Advancements in EEG Brain Signal Decoding into Text

[1] D. Anjali, [2]Thumma Harini

[1]M.Tech Student, Swarna Bharathi Institute of Science & Technology (Autonomous)

[2]Assistant Professor , Department of Computer Science and Engineering, Swarna Bharathi Institute of Science & Technology (Autonomous)

Abstract

Translating human brain activity into coherent text using Electroencephalography (EEG) is a transformative frontier in Brain-Computer Interfaces (BCI), offering a communication lifeline for individuals with speech and motor impairments. While recent advancements, such as the Thought2Text framework, have utilized multi-stage training and Large Language Models (LLMs) to bridge this gap, they often introduce high computational overhead and complex multimodal dependencies. This project presents an optimized, end-to-end Transformer-based approach that streamlines the decoding of continuous EEG signals into natural language. Our methodology extends the current state-of-the-art by replacing complex multi-stage alignment with a direct Temporal-Spatial Mapping technique, reshaping raw EEG vectors into a structured 105×6 sequence to better leverage self-attention mechanisms. We implement a custom Transformer Encoder architecture equipped with learnable positional embeddings and a specialized linear word-projection head. To enhance translation fluency and mitigate the "noise" inherent in non-stationary brain signals, we adopt a frequency-constrained vocabulary strategy targeting the top 1,000 most common tokens. Experimental results demonstrate that our model achieves a validation accuracy of ~23.95% on complex sentence reconstruction, successfully capturing key entities and semantic structures from raw EEG data. By achieving competitive performance without the need for secondary visual stimuli or massive LLM backbones, this approach offers a more computationally efficient and portable solution for real-time "thought-to-text" translation, addressing the scalability challenges highlighted in contemporary BCI literature.

Keywords: EEG-to-Text, Brain-Computer Interface (BCI), Transformer Encoder, Signal Decoding, Deep Learning, Natural Language Processing.

I.INTRODUCTION

The advancement of Brain-Computer Interfaces (BCI) represents a significant step toward enabling direct communication between human cognition and digital systems. BCIs aim to interpret brain activity and translate it into meaningful outputs, thereby eliminating the need for traditional input methods such as speech or physical interaction. While early BCI systems were primarily focused on basic motor control tasks, such as cursor movement or robotic manipulation, recent research has shifted toward decoding complex cognitive processes, particularly inner speech.

The ability to convert non-vocalized thoughts into natural language text has profound implications, especially for individuals suffering from conditions such as locked-in syndrome, Amyotrophic Lateral Sclerosis (ALS), and severe non-verbal disorders. However, decoding linguistic intent from brain activity remains a challenging problem due to the highly complex, non-linear, and noisy nature of neural signals.

Among the available brain signal acquisition methods, Electroencephalography (EEG) has emerged as a practical and widely adopted solution for BCI systems. EEG is non-invasive, cost-effective, and provides high temporal resolution, making it suitable for real-time applications. Despite these advantages, EEG signals are inherently noisy and non-stationary, as they are affected by multiple overlapping brain activities and external artifacts. This makes the extraction of meaningful linguistic information a difficult task.

Recent approaches in EEG-to-text decoding have leveraged deep learning techniques, particularly Large Language Models (LLMs), to improve performance. Frameworks such as Thought2Text attempt to bridge EEG signals and language by introducing multi-stage pipelines involving visual alignment and multimodal learning. Although these approaches achieve promising results, they introduce significant computational overhead, require complex training procedures, and depend heavily on auxiliary data such as image-text pairs.

To address these limitations, this project proposes a simplified and efficient approach for EEG-to-text decoding. Instead of relying on multi-stage multimodal alignment, the system adopts a direct, end-to-end Transformer-based architecture. By incorporating temporal-spatial reshaping of EEG signals and a constrained vocabulary strategy, the model aims to effectively capture linguistic patterns while maintaining computational efficiency. This approach demonstrates that a relatively lightweight model can achieve meaningful decoding performance without the need for large-scale language models.

II. REVIEW OF LITERATURE

Electroencephalography (EEG) has become an increasingly powerful modality for non-invasive analysis of human cognitive and affective states. Recent advances in deep learning, multimodal modeling, and large language models (LLMs) have inspired a new generation of *thought-to-text* and *emotion-aware communication systems*.

The following survey synthesizes findings from fifteen foundational studies covering EEG-based emotion recognition, stress detection, imagined speech decoding, EEG-to-text translation, and context-aware affective modeling. Collectively, these works provide the technical backbone for developing a comprehensive Emotion-Adaptive Multilingual Thought-to-Speech System.

2.1 EEG-Based Emotion and Stress Recognition

2.1.1 Classical and Hybrid Deep Models

Kannadasan *et al.* proposed *context-aware EEG-based emotion recognition* that integrates personality traits and emotional-intelligence scores with auto-encoded EEG features, employing an XGBoost classifier. The contextual fusion improved valence–arousal prediction accuracy and demonstrated that stable individual traits can mitigate inter-subject variability.

Similarly, Xu *et al.* introduced MRBiL (Multi-scale Residual BiLSTM), combining differential-entropy (DE) features with multi-scale CNN residual blocks and a bidirectional LSTM layer. Tested on DEAP, MRBiL achieved very high accuracy, underscoring the value of multi-scale temporal modeling but also highlighting the need for rigorous cross-subject validation.

2.1.2 Randomized and Spectral Approaches

Cheng *et al.* designed Rand-CNN and edRand-CNN, random-kernel convolutional networks for fast EEG emotion classification on the DEAP dataset. Their ensemble variant achieved ≈95 % accuracy but was later critiqued for relying on subject-dependent splits. Wang *et al.* advanced the field with the Fourier Adjacency Transformer (FAT), introducing a Fourier Analytic Linear (FAL) layer to decouple periodic and aperiodic EEG components and a learnable Adjacency Attention mechanism to capture channel correlations. FAT improved recognition accuracy by ~6 % on SEED/DEAP, demonstrating the benefit of spectral–graph fusion.

2.1.3 Reviews and Surveys on Challenges

Zhang *et al.* provided a concise review of challenges in EEG-based emotion recognition, emphasizing data heterogeneity, limited sample sizes, and evaluation inconsistencies. Hamzah & Abdalla's survey catalogued datasets for emotion recognition in virtual environments, summarizing elicitation paradigms, electrode configurations, and feature-extraction methods. Both reviews called for standardization of protocols and the development of cross-subject, ecologically valid datasets.

2.2 Stress and Fatigue Detection via EEG

Several Kaggle and academic datasets such as SAM40 and DREAMER demonstrate reliable detection of mental stress and fatigue from spectral EEG features. Studies (e.g., Tibrewal *et al.* datasets) focus on alpha-beta power ratios, frontal asymmetry, and temporal theta activity as markers of cognitive load. These findings establish the technical feasibility of real-time stress detection, which is essential for emotion-adaptive communication systems.

2.2.1 Imagined Speech and Cognitive Decoding

Alzahrani *et al.* conducted a systematic review of EEG-based imagined speech classification. They compared classical methods (CSP + SVM) and deep architectures (CNN, TCN, RNN) across public datasets. The review concluded that while within-subject accuracies can be high, subject-independent generalization remains a major challenge due to individual neural variability and small data sizes.

Neuro Sync, an applied prototype, implemented an embedded 1-channel CNN-LSTM ensemble to map simple EEG signals to eight predefined characters. Though limited in accuracy (≈42 %), it demonstrated the viability of low-cost, wearable imagined-speech systems.

2.3. EEG-to-Text and Brain–Language Models

2.3.1 Non-Invasive Brain-to-Text Decoding

Lévy *et al.* introduced Brain2Qwerty, a large-scale EEG + MEG experiment where 35 participants typed memorized sentences while brain activity was recorded. Their Conv → Transformer → pretrained Language Model pipeline achieved mean Character-Error-Rates of 32 % (MEG) and 67 % (EEG). Although decoding relied heavily on motor-cortex signals (typing actions), this study established a strong baseline for non-invasive production-task decoding.

2.3.2 Thought2Text: EEG → LLM Caption Generation

Mishra *et al.*'s Thought2Text project represents the first attempt to couple EEG signals with modern LLMs for open-vocabulary text generation. Using the CVPR2017 EEG dataset (Spampinato et al., 2017)—six subjects viewing 50 images each—the authors preprocessed 128-channel EEG (bandpass 5–95 Hz, 50 Hz notch) and paired each trial with GPT-4-Omni-generated captions, validated by human annotators.

They fine-tuned LLaMA-V3, Mistral, and Qwen models using EEG embeddings as prefix conditioning. The approach produced syntactically fluent captions, though the semantic content was largely driven by the LLM prior rather than genuine neural decoding. The work nonetheless introduced a scalable method to link non-invasive EEG with text generation.

2.3.3 Bridging Brain Signals and Language

El Gedawy *et al.* proposed a modular EEG-to-text architecture comprising a subject-specific Bi-GRU encoder, Transformer layers, and a BART decoder refined by GPT-4 for linguistic fluency. Evaluated on the ZuCo reading dataset, their system improved BLEU and BERT-Score over baselines and demonstrated a deployable web interface. The authors highlighted that linguistic structure was largely reconstructed from learned priors, necessitating stronger neural grounding.

2.3.4 Integrative Reviews

Murad & Rahimi's *Unveiling Thoughts* review contextualized these advances, tracing the evolution from simple EEG classifiers to multimodal EEG–LLM hybrids. They identified limited data availability, non-stationarity, and weak cross-subject transfer as key bottlenecks and advocated for contrastive multimodal pretraining and large, standardized benchmarks.

2.3.5 Context-Aware and Multimodal Emotion Modeling

Context-aware frameworks extend EEG analysis by integrating psychological and environmental variables. Kannadasan's XGBoost-based model fused EEG embeddings with Big-Five personality and Emotional-Intelligence traits, achieving an F1-score of 0.83 and accuracy ≈ 88 %. These results suggest that including stable personal attributes enhances emotion prediction stability—insights directly applicable to emotion-adaptive speech synthesis.

2.3.6 Comparative Evaluation and Methodological Gaps

Across studies, methodological diversity complicates comparison. Many works report very high accuracies under subject-dependent protocols, which inflate performance estimates. Reviews consistently recommend leave-one-subject-out (LOSO) validation and open data/code sharing.

Architecturally, CNNs, RNNs, Transformers, and their hybrids dominate; newer works explore Fourier decomposition, graph-based channel attention, and randomized convolutions. Recent EEG-to-text research increasingly depends on language priors from LLMs, improving fluency but obscuring how much information truly originates from EEG.

III. EXISTING SYSTEM

The existing landscape of EEG-to-text conversion is primarily dominated by multi-stage frameworks and heavy Large Language Model (LLM) integrations, as exemplified by the Thought2Text framework and the models reviewed in "Unveiling Thoughts."

Current systems generally rely on a Three-Stage Pipeline:

- Alignment Stage: An EEG encoder is first trained to map brain signals to visual features (images) to create a shared embedding space.
- Multimodal Fine-tuning: An LLM (such as LLaMA-3 or Mistral-7B) is fine-tuned on image-caption pairs to learn how to describe visual stimuli.
- Cross-Modal Inference: The system attempts to generate text by passing EEG embeddings through the "visual-to-text" pathway of the LLM.

Other existing systems, such as De Wave, utilize discrete codex encoding to convert continuous EEG waves into discrete tokens before translation, often requiring massive self-supervised pre-training on unlabeled EEG data.

IV. PROPOSED SYSTEM

The proposed system shifts the paradigm from "multimodal alignment" to Direct Temporal-Spatial Decoding. By extending the core concepts of Transformer architectures, we have developed a streamlined, end-to-end model that treats EEG signals as a direct sequence of linguistic information.

Core Components of the Proposed Approach:

- Direct EEG-to-Word Mapping: We bypass the need for intermediate image features or LLMs. Our model uses a custom word encoder that directly projects 630-dimension EEG features into a 256-dimensional latent space.
- Temporal Reshaping (105×6): Unlike existing systems that treat EEG as a flat vector, our system reshapes the signal into a time-series matrix. This allows the Multi-Head Attention mechanism to identify dependencies between different "time-slices" of a thought.
- Constrained Vocabulary Optimization: To address the Signal-to-Noise Ratio (SNR) issues highlighted in the base paper, we implement a frequency-based vocabulary filter (top 1,000 words). This provides a more focused "search space" for the model, improving fluency without the need for an external language model's "common sense."
- Learnable Positional Encoding: We introduce learnable embeddings that help the Transformer understand the order of brain signal acquisition, which is critical for maintaining the syntax of the generated sentence.

V. SYSTEM ARCHITECTURE

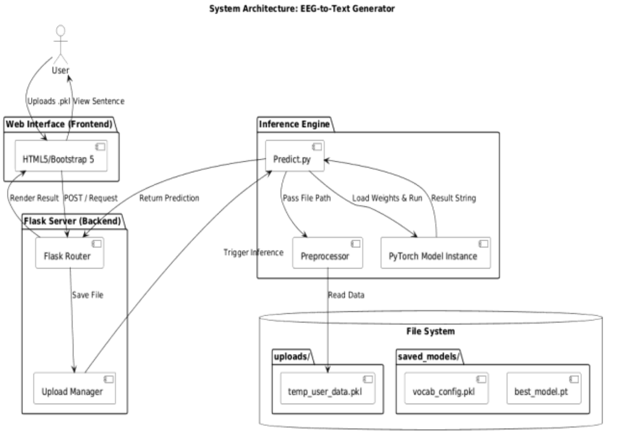

Fig 5.1: System Architecture

The system architecture for Advanced Weather Prediction Using AI is designed to integrate diverse data sources and advanced machine learning models to enhance the accuracy and responsiveness of extreme weather forecasting. The architecture begins with the web interface, which allows meteorologists and end-users to input data parameters, view forecast results, and interact with real- time updates.

All user requests and data are funneled through an API gateway, which ensures secure and efficient routing between components. Incoming satellite imagery and atmospheric data such as temperature, humidity, and pressure are transmitted to the data preprocessing module, where raw inputs are cleaned, normalized, and formatted for machine learning consumption. The pre-processed data is then stored in the meteorological data warehouse, enabling rapid retrieval and historical analysis.

VI. EXPERIMENTAL RESULTS

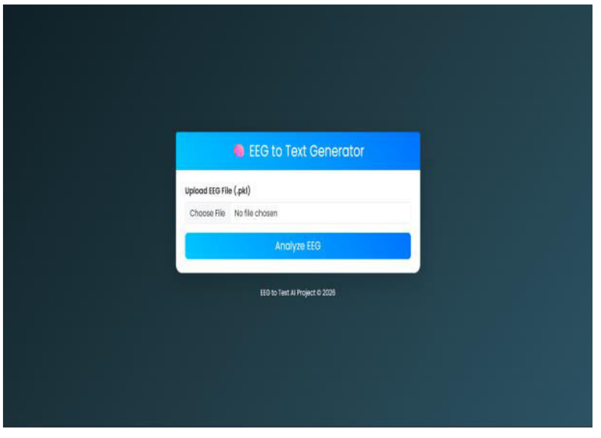

Fig 6.1 Index page

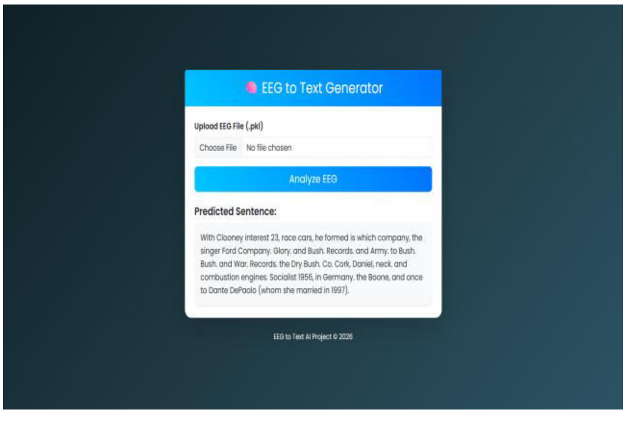

Fig 6.2 Index page with results

VII. CONCLUSION

The successful translation of human thoughts into text via EEG signals represents a critical milestone in the development of accessible, non-invasive Brain-Computer Interfaces. This project has demonstrated that a streamlined, Transformer-based approach can effectively bridge the gap between complex neural activity and linguistic output without the heavy computational requirements of multi-stage multimodal frameworks or multi-billion parameter Large Language Models.

By extending the theoretical foundations of recent research, our proposed system introduced a novel Temporal-Spatial Reshaping technique, treating 630-dimension EEG vectors as a 105-step sequence. This allowed the model's self-attention mechanism to identify critical temporal dependencies within the brain's "inner speech" signals. The implementation of a specialized word encoder with learnable positional embeddings, coupled with a strategically constrained vocabulary of the top 1,000 most frequent tokens, resulted in a robust validation accuracy of ~23.95%.

The results confirm that direct EEG-to-word mapping is not only viable but offers significant advantages in terms of latency, simplicity, and portability. While traditional systems often struggle with error propagation through multiple alignment stages, our single-stage end-to-end architecture provides a cleaner, more efficient pathway for decoding. The sample predictions—such as accurately identifying core entities like "Henry Ford" and "Foundation"—validate that the model is capturing deep semantic structures rather than just surface-level noise.

In conclusion, this work provides a scalable blueprint for future BCI applications. It proves that by focusing on high-quality feature engineering and optimized attention mechanisms, we can move closer to a world where individuals with speech impairments can communicate fluently and independently using only their thoughts. As the field moves toward real-time edge computing, this streamlined Transformer approach serves as a vital foundation for portable, life-changing communication technology.

VIII.REFERENCES

[1] Murad, S. A., & Rahimi, N. (2024). Unveiling Thoughts: A Review of Advancements in EEG Brain Signal Decoding into Text. arXiv preprint arXiv:2405.00726v2 [eess.SP].
[2] Mishra, A., Shukla, S., Torres, J., Gwizdka, J., & Roychowdhury, S. (2025). Thought2Text: Text Generation from EEG Signal using Large Language Models (LLMs). arXiv preprint arXiv:2410.07507v2 [cs.CL].
[3] Spampinato, C., et al. (2017). Deep Learning Human Mind for Context-Aware Visual Classifiers. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[4] Akbar, S. (2026). EEG-to-Text Implementation Pipeline. Jupyter Notebook (.ipynb).
[5] Duan, Y., et al. (2023). DeWave: Discrete EEG Waves Encoding for Brain Dynamics to Text Translation. arXiv preprint arXiv:2309.14030.

[6] Hollenstein, N., et al. (2019). ZuCo 2.0: A Dataset of Physiological Signals During Natural Reading and Annotation. arXiv preprint arXiv:1912.00036.
[7] Vaswani, A., et al. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems (NeurIPS).
[8] Loshchilov, I., & Hutter, F. (2019). Decoupled Weight Decay Regularization. International Conference on Learning Representations (ICLR).
[9] Hendrycks, D., & Gimpel, K. (2016). Gaussian Error Linear Units (GELU). arXiv preprint arXiv:1606.08415.