# A Causal Perspective in Brainwave Foundation Models

**Konstantinos Barmpas**
Imperial College London & Cogitat
konstantinos.barmpas16@imperial.ac.uk

**Yannis Panagakis**
National and Kapodistrian University of Athens &
Archimedes Research Unit & Cogitat

**Dimitrios A. Adamos**
Imperial College London & Cogitat

**Nikolaos Laskaris**
Aristotle University of Thessaloniki & Cogitat

**Stefanos Zafeiriou**
Imperial College London & Cogitat

## Abstract

Foundation Models have recently emerged as powerful tools in various domains of AI, showing potential for significant advancements in Brain-Computer Interfaces (BCIs). However, the initial implementations of Large Brainwave Models (LBMs) face significant challenges when applied to real-world scenarios, primarily due to various distribution shifts. This work examines the training process of these LBMs through a causal reasoning perspective, identifying key challenges that impact their performance. By doing so, we aim to provide insights that can guide the development of more robust and effective LBMs for BCI applications.

## 1 Introduction

Brain-Computer Interface (BCI) technology is revolutionizing human-computer interaction by creating direct connections between the human brain and computers. This advancement relies on analyzing brainwaves from electroencephalogram (EEG) recordings using sophisticated signal processing and machine learning techniques. BCIs have the potential to transform how we interact with our environment and each other, offering hope for people with disabilities to regain lost functions, augmenting human abilities, and opening new avenues for entertainment and communication [Irimia et al. (2012), Torres et al. (2020), Chaudhary et al. (2016), Luu et al. (2017), Biasiucci et al. (2018)].

Traditionally, analyzing brainwaves captured by EEG involved manual feature extraction by neuroengineers with extensive expertise in neuroscience and EEG analysis [Bashashati et al. (2007), Handy (2009), Rao (2013), Nam et al. (2018), McFarland et al. (2006)]. However, hand-crafted features often struggle to generalize well to data captured in real-world settings, limiting their effectiveness in practical BCI applications.

The rise of deep learning has reduced the need for manual feature extraction in BCIs, enabling data-driven approaches and leading to state-of-the-art performance in various BCI paradigms [Lawhern et al. (2018), Song et al. (2023), Barmpas et al. (2023a), Bakas et al. (2022)]. While deep learning models have shown promising results, they typically require substantial supervision and task-specific data collection, which can be time-consuming and resource-intensive.

Foundation Models have recently emerged as a promising approach to address the above mentioned limitations, showing remarkable results across various domains, particularly in Natural Language Processing and Computer Vision [Brown et al. (2020), Touvron et al. (2023), Mizrahi et al. (2023)]. Recognizing this potential, researchers have begun to apply similar large-scale models to BCIs,

NeurIPS 2024 Workshop on Causality and Large Models (CaLM).

which we refer to as Large Brainwave Models (LBMs) [Jiang et al. (2024), Cui et al. (2024), Yuan et al. (2024), Zhang et al. (2024)]. In the context of BCIs, LBMs offer several potential advantages: they can leverage knowledge across different brain-related tasks, potentially improve generalization, reduce the need for task-specific data collection, and create more robust and versatile BCI systems capable of adapting to various users and environments.

However, applying LBMs to BCIs presents unique challenges due to the nature of brain signals. These challenges include high inter-subject variability, where brain signals can vary significantly between individuals; non-stationarity of EEG signals, which can change over time even within the same individual; and susceptibility to various artifacts and noise sources. Furthermore, the high-dimensional and temporally dynamic nature of EEG data, with both spatial (multiple electrodes) and temporal aspects, poses significant challenges for efficient data representation and processing within these large-scale models. These factors collectively make it particularly challenging for LBMs to generalize effectively across different subjects, tasks, and recording conditions in BCI applications.

To address these challenges, we propose leveraging causal reasoning, which offers a promising solution to the limitations encountered by LBMs in BCI applications. Causal reasoning focuses on understanding cause-effect relationships, potentially allowing us to better model EEG data's complex and variable nature. Inspired by work in Representation Learning [Schölkopf et al. (2021)] and Medical Imaging [Castro et al. (2020)], [Barmpas et al. (2024)] presents a framework to analyze key challenges in brainwave modeling for BCIs. Our work here expands this framework to encompass LBMs, aiming to drive significant advancements in BCIs, paving the way for more robust, generalizable, and interpretable brainwave modeling techniques to better account for confounding influences and reduce the need for task-specific data.

## 2 Background on Causal Reasoning

[Barmpas et al. (2024)] introduces a causal framework that can break down every BCI paradigm to its independent components, according to the Principle of Independent causal mechanism [Reichenbach (1956)], and relies on two core properties:

- **The presence of task stimulus**: if the brain activity is modulated by the presence of external stimuli (Exogenous) or without any external stimulus (Endogenous)
- **The voluntary engagement of the subject**: when the subject willingly generates a particular brain activation or coactivation pattern (Voluntarily Engaged) or when the subject has no control of the generated brain activation pattern (Involuntarily Engaged)

Therefore, all BCI paradigms can be categorized using a combination of the above mentioned categories. For example, P300 speller [Won et al. (2019)] is exogeneous and voluntary engaged paradigm while seizure detection [Antoniades et al. (2016)] and workload estimation [Kosti et al. (2018)] are endogeneous and involuntary engaged paradigms.

According to [Barmpas et al. (2024)], the identified variables of interests in a BCI task are: the observed EEG brainwave signal $\mathsf{X}$, the labelled task $\mathsf{Y}$, the true underlying unobserved BCI-relevant brain activity $\mathsf{Z}$, the stimulus $\mathsf{S}$, the intention $\mathsf{I}$ of the subject and the environment $\mathsf{E}$. For each category, there is a specific directed acyclic graph (DAG) that represents the causal relations between the identified variables of interests:

- Voluntarily Engaged - Exogenous: $\mathsf{Y} \rightarrow \mathsf{S} \rightarrow \mathsf{Z} \rightarrow \mathsf{X}$
- Involuntarily Engaged - Exogenous: $(\mathsf{Y}, \mathsf{S}) \rightarrow \mathsf{Z} \rightarrow \mathsf{X}$
- Involuntarily Engaged - Endogenous: $\mathsf{E} \rightarrow \mathsf{Y} \rightarrow \mathsf{Z} \rightarrow \mathsf{X}$
- Voluntarily Engaged - Endogenous: $\mathsf{I} \rightarrow \mathsf{Y} \rightarrow \mathsf{Z} \rightarrow \mathsf{X}$

Different distribution shifts can influence different causal sub-modules, leading to sub-optimal performance of brainwave decoding. These shifts can originate from various sources, including the experimental setup, the quality and quantity of available EEG data, and the inherent complexities of the brain itself. For researchers in BCI systems, it is crucial to thoroughly understand these potential sources of distribution shifts that are linked to the core causal variables. This is essential when designing deep learning models for brainwave analysis, as it enables the development of more robust and accurate decoding systems [e.g. Barmpas et al. (2023b), Barmpas et al. (2022)]. Table 1 summarizes some of the various shifts in brainwave decoding as well as the causal sub-module they usually affect.

| **Type** | Related with | Causal Shift to | Affects |
|---|---|---|---|
| Stimulus Shift | Experimental Settings | $P(\mathsf{S})$ | Involuntarily Engaged and Exogenous |
| Shifts on Subject's Intention | Experimental Settings | $P(\mathsf{I})$ | Voluntarily Engaged and Endogenous |
| Shifts on Environmental Conditions | Experimental Settings | $P(\mathsf{E})$ | Involuntarily Engaged and Endogenous |
| Acquisition Shift | EEG Data | $P(\mathsf{X}\|\mathsf{Z})$ | All Cases |
| Subject Shift | Subjects | $P(\mathsf{Z}\|\cdot)$ | All Cases |
| Class Imbalance | Associated Labels | $P(\mathsf{Y})$ | Exogenous |

Table 1: All possible shifts in brainwave decoding for all causal BCI factorization scenarios. Table adapted from [Barmpas et al. (2024)].

## 3 Extended Causal Framework for Large Brainwave Model

Over the past decade, advancements in Deep Learning have led to the development of powerful Generative Large Foundation Model architectures that have achieved remarkable success. These foundation models have demonstrated great results in various downstream tasks while their capability to generate highly realistic data further enhances their efficacy and broadens their applicability.

In the field of BCIs, LBM architectures hold tremendous potential for EEG decoding. Early efforts within the BCI research community have already begun to explore this potential [Jiang et al. (2024), Cui et al. (2024), Yuan et al. (2024), Zhang et al. (2024)]. These models, with their extensive self-supervised pre-training on diverse, unlabeled datasets, are capable of capturing complex patterns and relationships within EEG data, enhancing their generalizability. As BCIs evolve from controlled laboratory environments to everyday user settings, there is a growing need for deep learning models that can adapt to a wide variety of consumer-grade devices and remain resilient to signal irregularities. LBMs offer promising solutions for predicting brain activities, reconstructing missing or corrupted segments of brain signals, and improving the removal of artifacts or noise from raw EEG data. These advancements not only enhance the robustness of brainwave decoding models but also pave the way for the development of fully device-agnostic models—capable of being deployed across a range of headsets with varying sensor configurations, straight out of the box.

Here, we will investigate these LBMs through the lens of causal reasoning. More specifically, using the framework described in Section 2, our goal is to identify challenges and possible distribution shifts that can hurt the generalization and performance of these large models. BCI researchers need to take into account these challenges when designing large brainwave decoding models. Our analysis is focused on two key aspects that characterize these large models:

1. Self-supervised pre-training of LBMs
2. Adaptation to downstream tasks

### 3.1 Self-Supervised Pre-training

LBMs are distinguished by their pre-training on vast and diverse datasets in a task-agnostic manner, often leveraging self-supervised learning techniques. This approach equips them with the flexibility and readiness for fine-tuning across a wide range of downstream tasks, eliminating the need for large, annotated datasets typically required for supervised learning.

As discussed in [Schölkopf et al. (2021)] and [Barmpas et al. (2024)], pre-training aims to broaden the diversity of the training set distribution without requiring additional examples from the joint distribution P(X,Y). The assumption is that pre-training on a large and diverse brainwave dataset, enables the model to capture valuable information from these varied distributions. This can enhance the model's ability to generalize in BCI systems. In contrast, self-supervised learning involves two key phases: first, training on a large, unlabeled dataset, followed by fine-tuning on a smaller, labeled set of brainwave data. The primary goal of this approach is to enrich P(X), which is especially advantageous for anti-causal models, predominantly used in brainwave decoding.

However, training of these LBMs relies on various masking choices and reconstruction capabilities. In other words, the knowledge captured by these large brainwave models is predominantly based on their ability to effectively predict masked parts of the input data from other parts. Two main dimensions can be masked:

1. Temporal dimension: Specific time segments are masked in the EEG time-series
2. Spatial dimension: Specific EEG sensors are masked

Therefore, bad masking choices can heavily affect the generalization properties and performance of these large models. Employing causal reasoning within the EEG data guided by the causal framework of [Barmpas et al. (2024)] can facilitate the design of efficient LBMs.

**Exogenous.** When a BCI paradigm is categorized as exogenous, it means that the brain's activity is influenced by external stimuli ($\mathsf{S} \rightarrow \mathsf{Z} \rightarrow \mathsf{X}$). The brain's response to these stimuli is what enables the decoding of brainwaves. For example, the P300-speller [Won et al. (2019)] is a type of BCI that allows users to spell words and sentences by focusing on specific flashing characters. This system relies on the P300 Event-Related Potential (ERP), a brain response that typically emerges around 300 milliseconds after the presentation of the stimulus mostly in the parietal electrodes.

If, during the training of a LBM, all electrodes — or even just the electrodes relevant to the stimulus - are masked during the stimulus time segment, the model will be unable to accurately reconstruct the brainwave patterns associated with the stimulus. This masking deprives the model of essential information, preventing it from learning the critical features and dynamics necessary for effective brainwave generation of the stimulus response.

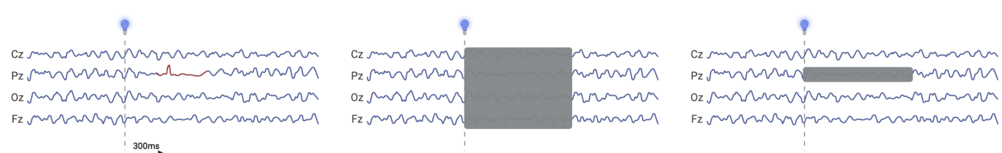


Figure 1: (Left) EEG signals before and after an external stimulus is applied in a P300-speller trial. Blue denotes the baseline neural activity and red denotes the event-related potential (ERP) response occurring 300 milliseconds after the stimulus. (Middle) Temporal masking in all electrodes during the stimulus time segment. (Right) Masking the electrode relevant to the stimulus during the stimulus time segment.

**Endogenous - Voluntarily Engaged.** When a BCI paradigm is categorized as endogenous - voluntarily engaged, it means that the intention $\mathsf{I}$ of the subject can be considered the source of the BCI-relevant neural activation pattern ($\mathsf{I} \rightarrow \mathsf{Y} \rightarrow \mathsf{Z} \rightarrow \mathsf{X}$). The neural patterns during the task is what enables the decoding of brainwaves. For example, a Motor-Imagery BCI [Rezaeitabar & Halici (2017)] is a system that decodes when a user imagines of moving their hands and / or feet with the main neural activity being concentrated in the motor cortex electrodes. Since these paradigms are user-driven, their duration can vary significantly, making the application of temporal masking particularly challenging.

If, during the training of a LBM, all electrodes — or even just the electrodes relevant to the task - are masked during the task's duration, the model will be unable to accurately reconstruct the brainwave patterns associated with the task.

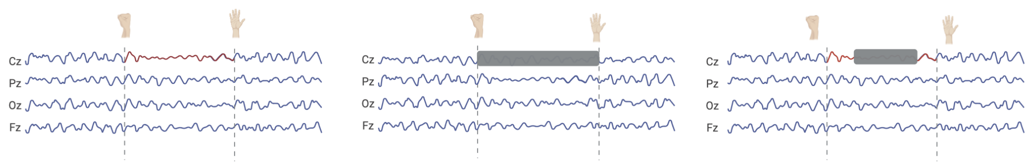


Figure 2: (Left) EEG signals during a Motor-Imagery (MI) trial. Blue denotes the baseline neural activity and red denotes the neural pattern related to the MI task. (Middle) Masking the electrode relevant to the MI task during the whole task duration.(Right) Partially masking the electrode relevant to the MI task during the task duration.

**Endogenous - Involuntarily Engaged.** When a BCI paradigm is categorized as endogenous - involuntarily engaged, it means that the brain's activity is influenced by the environment $\mathsf{E}$ ($\mathsf{E} \rightarrow \mathsf{Y} \rightarrow \mathsf{Z} \rightarrow \mathsf{X}$). These BCI paradigms often involve transitions between different brain states (such as baseline versus seizure [Antoniades et al. (2016)] or baseline versus drowsiness [Lin et al. (2010)]) that can persist over extended periods and are detectable across all or only specific electrode sites.

If, during the training of a LBM, a large masking time window is used that includes all electrodes — or even just the electrodes relevant to a state, the model will be unable to accurately reconstruct the brainwave patterns associated with that state.

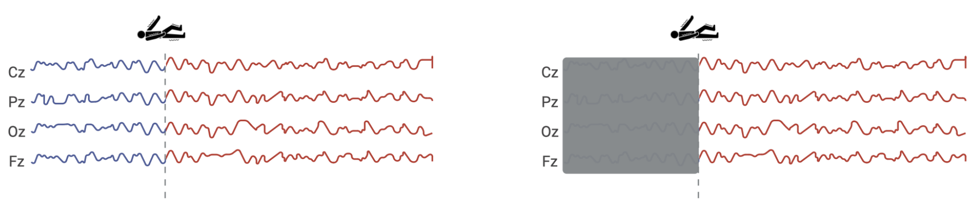


Figure 3: (Left) EEG signals before and during a seizure episode. Blue denotes the baseline neural activity and red denotes the neural pattern related to seizure. (Right) Temporal masking of all electrodes during the baseline state.

Consequently, choosing an ineffective spatial masking method for each BCI paradigm will prevent the model from accurately capturing the essential neural activity associated with each causal category. The variability in timing introduces additional complexity. This variability means that improperly applied temporal masking can obscure critical neural activity, potentially leading to inaccurate or incomplete data information.

Therefore, the BCI researcher developing LBMs should:

1. Ensure careful selection and retention of the relevant electrodes during the training phase
2. Carefully consider the timing and duration of masking to ensure that essential temporal information is preserved

### 3.2 Downstream Tasks

The vast majority of BCI models are used for decoding tasks. In other words, given an observed brain activity X (brainwave), the goal of these models is to predict the performed task Y corresponding to that particular brainwave pattern, i.e. P(Y|X). Therefore, when used for downstream tasks, these LBMs are no different (from a causality perspective) than any other deep learning decoding models (e.g. CNN-based brainwave decoders). Hence, they face the same challenges as described in Table 1 (e.g. inter-subject variability, device acquisition differences).

In downstream tasks, LBMs can adapt seamlessly to new tasks with minimal re-training, often achieved through the addition of simple, adaptable layers trained on a small amount of target data. As a result, the causal brainwave framework of [Barmpas et al. (2024)] can be successfully extended to these large brainwave models and BCI researchers need to take into account the guidelines and challenges described in [Barmpas et al. (2024)] during the adaptation training phase.

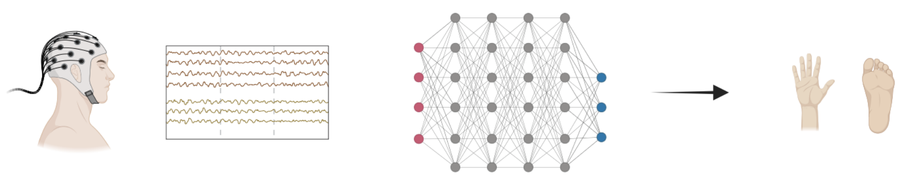

Figure 4: A Motor-Imagery (MI) BCI workflow: Classification of EEG signals into hand and foot movements. The decoding model can be either a deep MI decoding model or a large brainwave model in the MI downstream task.

## 4 Conclusion

In this work, we show that the integration of causal reasoning in the training as well as the adaptation and inference phase of Large Brainwave Models (LBMs) suggests a promising pathway towards more robust and generalizable BCI decoding systems. By extending the existing causal brainwave framework [Barmpas et al. (2024)] to these advanced models, BCI researchers can better navigate the challenges posed by distribution shifts and masking choices that severely impact these large models' performance. We showcase that during adaptation and inference these large models suffer from the distribution shifts as other deep models, while the careful masking selection, both temporal and spatial, during the training phase are essential for capturing critical neural activity accurately during training. As BCIs continue to evolve, the application of this causal framework has the potential to improve BCIs and will be crucial in driving the development of efficient brainwave deep decoding models.

## References


Andreas Antoniades, Loukianos Spyrou, Clive Cheong Took, and Saeid Sanei. Deep learning for epileptic intracranial eeg data. In *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, 2016. doi: 10.1109/MLSP.2016.7738824.

Stylianos Bakas, Siegfried Ludwig, Konstantinos Barmpas, Mehdi Bahri, Yannis Panagakis, Nikolaos Laskaris, Dimitrios A. Adamos, and Stefanos Zafeiriou. Team cogitat at neurips 2021: Benchmarks for eeg transfer learning competition, 2022.

Konstantinos Barmpas, Yannis Panagakis, Dimitrios A. Adamos, Nikolaos Laskaris, and Stefanos Zafeiriou. A CAUSAL VIEWPOINT ON MOTOR-IMAGERY BRAINWAVE DECODING. *ICLR2022 Workshop on the Elements of Reasoning: Objects, Structure and Causality*, 2022.

Konstantinos Barmpas, Yannis Panagakis, Dimitrios A Adamos, Nikolaos Laskaris, and Stefanos Zafeiriou. Brainwave-scattering net: a lightweight network for eeg-based motor imagery recognition. *Journal of Neural Engineering*, 20(5):056014, September 2023a. ISSN 1741-2552.

Konstantinos Barmpas, Yannis Panagakis, Stylianos Bakas, Dimitrios A. Adamos, Nikolaos Laskaris, and Stefanos Zafeiriou. Improving generalization of cnn-based motor-imagery eeg decoders via dynamic convolutions. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31: 1997–2005, 2023b. doi: 10.1109/TNSRE.2023.3265304.

Konstantinos Barmpas, Yannis Panagakis, Georgios Zoumpourlis, Dimitrios A Adamos, Nikolaos Laskaris, and Stefanos Zafeiriou. A causal perspective on brainwave modeling for brain–computer interfaces. *Journal of Neural Engineering*, 21(3):036001, may 2024.

Ali Bashashati, Mehrdad Fatourechi, Rabab K Ward, and Gary E Birch. A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals. *Journal of Neural Engineering*, 4(2):R32–R57, mar 2007. doi: 10.1088/1741-2560/4/2/r03.

A. Biasiucci, R. Leeb, I. Iturrate, S. Perdikis, A. Al-Khodairy, T. Corbet, A. Schnider, T. Schmidlin, H. Zhang, M. Bassolino, D. Viceic, P. Vuadens, A.G. Guggisberg, and J. d. R. Millán. Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke. *Nat Commun*, 9, 2018. doi: 10.1038/s41467-018-04673-z.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020.

D.C. Castro, I. Walker, and B. Glocker. Causality matters in medical imaging. Commun 11, 3673. 2020. doi: https://doi.org/10.1038/s41467-020-17478-w.

U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday. Brain–computer interfaces for communication and rehabilitation. *Nat Rev Neurol*, 12, 2016. doi: 10.1038/nrneurol.2016.113.

Wenhui Cui, Woojae Jeong, Philipp Thölke, Takfarinas Medani, Karim Jerbi, Anand A Joshi, and Richard M Leahy. Neuro-gpt: Towards a foundation model for eeg. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5. IEEE, 2024.

Todd C. Handy. *Brain signal analysis advances in neuroelectric and neuromagnetic methods. Cambridge, Mass, MIT Press.* 2009.

D.C. Irimia, R. Ortner, G. Krausz, C. Guger, and M.S. Poboroniuc. Bci application in robotics control. *IFAC Proceedings Volumes*, 45(6):1869–1874, 2012. ISSN 1474-6670. doi: 10.3182/20120523-3-RO-2023.00432. 14th IFAC Symposium on Information Control Problems in Manufacturing.

Weibang Jiang, Liming Zhao, and Bao liang Lu. Large brain model for learning generic representations with tremendous EEG data in BCI. In *The Twelfth International Conference on Learning Representations*, 2024.

Makrina Viola Kosti, Kostas Georgiadis, Dimitrios A. Adamos, Nikos Laskaris, Diomidis Spinellis, and Lefteris Angelis. Towards an affordable brain computer interface for the assessment of programmers' mental workload. *International Journal of Human-Computer Studies*, 115:52–66, 2018. ISSN 1071-5819. doi: https://doi.org/10.1016/j.ijhcs.2018.03.002.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5):056013, jul 2018.

Chin-Teng Lin, Che-Jui Chang, Bor-Shyh Lin, Shao-Hang Hung, Chih-Feng Chao, and I-Jan Wang. A real-time wireless brain–computer interface system for drowsiness detection. *IEEE Transactions on Biomedical Circuits and Systems*, 4(4):214–222, 2010. doi: 10.1109/TBCAS.2010.2046415.

T.P. Luu, S. Nakagome, Y. He, and J.L. Contreras-Vidal. Real-time eeg-based brain-computer interface to a virtual avatar enhances cortical involvement in human. *Sci Rep*, 7, 2017. doi: 10.1038/s41598-017-09187-0.

D.J. McFarland, C.W. Anderson, K.-R. Muller, A. Schlogl, and D.J. Krusienski. Bci meeting 2005-workshop on bci signal processing: feature extraction and translation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):135–138, 2006. doi: 10.1109/TNSRE.2006.875637.

David Mizrahi, Roman Bachmann, Oguzhan Fatih Kar, Teresa Yeo, Mingfei Gao, Afshin Dehghan, and Amir Zamir. 4m: Massively multimodal masked modeling. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL `https://openreview.net/forum?id=TegmlsD8oQ`.

Chang S. Nam, Anton Nijholt, and Fabien Lotte. *Brain–Computer Interfaces Handbook: Technological and Theoretical Advances, CRC Press*. 2018.

Rajesh P N Rao. *Brain-computer interfacing: an introduction*. 2013.

Hans Reichenbach. *The Direction of Time*. Dover Publications, 1956.

Yousef Rezaeitabar and Ugur Halici. A novel deep learning approach for classification of eeg motor imagery signals. *Journal of Neural Engineering*, 14:016003, 02 2017. doi: 10.1088/1741-2560/14/1/016003.

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021. doi: 10.1109/JPROC.2021.3058954.

Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. Eeg conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2023. doi: 10.1109/TNSRE.2022.3230250.

Edgar P. Torres, Edgar A. Torres, Myriam Hernández-Álvarez, and Sang Guun Yoo. Eeg-based bci emotion recognition: A survey. *Sensors*, 20(18), 2020. ISSN 1424-8220. doi: 10.3390/s20185083.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. Llama: Open and efficient foundation language models, 2023.

Kyungho Won, Moonyoung Kwon, Sehyeon Jang, Minkyu Ahn, and Sung Chan Jun. P300 speller performance predictor based on rsvp multi-feature. *Frontiers in Human Neuroscience*, 13, 2019. ISSN 1662-5161. doi: 10.3389/fnhum.2019.00261.

Zhizhang Yuan, Daoze Zhang, Junru Chen, Geifei Gu, and Yang Yang. Brant-2: Foundation model for brain signals. *arXiv preprint arXiv:2402.10251*, 2024.

Daoze Zhang, Zhizhang Yuan, Yang Yang, Junru Chen, Jingjing Wang, and Yafeng Li. Brant: Foundation model for intracranial neural signal. *Advances in Neural Information Processing Systems*, 36, 2024.