enters a sensitive ethical landscape that demands careful navigation. The ability to "read minds" is no longer purely the domain of science fiction, and its development carries profound societal responsibilities.

- **Mental Privacy:** This is the foremost concern. EEG data is a direct, albeit noisy, window into a person's cognitive and emotional state. The potential for this technology to be used for surveillance, to infer personal beliefs, or to access a person's inner thoughts without explicit and fully informed consent raises fundamental questions about the right to cognitive liberty and the final frontier of privacy.[6]
- **Data Security and Ownership:** EEG data is arguably the most sensitive category of personal information. As commercial entities like Neuralink and Meta invest heavily in this space, clear and robust regulations are needed to govern data ownership, security, and use. Research has shown that even "anonymized" EEG patterns can be used to re-identify individuals, necessitating advanced privacy-preserving techniques like differential privacy and secure encryption to prevent misuse.[6]
- **Consent and Autonomy:** The nature of informed consent becomes highly complex, particularly when dealing with vulnerable populations who may be the primary beneficiaries of this technology. Furthermore, bidirectional BCIs—which can not only read from but also write to the brain—raise profound questions about personal identity and autonomy. The potential for such systems to subtly influence a user's thoughts or decisions must be carefully considered and safeguarded against.[6]
- **Bias and Inequality:** There is a significant risk that BCI technology could exacerbate existing social inequalities. If access to high-performance systems is limited to the wealthy, it could create a "neuro-divide." Moreover, AI models trained on insufficiently diverse datasets may perform poorly for underrepresented populations, leading to algorithmic bias that could further marginalize certain groups.[6]

5.4 Conclusion and Future Outlook

The field of AI-powered speech decoding from EEG signals stands at a thrilling and critical juncture. Driven by the confluence of neuroscience and machine learning, research has progressed from the simple classification of isolated words to the demonstrated synthesis of intelligible speech and the generation of continuous text from non-invasive brain recordings. The development of sophisticated deep learning architectures—particularly hybrid CNN-RNN models and, more recently, Transformers integrated with large language models—has been pivotal in overcoming the formidable challenge of extracting meaningful information from noisy EEG data.

Despite these advances, the path to widespread, practical application remains steep. The primary obstacles are not conceptual but practical: the critical lack of large, diverse, and standardized datasets continues to hamper the development of generalizable models that can perform reliably across different individuals without extensive recalibration. The performance of current systems is a direct function of this data bottleneck; impressive headline accuracies are often achieved only under highly constrained laboratory conditions that do not reflect real-world variability.

The future of this transformative technology lies at the intersection of three key domains of innovation. First, **algorithmic advancement** will continue, with a focus on more sophisticated generative and self-supervised models that can learn robust representations from limited or unlabeled data. Second, **hardware and sensor technology** must evolve toward comfortable, reliable, and wearable dry-electrode systems that can be seamlessly integrated into daily life. Third, and most importantly, the community must continue to build and share **large-scale, public datasets** to enable the rigorous benchmarking and training of truly generalizable AI.

As these technological frontiers are pushed, the profound ethical questions they raise must be addressed with equal rigor and foresight. The ultimate goal of this research is not merely to decode words from brainwaves, but to restore the fundamental human capacity for communication, connection, and self-expression. Achieving this goal responsibly requires a balanced and concerted effort to advance the science while proactively building the ethical and regulatory frameworks needed to ensure this powerful technology serves humanity equitably and safely.

Works cited

1. Decoding EEG Speech Perception with Transformers and VAE-based Data

Augmentation, accessed June 16, 2025, https://arxiv.org/html/2501.04359v1
2. Bridging Brain Signals and Language: A Deep Learning Approach to EEG-to-Text Decoding, accessed June 16, 2025, https://arxiv.org/html/2502.17465v1
3. Imagined Speech Using EEG Signals - MIT Solve, accessed June 16, 2025, https://solve.mit.edu/challenges/cure-challenge/solutions/81583
4. Speech Decoding from EEG Signals - The Science and Information (SAI) Organization, accessed June 16, 2025, https://thesai.org/Downloads/Volume16No4/Paper_85-Speech_Decoding_from_EEG_Signals.pdf
5. Invasive Brain–Computer Interface for Communication: A Scoping Review - MDPI, accessed June 16, 2025, https://www.mdpi.com/2076-3425/15/4/336
6. arxiv.org, accessed June 16, 2025, https://arxiv.org/html/2502.12048v1
7. Decoding Covert Speech From EEG-A Comprehensive Review - PMC - PubMed Central, accessed June 16, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC8116487/
8. Unlocking Potential with BCI, accessed June 16, 2025, https://www.numberanalytics.com/blog/ultimate-guide-bci-assistive-technology
9. A State-of-the-Art Review of EEG-Based Imagined Speech Decoding - PMC, accessed June 16, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC9086783/
10. Decoding Covert Speech From EEG-A Comprehensive Review - Frontiers, accessed June 16, 2025, https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2021.642251/full
11. EEG-Based Silent Speech Interface and its Challenges: A Survey - The Science and Information (SAI) Organization, accessed June 16, 2025, https://thesai.org/Downloads/Volume13No11/Paper_73-EEG_Based_Silent_Speech_Interface_and_its_Challenges.pdf
12. The Role of Artificial Intelligence in Decoding Speech from EEG Signals: A Scoping Review, accessed June 16, 2025, https://www.mdpi.com/1424-8220/22/18/6975
13. Improving Electroencephalography-Based Imagined Speech Recognition with a Simultaneous Video Data Stream - ScholarWorks@UARK, accessed June 16, 2025, https://scholarworks.uark.edu/cgi/viewcontent.cgi?article=1038&context=csceuht
14. Neural Decoding of EEG Signals with Machine Learning: A Systematic Review - MDPI, accessed June 16, 2025, https://www.mdpi.com/2076-3425/11/11/1525
15. Decoding Imagined Speech from EEG Data: A Hybrid Deep Learning Approach to Capturing Spatial and Temporal Features - MDPI, accessed June 16, 2025, https://www.mdpi.com/2075-1729/14/11/1501
16. Speech decoding from stereo-electroencephalography (sEEG) signals using advanced deep learning methods - PubMed, accessed June 16, 2025, https://pubmed.ncbi.nlm.nih.gov/38885688/
17. Brain-to-voice neuroprosthesis restores naturalistic speech - Berkeley Engineering, accessed June 16, 2025, https://engineering.berkeley.edu/news/2025/03/brain-to-voice-neuroprosthesis-restores-naturalistic-speech/

18. Synthesizing intelligible utterances from EEG of imagined ... - Frontiers, accessed June 16, 2025, https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2025.1565848/pdf
19. Toward Fully-End-to-End Listened Speech Decoding from EEG Signals - arXiv, accessed June 16, 2025, https://arxiv.org/pdf/2406.08644
20. Decoding imagined speech with delay differential analysis - Frontiers, accessed June 16, 2025, https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2024.1398065/full
21. Neural Decoding of Spontaneous Overt and Intended Speech - ASHA Journals, accessed June 16, 2025, https://pubs.asha.org/doi/10.1044/2024_JSLHR-24-00046
22. Automatic Speech Recognition from Neural Signals: A Focused Review - Frontiers, accessed June 16, 2025, https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2016.00429/full
23. Differences between EEG, NIRS, fMRI and MEG - Brain Latam, accessed June 16, 2025, https://brainlatam.com/blog/differences-between-eeg-nirs-fmri-and-meg-924
24. (PDF) Decoding Performance for Hand Movements: EEG vs. MEG - ResearchGate, accessed June 16, 2025, https://www.researchgate.net/publication/224292130_Decoding_Performance_for_Hand_Movements_EEG_vs_MEG
25. Using AI to decode language from the brain and advance our understanding of human communication - Meta AI, accessed June 16, 2025, https://ai.meta.com/blog/brain-ai-research-human-communication/
26. Automatic Speech Recognition from Neural Signals: A Focused Review - Frontiers, accessed June 16, 2025, https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2016.00429/pdf
27. Concurrent fNIRS and EEG for Brain Function Investigation: A Systematic, Methodology-Focused Review - MDPI, accessed June 16, 2025, https://www.mdpi.com/1424-8220/22/15/5865
28. Researchers train AI to read minds—by decoding brain signals into ..., accessed June 16, 2025, https://www.news-medical.net/news/20250304/Researchers-train-AI-to-read-mindse28094by-decoding-brain-signals-into-text.aspx
29. AN EMPIRICAL STUDY OF SPEECH PROCESSING IN THE BRAIN BY ANALYZING THE TEMPORAL SYLLABLE STRUCTURE IN SPEECH-INPUT INDUCED EEG Ri - DSpace@MIT, accessed June 16, 2025, https://dspace.mit.edu/bitstream/handle/1721.1/138047.2/syllable_classification_from_EEG.pdf?sequence=4
30. Using Brain Waves and Computer Interface Technology as a Communication System, accessed June 16, 2025,

https://www.gtec.at/2025/03/10/using-brain-waves-as-a-communication-system/
31. A spatial and temporal transformer-based EEG emotion recognition in VR environment, accessed June 16, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC11897567/
32. A spatial and temporal transformer-based EEG emotion recognition in VR environment, accessed June 16, 2025, https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2025.1517273/full
33. EEG-Based Speech Decoding Using a Machine Learning Pipeline - DiVA portal, accessed June 16, 2025, https://www.diva-portal.org/smash/get/diva2:1731611/FULLTEXT01.pdf
34. (PDF) Decoding Imagined Speech from EEG Using Transfer Learning - ResearchGate, accessed June 16, 2025, https://www.researchgate.net/publication/354820101_Decoding_Imagined_Speech_from_EEG_Using_Transfer_Learning
35. Exploratory methods for high-performance EEG speech decoding - Salk Institute, accessed June 16, 2025, https://papers.cnl.salk.edu/PDFs/Exploratory%20methods%20for%20high-performance%20EEG%20speech%20decoding%202021-4628.pdf
36. Transformers in EEG Analysis: A Review of Architectures and Applications in Motor Imagery, Seizure, and Emotion Classification - MDPI, accessed June 16, 2025, https://www.mdpi.com/1424-8220/25/5/1293
37. Toward Fully-End-to-End Listened Speech Decoding from ... - arXiv, accessed June 16, 2025, https://arxiv.org/abs/2406.08644
38. Kumar's EEG Imagined speech - Kaggle, accessed June 16, 2025, https://www.kaggle.com/datasets/ignazio/kumars-eeg-imagined-speech
39. EEG based Continuous Speech Recognition using Transformers, accessed June 16, 2025, https://arxiv.org/abs/2001.00501
40. Decode Neural signal as Speech - arXiv, accessed June 16, 2025, https://arxiv.org/html/2403.01748v1
41. Virtual Electroencephalogram Acquisition: A Review on Electroencephalogram Generative Methods - MDPI, accessed June 16, 2025, https://www.mdpi.com/1424-8220/25/10/3178
42. Generative modeling and augmentation of EEG signals using improved diffusion probabilistic models - PubMed, accessed June 16, 2025, https://pubmed.ncbi.nlm.nih.gov/39693767/
43. Scaling Law in Neural Data: Non-Invasive Speech Decoding with 175 Hours of EEG Data, accessed June 16, 2025, https://arxiv.org/html/2407.07595v1
44. (PDF) Text and image generation from intracranial electroencephalography using an embedding space for text and images - ResearchGate, accessed June 16, 2025, https://www.researchgate.net/publication/380024005_Text_and_image_generation_from_intracranial_electroencephalography_using_an_embedding_space_for_text_and_images

45. [2302.01736] Relating EEG to continuous speech using deep neural networks: a review, accessed June 16, 2025, https://arxiv.org/abs/2302.01736
46. Rethinking the Methods and Algorithms for Inner Speech Decoding and Making Them Reproducible - PMC, accessed June 16, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC11523721/
47. Portable, non-invasive, mind-reading AI turns thoughts into text - University of Technology Sydney, accessed June 16, 2025, https://www.uts.edu.au/news/2023/12/portable-non-invasive-mind-reading-ai-turns-thoughts-text
48. AI and EEG Transform Silent Thoughts to Text - Neuroscience News, accessed June 16, 2025, https://neurosciencenews.com/ai-eeg-thoughts-to-text-25343/
49. ClinClip: a Multimodal Language Pre-training model integrating EEG data for enhanced English medical listening assessment - PubMed, accessed June 16, 2025, https://pubmed.ncbi.nlm.nih.gov/39850622/
50. Systematic Review of EEG-Based Imagined Speech Classification Methods - MDPI, accessed June 16, 2025, https://www.mdpi.com/1424-8220/24/24/8168
51. EEG dataset for speech decoding - OpenNeuro, accessed June 16, 2025, https://openneuro.org/datasets/ds006104
52. SparrKULee: A Speech-Evoked Auditory Response Repository from ..., accessed June 16, 2025, https://www.mdpi.com/2306-5729/9/8/94
53. SparrKULee: A Speech-evoked Auditory Response Repository of the KU Leuven, containing EEG of 85 participants | bioRxiv, accessed June 16, 2025, https://www.biorxiv.org/content/10.1101/2023.07.24.550310v1.full-text
54. EEG Speech-Robot Interaction Dataset Dataset | Papers With Code, accessed June 16, 2025, https://paperswithcode.com/dataset/eeg-speech-robot-interaction-dataset
55. EEG data recorded during spoken and imagined speech interaction with a simulated robot, accessed June 16, 2025, https://zenodo.org/records/14645653
56. EEG and audio dataset for auditory attention decoding - Zenodo, accessed June 16, 2025, https://zenodo.org/records/1199011
57. Why Consumer EEG Embedded Hardware Is the Next Big Platform: My Predictions as a CTO | Arctop Deep Dives, accessed June 16, 2025, https://arctop.com/deep-dives/consumer-eeg-hardware
58. No longer science fiction: Mind reading through EEG could soon become reality - Frontiers, accessed June 16, 2025, https://www.frontiersin.org/news/2021/04/29/frontiers-mind-reading-eeg-electroencephalography-panachakel-ganesan-indian-institute-of-science
59. Brain Computer Interfaces: Assistive Technology Meets Consumer Devices | IDTechEx Research Article, accessed June 16, 2025, https://www.idtechex.com/en/research-article/brain-computer-interfaces-assistive-technology-meets-consumer-devices/32318
60. Advancements in dry and semi-dry EEG electrodes: Design, interface characteristics, and performance evaluation - AIP Publishing, accessed June 16, 2025, https://pubs.aip.org/adv/article/15/4/040703/3345166/Advancements-in-dry-and-

semi-dry-EEG-electrodes
61. Next-Generation EEG Hardware - Zander Labs, accessed June 16, 2025, https://www.zanderlabs.com/blog/next-generation-eeg-hardware
62. Noninvasive Sensors for Brain–Machine Interfaces Based on Micropatterned Epitaxial Graphene | ACS Applied Nano Materials - ACS Publications, accessed June 16, 2025, https://pubs.acs.org/doi/10.1021/acsanm.2c05546
63. Brain Recording, Mind-Reading, and Neurotechnology: Ethical Issues from Consumer Devices to Brain-Based Speech Decoding - PubMed Central, accessed June 16, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7417394/
64. Ethics of neurotechnology - UNESCO, accessed June 16, 2025, https://www.unesco.org/en/ethics-neurotech
65. Full article: Neurorights: The Land of Speculative Ethics and Alarming Claims?, accessed June 16, 2025, https://www.tandfonline.com/doi/full/10.1080/21507740.2024.2328244
66. Ethics, Human Rights & Privacy in NeuroTech using EEG Brain Data & (BCi) Brain Computer Interfaces - ResearchGate, accessed June 16, 2025, https://www.researchgate.net/publication/370068015_Ethics_Human_Rights_Privacy_in_NeuroTech_using_EEG_Brain_Data_BCi_Brain_Computer_Interfaces
67. Neurodata Consent Frameworks: Managing EEG/Brain-Computer Interface Data Under GDPR/CCPA - Secure Privacy, accessed June 16, 2025, https://secureprivacy.ai/blog/neurodata-consent-eeg-brain-computer-interface-data-gdpr-ccpa
68. EEG Data Privacy Enhancement using Differential Privacy in WGAN-based Federated Learning - ResearchGate, accessed June 16, 2025, https://www.researchgate.net/publication/377333161_EEG_Data_Privacy_Enhancement_using_Differential_Privacy_in_WGAN-based_Federated_Learning
69. Protecting Multiple Types of Privacy Simultaneously in EEG-based Brain-Computer Interfaces - arXiv, accessed June 16, 2025, https://arxiv.org/html/2411.19498v1