Language for Thoughts: A NLP approach to EEG-to-MEG

Francesco Capuano [1] **Liam Ludington** [1]

Abstract

Electroencephalography (EEG) offers high temporal resolution, portability, and cost-efficiency but suffers from poor spatial resolution due to the skull's resistive properties. In contrast, magnetoencephalography (MEG) provides improved spatial resolution by detecting magnetic fields, although it is less accessible. Converting EEG data to MEG-like insights could significantly enhance brain activity analysis using already widespread EEG devices.

Motivated by the recent success of encoder-decoder transformer architectures in machine translation, we explore their application in predicting MEG signals from EEG recordings, inspired by the similarities between this task and translating sequences with similar meaning expressed in different languages. In this first report, we detail our methodology, present our experiments training on openly available EEG-MEG paired data, and discuss the key challenges we faced in extracting high spatial fidelity information directly from EEG.

1. Introduction

Electroencephalography is a broadly adopted technique for measuring electrical signals generated by neural activity in the brain. Compared to other common neuro-imaging methods, namely magnetoencephalography and functional magnetic resonance imaging, EEG distinguishes itself by its relatively high time resolution (100-1000 Hz) and cost-efficiency. In comparison, because fMRI relies on blood flow to measure changes in neural firing, it has a temporal resolution on the order of seconds. It is also much less demanding for subjects, who must remain completely still for long periods in MEG and fMRI setups, while EEG allows limited mobility for patients, which can be useful in experimental setups. For these reasons, EEG is the go-to first approach for neuroscientists when testing hypotheses about human brain function and neurologists when measuring patients' brain activity. Unlike MEG, which requires an electromagnetically isolated room and a massive patient-immobilizing detection apparatus to reliably register minuscule magnetic signals, with an upfront cost on the order of millions of dollars, EEG setups usually run on the order of several thousands.

Despite EEG's easy setup and low cost, it has lower spatial resolution than MEG and fMRI. The electrical signals that EEG registers are dispersed by both the cerebro-spinal liquid and the cranium due to their electrically resistive nature. In contrast, the magnetic waves picked up by MEG are less affected by resistors. This gives MEG better spatial resolution, despite its costliness.

EEG and MEG are widely used in the diagnosis and treatment planning of many psychiatric and neurological conditions, such as epilepsy, brain cancer, Alzheimer's, and schizophrenia, to localize pathological brain regions such as tumors or lesions. MEG's higher spatial resolution and higher sensitivity to deep brain regions makes it better suited for pathological region identification than EEG. In particular, inclusion of MEG with EEG data has been shown to improve the diagnosis of epilepsy-causing areas in patients (Aydin et al., 2015). However, the 100 1000x price differential of an MEG system means that access to MEG diagnostics remains relatively rare. The ability to robustly translate EEG signal to MEG signal would help neuroscientists and neurologists locate pathological brain areas with greater precision quicker and cheaper.

In this report, we present our approach to the problem of EEG-MEG translation with an encoder-decoder transformer-based approach. Encoder-decoder architectures emerged in the last 10 years (Sutskever et al., 2014) as a powerful method for translation between two sequential data modalities. The introduction of the Transformer architecture (Vaswani et al., 2017) in 2017, along with the availability of large online data sets in source and target modality, and the algorithmic and computational ability to train deep networks with billions of parameters, allowed the encoder-decoder framework to become much more performant. For

[1] Deparment of Mathematics, École Normale Supérieure Paris-Saclay, Gif-sur-Yvette, France. Correspondence to: Francesco Capuano <francesco.capuano@ens-paris-saclay.fr>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

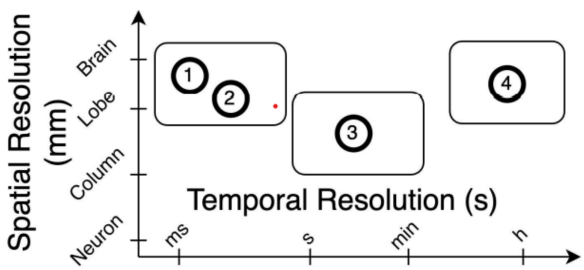

Figure 1. Comparison of the Spatial and Temporal Resolutions of Various Brain Imaging Techniques. **1**: EEG. The highest temporal resolution and the highest spatial resolution **2**: MEG. Higher spatial resolution that EEG. **3**: fMRI. Measures bloodflow with Magnetic Resonance Imagine, with a delay and time resolution on the order of seconds. **4**: PET

instance, the inclusion of pre-trained language transformers in machine translation systems (Zhu et al., 2020) enabled a step change in the performance of automatic translation systems.

Unlike language-to-language translation, EEG-to-MEG does not yet benefit from a wealth of strongly-paired sequences. The different setups required for EEG and MEG recording make simultaneous recording impossible. However, Wakeman & Henson (2015) present a data set of weakly paired EEG and MEG recordings from the same group of subjects under the same experimental conditions recorded first in EEG and subsequently in MEG. Also, in strong contrast to language, EEG and MEG are multi-variate time series (MTS)–they are both composed of various channels (sensor locations) that change independently of each other. Effectively applying transformers to MTS is central to the difficulty of EEG to MEG translation. To the best of our knowledge, a transformer-based encoder-decoder approach has not yet been applied to the EEG to MEG translation problem.

In our method, we surmount the challenge posed by multiple channels by training a Crossformer (Zhang & Yan, 2023) encoder-decoder architecture on the Wakeman data set. We hope to validate our method against real-world MEG recordings, demonstrating our model's promising ability to translate across different sequence types. Our main contribution is the conceptual framing of converting EEG to MEG as a seq2seq translation problem, which allows us to draw from the wealth of effective approaches developed in that literature, including encoder-decoder models and transformers for multi-variate time series.

2. Related Work

Although we are the first to expressly attempt the problem of EEG to MEG translation, there is a growing literature on multi-modal foundation models for neural signals that present interesting methods for solving similar, if not identical, problems. Multimodal foundation models (MMFMs) (**?**) are machine-learning models pre-trained on multiple data modalities (e.g., vision and language) that generate representations useful for a range of downstream tasks. MMFMs leverage paired data to align internal representations across modalities, enabling cross-modal translation tasks such as generating text descriptions from images or vice versa. They often exhibit impressive performance on zero-shot tasks for which have not figured in their pretraining. MMFMs have been applied to medical domains such as pathological slides (**?**) where they show an ability to prediction how perturbations in one modality effect perturbations in another.

Several recent studies attempt to fuse latent representations of EEG and MEG signals, though none focus on direct signal to signal translation. A key challenge in neural signal processing is the inter- and intra-modal variability in channel couunts, which has been addressed in various clever ways by these studies. For example, Défossez et al. (2023) implement a spatial attention map with Fourier parametrization, followed by subject-specific 1×1 channel convolutions and dilated convolutional blocks on the time dimension. They employ a contrastive CLIP loss to align EEG and MEG signals but focus on speech perception decoding rather than direct EEG-to-MEG translation. Although our approach differs in objective, their method provides valuable methods for instantiating inductive biases that create spatially and temporally aware representations.

The task of EEG to MEG translation entails upsampling the spatial resolution of neural measurements, providing a potential augmentation method for multimodal neural foundation models. Notably, Défossez et al. (2023) report that speech decoding from EEG signals is significantly less accurate than from MEG. If our approach proves reliable, it could enhance existing multimodal decoding pipelines by generating high-resolution neural data to improve downstream tasks.

Extending this line of work on decoding speech from neural signals, NeuGPT (Yang et al., 2024) introduces a multimodal neural foundation model aimed at decoding speech from ECoG, SEEG, fMRI, and fNIRS data, as well as M/EEG recordings. Like our work, NeuGPT uses a quantizer to tokenize signals before feeding them into a transformer-based architecture. However, its signal processing pipeline is tailored for decoding speech signals rather than neural signals. Though NeuGPT would in theory enable EEG-to-MEG translation with minor modifications, the authors do not attempt this in their paper. Future versions of this paper would benefit from using NeuGPT modified to translate EEG to MEG as a comparative baseline for experimental validation.

Other works have targeted visual rather than audio perceptual decoding. Ferrante et al. (2024) explore EEG-to-MEG conversion in the context of multimodal neural decoding. Their work aims to identify visual images from the brain imaging signals of subjects presented with certain photos. They encode M/EEG signals into a fused latent space and translate between modalities indirectly by retrieving nearest-neighbor representations rather than performing direct seq2seq translation. Unlike our work, which forecasts MEG time series from raw EEG signals, their approach lacks an explicit decoder for the reconstruction of neural signal.

Despite the fact that our work attacks a problem fundamentally different from the ones targeted by the above-mentioned papers, we draw heavily from similar concepts. The transformers in our architecture will implement a geometrically grounded attention map inspired by Défossez et al. (2023). Although our training will not rely on the contrastive CLIP loss, but rather a next-token prediction loss common in the seq2seq framework, we use an MTS transformer, akin to NeuGPT, although here specifically designed for neural signal translation instead of speech processing.

3. Background

Our work, as is always the case in research, stands on the shoulders of giants. Specifically, the entire endeavor would be meaningless without the vast literature demonstrating the clear clinical benefits of MEG signal. It also relies on the Wakeman data set, without which a seq2seq translation approach to this problem would be fruitless. On the machine learning side, machine translation is greatly aided by previous work on effective preprocessing of M/EEG signal as well as efficient attention mechanisms for spatio-temporal data.

3.1. Neuroscientific Significance of M/EEG

Because of MEG's higher spatial resolution, it is a more potent representation of brain activity than EEG. MEG signal has been widely found to be better than EEG signal for decoding perceptual stimuli (Défossez et al., 2023). Research has also shown that combined MEG/EEG signals outperform pure EEG signals in epileptic source reconstruction for epileptic patients (Aydin et al., 2015). Thus, the ability to generate more expensive and spatially resolved MEG signals from EEG signals has the potential to improve treatment for the many patients worldwide suffering from neurological or psychiatric diseases but without access to costly MEG machines (Næss et al., 2020; da Silva, 2013).

3.2. Paired EEG MEG signals

It is very difficult to obtain EEG and MEG signals from the same subject at the same time. Nevertheless, we have access to weakly paired data thanks to Wakeman & Henson (2015). They present EEG and MEG data acquired from the same nineteen healthy volunteers recorded while the subjects performed multiple runs of a simple perceptual task on images of familiar, unfamiliar and scrambled faces. Though this data comprises only nineteen hours of total recordings, we hope it will suffice for training our encoder-decoder architecture on the EEG to MEG translation task. Despite the fact that for this report we have not , we envision experimenting with pre-trained EEG and MEG modules in future iterations of this work.

3.3. Neural Transformers

The advent of the Transformer architecture has revolutionized machine sequence to sequence translation (**?**). In this report, we apply the sequence to sequence framework to the problem of EEG to MEG time series translation. Although this problem has not been approached before, there do exist successful efforts to apply transformers to neural signal processing. In particular, Zhou & Liu (2024) presents a method of masked auto-encoding (MAE) for effective pre-training of EEG encoder and decoder modules designed specifically for EEG signal (**?**). They show that this pre-training improves precision and learning efficiency on downstream tasks such as gaze estimation. Additionally, Duan et al. (2023) demonstrates the ability of a variational quantized encoder to learn representations alignable with pre-trained language transformers.

In addition to EEG signal processing, there has also been significant work on applying transformer architectures to MEG signals. Csaky et al. (2024) not only shows the powerful inferential ability of foundational models applied to MEG signal but also establishes a principled and effective method of MEG signal pre-processing, which we in large part recreate in our methodology.

3.4. Multi-Channel Approaches

The original Transformer architecture is designed for dealing with a single sequence of tokens. The existence of multiple channels of time-varying tokens, as is the case here, complicates the application of the convention Transformer architecture to this problem. Fortunately, there is already extensive work on the adaptation of Transformers and especially attention maps to multi-channel time series. For instance, the Crossformer architecture factors the attention operation into spatial and temporal attention functions, which allows for improved forecasting without prohibitive computational asymptotics (Zhang & Yan, 2023). We implement our own Crossformer with a two-stage attention mechanism, applied across sensor channels and time steps sequentially. Other approaches to efficient attention calculation in multi-variate time series include low-rank factorization of spatial and temporal dimensions and hybrid multi-dimensional attention with autoencoders (Zhang et al., 2025), (Chang et al., 2021).

4. Method

In this section we outline the pre-processing steps we apply to our M/EEG neural signals as well as the various architectures we use for our MTS encoder-decoder model.

4.1. Preprocessing

The preprocessing is accomplished in two main steps, closely following the methodology of Csaky et al. (2024). The first step uses the MNE library's automated interpolation, filtering, and other methods to establish a consistent signal, while the second step normalizes the signal before quantizing it into tokens.

Signal Cleaning

To ensure the integrity and consistency of the EEG and MEG data, we perform several preprocessing steps before model training.

- Bad Channel Detection and Interpolation. Faulty channels, often caused by sensor malfunctions or artifacts, can degrade data quality and introduce inconsistencies across subjects and recording sessions. We identify bad channels as those with high low-frequency noise using an automated detection algorithm and replace them via spatial interpolation using MNE's `interpolate_bads` function. This procedure ensures that all channels contribute meaningful signals to the analysis.

- Bandpass and Notch Filtering. To retain relevant neural activity while minimizing noise, we apply a bandpass filter (0.1–100 Hz) to eliminate unwanted frequency components outside the typical neural signal range. Additionally, we use a notch filter at 50 Hz and its harmonics (100 Hz, 150 Hz) to remove power line interference. The effectiveness of this filtering step is verified using power spectral density (PSD) plots to confirm that signal power is retained within the desired frequency range.

- Downsampling. Signals are downsampled along the time domain to a desired frequency. This both reduces the data size and allows for different EEG and MEG setups with variable sampling rates to be used together.

- Epoching. The continuous data stream is segmented into event-related epochs to align neural activity with specific experimental conditions. Using event markers, we extract epochs within a predefined time window (e.g., [-0.2, 0.8] s relative to the event onset). This process ensures consistency across trials and allows for direct comparisons across different conditions. The resulting data is structured as a three-dimensional matrix.

- Baseline Correction. To remove low-frequency drifts and emphasize event-related activity, we apply baseline correction using a pre-event time window (e.g., -0.2 to 0 s). For each channel, the mean signal value within the baseline period is subtracted from the entire epoch, ensuring that signals are centered around zero prior to event onset.

Tokenization

To prepare the neural signals for processing within our transformer-based model, we employ a multi-step tokenization procedure inspired by speech processing pipelines.

- Signal Normalization. Neural signals can exhibit large amplitude variations across channels, recording sessions, and subjects, which can hinder direct comparison and model convergence. Data is normalized with percentile clamp and scale either per channel or across channels. Though both practices are common in the literature, they both have their advantages and disadvantages. Here, we favor per channel normalization, in order not to scale some channels to much smaller size than other channels.

- μ-Law Compression. We employ μ-law compression, a nonlinear transformation commonly used in audio signal processing, to enhance resolution in low-amplitude regions while preserving high-amplitude components. This step improves the representation of fine-grained neural dynamics, which are crucial for effective EEG-to-MEG translation.

- Quantization. Finally, we discretize the continuous neural signals into a finite set of 255 quantization levels, mapping real-valued signals to integer representations. This transformation is essential for models that operate on discrete tokens and allows for efficient representation learning while preserving signal fidelity.

4.2. Model

We implement four different architectures for encoding and decoding of the M/EEG signals. These include more traditional approaches like convolutional nets, more modern architectures like the Crossformer (Zhang & Yan, 2023), and hybrid approaches like a convolutional attention network. We also implement a continuous Crossformer so that we can avoid quantization of our data.

4.2.1. Convolutional Network

The convolutional network consists of two initial convolutions, one for frequency and another for temporal features. These are followed by multi-head attention blocks to generate the three components of MEG signal, which are then concatenated together. The convolutional architecture comprises the following key components:

Frequency and Temporal Convolutional Layers: The model has two initial convolutional layers for frequency feature extraction and temporal feature extraction. Each layer contains sequences of 1D convolutional layers, ReLU activations, and batch normalization layers. The frequency convolution includes a 256-channel 1D convolution followed by a 128-channel 1D convolution, while the temporal convolution follows a similar structure but with different kernel sizes.

Attention Mechanisms: The model incorporates multi-head attention layers to process combined feature representations from the convolutional layers. These attention layers help the model focus on different parts of the input sequence and enhance the feature extraction process.

Residual Blocks: The architecture includes several residual blocks integrated with the attention layers and output layers. Each residual block consists of two convolutional layers with batch normalization and ReLU activations, facilitating skip connections to improve gradient flow and feature learning.

Output Layers: The final output layers for the magnetic, gradient1, and gradient2 predictions also include residual blocks followed by 1D convolutional layers. These layers generate the final predictions by combining the features extracted through the previous layers.

4.2.2. Crossformer

The Crossformer is a speial case of an encoder-decoder Transformer model, incorporating a two-stage attention mechanism on time-series data. The model processes input data in the shape (B, T, D), where B is the batch size, T is the sequence length, and D is the dimension of the time series. The model architecture is designed to handle masking exclusively in the time dimension and considers the dimensions of the time series as permutation-invariant, thus excluding them from the positional encoding. The class initializes with parameters such as the source and target sequence dimensions, embedding dimensions, number of attention heads, and feedforward dimensions. It creates positional encodings for the source and target channels and uses these to embed the input sequences before passing them through the encoder and decoder layers.

The encoder and decoder are instantiated as stacks of two stage attention layer modules, which perform the two-stage attention mechanism. During the forward pass, the source and target sequences are embedded with positional and channel embeddings and scaled. The source sequence is then passed through the encoder layers, which apply attention mechanisms to capture dependencies in the data. The decoder layers process the target sequence, utilizing self-attention and cross-attention mechanisms to incorporate information from the encoder outputs. The model reshapes the encoder output for dimension-wise cross-attention and finally returns the output after passing through the decoder layers.

Two-Stage Attention: The two stage attention layer is designed to handle input data in the shape of `[batch_size, D, L, d_model]`, where D is the number of channels, L is the number of time segments, and d_{model} is the embedding dimension. The first stage, called Cross Time Stage, applies Multi-Head Self-Attention (MSA) across the time dimension, followed by feed-forward layers and normalization. The output is reshaped and passed to the second stage, Cross Dimension Stage, which uses a smaller router matrix to distribute messages and build channel-to-channel connections. This stage involves two additional Multi-Head Attention layers: one for attention from time to the router and another for attention from the router back to the channels. The result is a refined output that captures dependencies across both time and channel dimensions.

Encoder and Decoder: The transformer's encoder and decoder modules are implemented by stacks of transformer blocks with pre-normalization and dropout for stable training. Each block consists of a multi-head self-attention mechanism followed by a feed-forward neural network. The input tensor is first normalized and passed through the self-attention layer, which can handle optional attention and key padding masks. The output is then added back to the input tensor (residual connection) after applying dropout. This process is repeated for the feed-forward network, ensuring robust and effective encoding of the input data.

4.2.3. Continuous Crossformer

The continuous Crossformer, while rumored to exist in top research labs, remains too be implemented in our work.

5. Future Work

One of the challenges of this sequence-to-sequence translation task is the variable number of recording channels in EEG and MEG set-ups. Beyond that, individual EEG and

MEG channel sensors are placed at different points on subjects skull in different experiments. Going forward, to make signals across subjects comparable, we propose learning embeddings from graph neural networks to take into account the strong geometric information present in these signals.

6. Code

All of our code will be made publicly available at https://github.com/fracapuano/brainformer. The paired EEG-MEG dataset is available at https://openneuro.org/datasets/ds000117

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Aydin, Ü., Vorwerk, J., Dümpelmann, M., Küpper, P., Kugel, H., Heers, M., Wellmer, J., Kellinghaus, C., Haueisen, J., Rampp, S., et al. Combined eeg/meg can outperform single modality eeg or meg source reconstruction in presurgical epilepsy diagnosis. *PloS one*, 10(3):e0118753, 2015.

Chang, F.-J., Radfar, M., Mouchtaris, A., King, B., and Kunzmann, S. End-to-end multi-channel transformer for speech recognition. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5884–5888. IEEE, 2021.

Csaky, R., van Es, M. W., Jones, O. P., and Woolrich, M. Foundational gpt model for meg. *arXiv preprint arXiv:2404.09256*, 2024.

da Silva, F. L. Eeg and meg: relevance to neuroscience. *Neuron*, 80(5):1112–1128, 2013.

Défossez, A., Caucheteux, C., Rapin, J., Kabeli, O., and King, J.-R. Decoding speech perception from non-invasive brain recordings. *Nature Machine Intelligence*, 5(10):1097–1107, 2023.

Duan, Y., Zhou, J., Wang, Z., Wang, Y.-K., and Lin, C.-T. Dewave: Discrete eeg waves encoding for brain dynamics to text translation. *arXiv preprint arXiv:2309.14030*, 2023.

Ferrante, M., Boccato, T., Rashkov, G., and Toschi, N. Towards neural foundation models for vision: Aligning eeg, meg, and fmri representations for decoding, encoding, and modality conversion. *arXiv preprint arXiv:2411.09723*, 2024.

Næss, S., Halnes, G., Hagen, E., Hagler Jr, D. J., Dale, A. M., Einevoll, G. T., and Ness, T. V. Biophysical modeling of the neural origin of eeg and meg signals. *bioRxiv*, pp. 2020–07, 2020.

Sutskever, I., Vinyals, O., and Le, Q. V. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27, 2014.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Wakeman, D. G. and Henson, R. N. A multi-subject, multi-modal human neuroimaging dataset. *Scientific data*, 2(1): 1–10, 2015.

Yang, Y., Duan, Y., Jo, H., Zhang, Q., Xu, R., Jones, O. P., Hu, X., Lin, C.-t., and Xiong, H. Neugpt: Unified multi-modal neural gpt. *arXiv preprint arXiv:2410.20916*, 2024.

Zhang, Y. and Yan, J. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The eleventh international conference on learning representations*, 2023.

Zhang, Y., Liu, Y., Yuan, H., Qin, Z., Yuan, Y., Gu, Q., and Yao, A. C.-C. Tensor product attention is all you need. *arXiv preprint arXiv:2501.06425*, 2025.

Zhou, Y. and Liu, S. Enhancing representation learning of eeg data with masked autoencoders. In *International Conference on Human-Computer Interaction*, pp. 88–100. Springer, 2024.

Zhu, J., Xia, Y., Wu, L., He, D., Qin, T., Zhou, W., Li, H., and Liu, T.-Y. Incorporating bert into neural machine translation. *arXiv preprint arXiv:2002.06823*, 2020.