Position: Addressing Ethical Challenges and Safety Risks in GenAI-Powered Brain-Computer Interfaces

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

Georgios Zoumpourlis*
Cogitat

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

The use of Generative AI (GenAI) in developing large brainwave foundation models for Brain-Computer Interfaces (BCIs) offers enormous potential but also comes with several key safety and ethical concerns. This work identifies these challenges and highlights cases of potential misuse of GenAI in BCIs, including synthetic neural activity, behaviour profiling, privacy and equality risks. Finally, it emphasizes the importance of essential safeguarding techniques to mitigate these risks, such as innovative technological solutions and proper regulatory and ethical frameworks.

1 Introduction

Brain-Computer Interface (BCI) technology (Vidal, 1973; Wolpaw et al., 2002) represents a groundbreaking advancement in human-computer interaction, enabling direct communication and control between the human brain and computers. BCI systems typically leverage neuroimaging techniques, with electroencephalography (EEG) being the most common, to monitor brain activity and employ advanced signal processing and learning methods for analysing the collected EEG signals (Yadav et al., 2020).

The breadth of BCI applications spans several domains, including: neurorehabilitation (Sitaram et al., 2017); sleep staging (Phan et al., 2021); mental workload monitoring (Mühl et al., 2014); control of effectors, such as computer cursors or prosthetic limbs (Vilela & Hochberg, 2020).This wide range of applications indicates the capacity of BCIs to revolutionize how we interact with our surroundings and each other.

Historically, the analysis of EEG signals relied on the extraction of manually designed handcrafted features (Pfurtscheller & Da Silva, 1999). This detailed process, executed by neuroengineers with specialized knowledge in neuroscience and EEG signal processing, was oriented around the "expert scientist" approach. The advent of deep learning has transformed this landscape (Roy et al., 2019), enabling data-driven analysis of raw EEG signals. Deep learning techniques have replaced handcrafted features with automatically learned features (Lawhern et al., 2018; Song et al., 2023; Barmpas et al., 2023a), leading to the development of deep neural networks with significantly improved performance

*These authors contributed equally to this work.

GenAI for Health Workshop @ NeurIPS 2024, Vancouver.

in several benchmarks, especially in terms of cross-subject generalization (Pérez-Velasco et al., 2022; Barmpas et al., 2023b; Wei et al., 2021).

The emergence of Generative Artificial Intelligence (GenAI) and particularly Large Foundation Models has brought significant advancements to various fields, including Natural Language Processing (Devlin et al., 2019) and Computer Vision (Radford et al., 2021). Key features of these models include: massive scale (hardware requirements, volume of training data, and number of trainable model parameters) and strong transfer learning capabilities.

Recently, a series of studies combining generative transformers and masked pre-training (Radford et al., 2018; He et al., 2022) have proposed training foundation models on EEG data (Cui et al., 2024; Zhang et al., 2024; Yuan et al., 2024; Jiang et al., 2024). These Large Brainwave Models (LBMs) serve as powerful backbones for several downstream tasks, having a profound impact on the field of BCIs that was previously dominated by uni-task deep learning models for EEG decoding. Apart from their impressive performance on classification tasks, these models also excel in generative tasks, possessing the ability to synthesize highly realistic data, which further enhances their efficacy and broadens their applicability.

However, research on LBMs for BCIs, as well as their deployment, pose numerous challenges. The safety and ethical concerns that arise include the potential for misuse and the need for robust ethical guidelines.

In this work, we aim to:

1. Highlight the risks associated with LBMs in BCI applications
2. Propose mitigation strategies to address these risks
3. Contribute to the responsible development and application of these advanced powerful large models

By raising awareness and providing recommendations, we hope to ensure that the societal benefits of this technology are realized without adverse consequences.

2 Background

Artificial Intelligence (AI) is one of the most advanced technological innovations of the last decades and has the potential to truly transform our lives and society. Today, AI can be met in several aspects of our daily activities with its applications ranging from social media to education. As AI becomes increasingly embedded in our society, its impact will continue to grow, offering opportunities but also raising safety and ethical challenges. Identifying and dealing with these issues has been an active field of research in the area of AI, especially in domains where deep learning models are deployed in healthcare applications (Lee et al., 2023; Kelly et al., 2019; Zhang & Metaxas, 2023). Some of these challenges are more evident and usually common across fields while others are less obvious and domain-specific.

Over the last couple of years, GenAI and large models have dominated the field, serving as powerful backbones for downstream tasks and demonstrating superior performance across several benchmarks. However, GenAI advancements come with significant ethical and safety challenges that must be addressed to ensure the responsible development and deployment of these large models (Weidinger et al., 2023; Stahl & Eke, 2024), especially when dealing with medical data (Ning et al., 2023).

In the field of neuroscience, the discussion around the safety and ethical implications of advances in neurotechnology and BCI technology traces back to the 2000s (Marcus, S. and Charles A. Dana Foundation, 2002), long before the modern breakthroughs in AI - let alone GenAI and large foundation models for brainwave data. Since then, a number of institutions, bodies and countries have produced reports mapping the landscape of these issues (Ienca, 2021; UNESCO, 2023) and passed bills for protecting brain rights (McCay, 2024).

Nowadays, the advent of deep learning has enabled significant advancements to several BCI paradigms, paving the way for wide-spread adoption of brainwave recording devices. As BCI systems transition from controlled lab environments to everyday use (Niforatos et al., 2024; Banville et al., 2024), there is an increasing demand for deep learning models that can handle arbitrary EEG devices (from research-grade to consumer-grade headsets of various sensor configurations) and manage noisy, corrupted or missing EEG data (Banville et al., 2022). This demand for robust and versatile BCI models can be fulfilled by Large Brainwave Models (LBMs). These models are capable of identifying complex patterns and relationships in EEG data, thanks to their extensive self-supervised pre-training on a wide array of unlabeled datasets. Thus, LBMs are not only equipped with strong adaptability to novel downstream tasks, but are also able to generate synthetic data of high-quality, offering promising solutions for predicting brain activity and reconstructing corrupted brain signals.

The deployment of LBMs in many BCI paradigms raises concerns about the potential misuse of synthetic brain data, user profiling, privacy and equality. As these models become more and more advanced, the risks associated with them grow and their mitigation becomes of imperative importance. In this work, we discuss the safety and ethical challenges that arise at the intersection of LBMs (GenAI) and BCIs. Furthermore, we highlight safeguarding techniques that all stakeholders, ranging from BCI researchers to policymakers, should follow to ensure an alignment between the benefits of LBMs and the necessary protections.

3 Challenges

The development of LBMs comes with numerous safety challenges that demand serious consideration. These challenges are not only technical but also ethical, as they have significant consequences for privacy, security and societal harmony. As large models of this type evolve further, their potential to influence or compromise individual and collective interests grows, making it crucial to address these issues effectively.

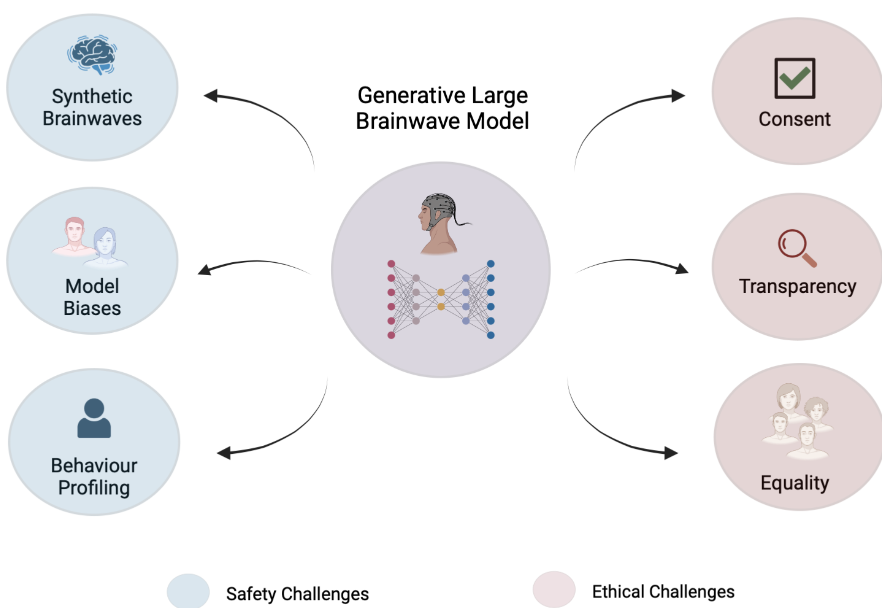

Figure 1: Identified challenges when using Large Brainwave Foundation Models. Blue denotes the identified safety challenges and pink denotes the identified ethical challenges.

3.1 Safety Challenges

Synthetic Brainwaves GenAI has the potential to create data with highly convincing but entirely fictional content, for example a fake image or video (Nguyen et al., 2022). This capability presents an alarming risk when applied to brainwave signals. LBMs, trained on massive diverse EEG datasets, could potentially generate synthetic realistic brainwave signals and accurately recreate neural activity patterns, leading to realistic generated representations of the states, actions or intentions of a person.

The exploitation of such technology includes the generation of synthetic neural activity and the creation of deceptive content that could be used to manipulate systems and/or individuals. The ability to create synthetic brainwave signals also raises concerns about the authenticity of brainwaves and the potential for misuse in sensitive contexts, such as settings where brainwaves are used as a biometric security feature (Zhang et al., 2021).

Model Biases LBMs are trained on large volumes of EEG data which may reflect certain biases, resulting in unfair model outputs. If data selection and cleaning are not performed with bias prevention in mind (Verma et al., 2021), these large models might amplify these biases, when utilised in real-life applications outside the controlled lab environment. For example, if a BCI system is used to decode neural signals for decision-making purposes (Bhattacharyya et al., 2021; Poli et al., 2014), biases in the underlying model could lead to unfair or discriminatory outcomes, affecting certain groups and reinforcing existing inequalities.

Behavioral Profiling LBMs have the potential to provide unprecedented insights into certain mental states, preferences and behavioral patterns of an individual. While this capability can be used in various applications offering significant benefits, such as improved interaction for those with disabilities (Padfield et al., 2024; López-Hernández et al., 2019; Millán et al., 2010), it also poses significant risks related to unauthorized behavioral profiling. For example, insights about personal preferences (Costa-Feito et al., 2023) can be exploited for targeted advertising or manipulation.

3.2 Ethical Challenges

Consent As in every deep learning pipeline that uses data for training as well as inference purposes, informed consent and data sovereignty are critical ethical concerns. Questions about who owns this data and how it is used are often raised by the users. These questions are particularly important when sensitive data is involved, such as brainwave signals. Proper and informed consent becomes particularly important when dealing with LBMs that have the potential to access and interpret these sensitive brainwave data. Individuals must fully understand the implications of data collection and use of their brainwave signals (Bannier et al., 2021; Eke et al., 2022): how this data will be used and the potential risks involved - for example unauthorized access to neural signals could enable the profiling of the emotional states or cognitive processes of a user, potentially leading to privacy violations and exploitation. Ensuring that users have control over their data and that it is handled according to their preferences and legal standards is essential for maintaining ethical integrity.

Transparency Transparency is a critical element when deploying these LBMs, especially in cases where BCI systems are part of decision-making processes (Tveit et al., 2023). When these large models make decisions that impact their users, or provide information that support human decision-making, understanding the mechanisms behind them is critical for accountability (Cheong, 2024). However, many LBMs are perceived as "black boxes", making the decision-making process ambiguous and incomprehensible while undermining the trustworthiness of BCI systems as well as their potential benefits.

Equality LBMs have the potential to transform the way we connect, communicate and interact with one another. They hold the promise to augment our abilities by enabling users to process information, express ideas, and understand each other in ways that were previously unimaginable. However, their high cost combined with the expensive EEG devices and headsets required for their use, make them accessible only to those who can afford them, raising concerns about inequality. This could lead to a society where only a few privileged benefit from GenAI-powered BCIs, deepening social and economic disparities.

4 Safeguarding Techniques

Addressing these safety and ethical challenges requires a multidisciplinary approach that includes innovative technological solutions, proper regulatory frameworks and critical ethical considerations. As LBMs continue to advance, significant efforts to mitigate these risks will be essential in ensuring their responsible and beneficial deployment to real-life applications. The balance between technological advancement and responsible deployment is vital in ensuring that LBMs benefit society while avoiding harmful consequences.

Robust Models Deploying these LBMs in real-world applications requires ensuring not only the effectiveness and reliability of their decoding performance, but also its robustness and security against potential sophisticated attacks (Meng et al., 2023). Using adversarial training methods (Chen et al., 2024), these models can become more robust and capable of handling malicious inputs in the real-world. Adversarial training occurs during the learning phase of the model, where it is exposed not only to standard input data but also to carefully crafted adversarial examples (Goodfellow et al., 2015). It can be used in conjunction with other regularization techniques to improve the robustness and generalization capabilities of a model. Additionally, proactive red teaming (Perez et al., 2022) plays a crucial role in identifying potential safety risks in models before they can be exploited. Organizations developing LBMs should incorporate red teaming practices early in the development lifecycle to effectively mitigate these risks.

Interpretable and Explainable Models When deployed in real-world applications, LBMs should possess certain properties, such as interpretability and explainability, which are essential for building trust and ensuring their safe deployment. Moving away from the "black-box" approach where model decisions are not comprehensible by its users, GenAI BCI researchers should develop LBMs featuring interpretable insights to further validate their capabilities based on well-established neuroscience knowledge. Ensuring that LBMs are explainable and interpretable can help in identifying and potentially resolving issues like biases or inaccuracies. Finally, these properties can further increase the trust of users to such BCI systems, thus making them more reliable and easier to comprehend.

Ethical AI Frameworks As AI, and more specifically large models, become an integral part of our everyday lives, many governments around the globe have started designing legislation and frameworks to ensure that deep learning models operate within ethical boundaries, prioritizing user safety, privacy and fairness. BCI researchers should follow comprehensive guidelines, such as those provided in the AI Act of the EU (European Union, 2024) and the AI regulation framework of the UK (Gov, UK, 2023), when designing and developing LBMs. Such guidelines can assist in addressing issues like bias, accountability and the responsible use of these large models. By implementing these guidelines, BCI researchers ensure that ethical considerations are embedded in the design and operational stages of LBMs, promoting responsible technological advancement and limiting the associated risks.

Privacy When sensitive personal data is involved, such as brainwave signals, certain mechanisms need to be put in place during the data collection as well as the deployment phase to ensure the privacy of the users. Methods such as brainwave data encryption (Liu et al., 2020) and anonymization should be employed in both phases to protect sensitive biometric data and prevent unauthorized access, for example in security settings. These encryption techniques should be implemented in every step of the BCI pipeline ranging from on-device operations to the cloud provider since LBMs typically require considerable computational power to operate. Additionally, proper and informed consent should be obtained from all users prior to their involvement with a BCI system. Especially during the data collection phase, rigorous consent protocols should be implemented to ensure that the brainwave data will be used according to user expectations and ethical standards (World Medical Association, 2013).

5 Conclusion

In this work, we explored Large Brainwave Foundation Models through the lens of safety. While Generative AI and LBMs introduce unique opportunities in the field of BCIs, these models also come with several safety and ethical concerns. Implementing appropriate training and deployment techniques, ensuring the interpretability and reliability of models and addressing ethical disparities are vital to developing these large models so that they are both powerful and responsible. These efforts will help ensure that such technology benefits society more broadly while minimizing risks and ethical concerns.

References

Elise Bannier, Gareth Barker, Valentina Borghesani, Nils Broeckx, Patricia Clement, Kyrre E Emblem, Satrajit Ghosh, Enrico Glerean, Krzysztof J Gorgolewski, Marko Havu, et al. The open brain consent: Informing research participants and obtaining consent to share brain imaging data, 2021.

Hubert Banville, Sean UN Wood, Chris Aimone, Denis-Alexander Engemann, and Alexandre Gramfort. Robust learning from corrupted EEG with dynamic spatial filtering. *NeuroImage*, 251: 118994, 2022.

Hubert Banville, Maurice Abou Jaoude, Sean UN Wood, Chris Aimone, Sebastian C Holst, Alexandre Gramfort, and Denis-Alexander Engemann. Do try this at home: Age prediction from sleep and meditation with large-scale low-cost mobile EEG. *Imaging Neuroscience*, 2:1–15, 2024.

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.

Saugat Bhattacharyya, Davide Valeriani, Caterina Cinel, Luca Citi, and Riccardo Poli. Anytime collaborative brain–computer interfaces for enhancing perceptual group decision-making. *Scientific Reports*, 11(1), August 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-96434-0.

Xiaoqing Chen, Ziwei Wang, and Dongrui Wu. Alignment-Based Adversarial Training (ABAT) for improving the robustness and accuracy of EEG-based BCIs. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.

Ben Chester Cheong. Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, 6, July 2024. ISSN 2673-2726. doi: 10.3389/fhumd.2024.1421273.

Aroa Costa-Feito, Ana M González-Fernández, Carmen Rodríguez-Santos, and Miguel Cervantes-Blanco. Electroencephalography in consumer behaviour and marketing: a science mapping approach. *Humanities and Social Sciences Communications*, 10(1):1–13, 2023.

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.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423.

Damian O Eke, Amy Bernard, Jan G Bjaalie, Ricardo Chavarriaga, Takashi Hanakawa, Anthony J Hannan, Sean L Hill, Maryann E Martone, Agnes McMahon, Oliver Ruebel, Sharon Crook, Edda Thiels, and Franco Pestilli. International data governance for neuroscience. *Neuron*, 110(4): 600–612, 2022.

European Union. EU Artificial Intelligence Act, 2024. URL `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689`. Accessed: 2024-08-28.

Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In Yoshua Bengio and Yann LeCun (eds.), *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

Gov, UK. A pro-innovation approach to AI regulation. *GOV.UK*, 2023. URL `https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach`.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16000–16009, 2022.

Marcello Ienca. Common human rights challenges raised by different applications of neurotechnologies in the biomedical field. *Report commissioned by the Council of Europe*, 2021.

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.

Christopher J Kelly, Alan Karthikesalingam, Mustafa Suleyman, Greg Corrado, and Dominic King. Key challenges for delivering clinical impact with artificial intelligence. *BMC medicine*, 17:1–9, 2019.

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, 2018.

Peter Lee, Sebastien Bubeck, and Joseph Petro. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *New England Journal of Medicine*, 388(13):1233–1239, 2023.

Yongshuang Liu, Haiping Huang, Fu Xiao, Reza Malekian, and Wenming Wang. Classification and recognition of encrypted EEG data based on neural network. *Journal of Information Security and Applications*, 54:102567, October 2020. ISSN 2214-2126. doi: 10.1016/j.jisa.2020.102567.

Jesús Leonardo López-Hernández, Israel González-Carrasco, José Luis López-Cuadrado, and Belén Ruiz-Mezcua. Towards the recognition of the emotions of people with visual disabilities through brain–computer interfaces. *Sensors*, 19(11):2620, 2019.

Marcus, S. and Charles A. Dana Foundation. *Neuroethics: Mapping the Field : Conference Proceedings, May 13-14, 2002, San Francisco, California*. Dana Press, 2002. ISBN 9780972383004.

Allan McCay. Neurorights: the Chilean constitutional change. *AI & SOCIETY*, 39(2):797–798, 2024.

Lubin Meng, Xue Jiang, Jian Huang, Zhigang Zeng, Shan Yu, Tzyy-Ping Jung, Chin-Teng Lin, Ricardo Chavarriaga, and Dongrui Wu. EEG-based brain–computer interfaces are vulnerable to backdoor attacks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31: 2224–2234, 2023.

J d R Millán, Rüdiger Rupp, Gernot R Müller-Putz, Roderick Murray-Smith, Claudio Giugliemma, Michael Tangermann, Carmen Vidaurre, Febo Cincotti, Andrea Kübler, Robert Leeb, et al. Combining brain–computer interfaces and assistive technologies: state-of-the-art and challenges. *Frontiers in neuroscience*, 4:161, 2010.

Christian Mühl, Camille Jeunet, and Fabien Lotte. EEG-based workload estimation across affective contexts. *Frontiers in neuroscience*, 8:114, 2014.

Thanh Thi Nguyen, Quoc Viet Hung Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen, Thien Huynh-The, Saeid Nahavandi, Thanh Tam Nguyen, Quoc-Viet Pham, and Cuong M. Nguyen. Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding*, 223:103525, October 2022. ISSN 1077-3142. doi: 10.1016/j.cviu.2022.103525.

Evangelos Niforatos, Tianhao He, Athanasios Vourvopoulos, and Michail Giannakos. Democratizing EEG: Embedding electroencephalography in a head-mounted display for ubiquitous brain-computer interfacing. *International Journal of Human–Computer Interaction*, pp. 1–25, 2024.

Yilin Ning, Salinelat Teixayavong, Yuqing Shang, Julian Savulescu, Vaishaanth Nagaraj, Di Miao, Mayli Mertens, Daniel Shu Wei Ting, Jasmine Chiat Ling Ong, Mingxuan Liu, Jiuwen Cao, Michael Dunn, Roger Vaughan, Marcus Eng Hock Ong, Joseph Jao-Yiu Sung, Eric J Topol, and Nan Liu. Generative Artificial Intelligence in healthcare: Ethical considerations and assessment checklist. *ArXiv*, abs/2311.02107, 2023.

Natasha Padfield, Andrei Agius Anastasi, Tracey Camilleri, Simon Fabri, Marvin Bugeja, and Kenneth Camilleri. BCI-controlled wheelchairs: end-users' perceptions, needs, and expectations, an interview-based study. *Disability and Rehabilitation: Assistive Technology*, 19(4):1539–1551, 2024.

Ethan Perez, Saffron Huang, H. Francis Song, Trevor Cai, Roman Ring, John Aslanides, Amelia Glaese, Nat McAleese, and Geoffrey Irving. Red teaming language models with language models. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (eds.), *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pp. 3419–3448. Association for Computational Linguistics, 2022. doi: 10.18653/V1/2022.EMNLP-MAIN.225.

Sergio Pérez-Velasco, Eduardo Santamaría-Vázquez, Víctor Martínez-Cagigal, Diego Marcos-Martínez, and Roberto Hornero. EEGSym: Overcoming inter-subject variability in motor imagery based BCIs with deep learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:1766–1775, 2022.

Gert Pfurtscheller and FH Lopes Da Silva. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical neurophysiology*, 110(11):1842–1857, 1999.

Huy Phan, Oliver Y Chén, Minh C Tran, Philipp Koch, Alfred Mertins, and Maarten De Vos. XSleepNet: Multi-view sequential model for automatic sleep staging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5903–5915, 2021.

Riccardo Poli, Davide Valeriani, and Caterina Cinel. Collaborative brain-computer interface for aiding decision-making. *PLoS ONE*, 9(7):e102693, July 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0102693.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. 2018.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PMLR, 2021.

Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering*, 16(5):051001, 2019.

Ranganatha Sitaram, Tomas Ros, Luke Stoeckel, Sven Haller, Frank Scharnowski, Jarrod Lewis-Peacock, Nikolaus Weiskopf, Maria Laura Blefari, Mohit Rana, Ethan Oblak, et al. Closed-loop brain training: the science of neurofeedback. *Nature Reviews Neuroscience*, 18(2):86–100, 2017.

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.

Bernd Carsten Stahl and Damian Eke. The ethics of ChatGPT – Exploring the ethical issues of an emerging technology. *International Journal of Information Management*, 74:102700, February 2024. ISSN 0268-4012. doi: 10.1016/j.ijinfomgt.2023.102700.

Jesper Tveit, Harald Aurlien, Sergey Plis, Vince D Calhoun, William O Tatum, Donald L Schomer, Vibeke Arntsen, Fieke Cox, Firas Fahoum, William B Gallentine, et al. Automated interpretation of clinical electroencephalograms using artificial intelligence. *JAMA neurology*, 80(8):805–812, 2023.

State University of New York (SUNY) UNESCO, Milan-Bicocca University. *The Risks and Challenges of Neurotechnologies for Human Rights*. UNESCO Publishing, 2023. ISBN 9789231005671.

Sahil Verma, Michael Ernst, and Rene Just. Removing biased data to improve fairness and accuracy. *arXiv preprint arXiv:2102.03054*, 2021.

Jacques J Vidal. Toward direct brain-computer communication. *Annual review of Biophysics and Bioengineering*, 2(1):157–180, 1973.

Marco Vilela and Leigh R Hochberg. Applications of brain-computer interfaces to the control of robotic and prosthetic arms. *Handbook of clinical neurology*, 168:87–99, 2020.

Xiaoxi Wei, A. Aldo Faisal, Moritz Grosse-Wentrup, Alexandre Gramfort, Sylvain Chevallier, Vinay Jayaram, Camille Jeunet, Stylianos Bakas, Siegfried Ludwig, Konstantinos Barmpas, Mehdi Bahri, Yannis Panagakis, Nikolaos A. Laskaris, Dimitrios A. Adamos, Stefanos Zafeiriou, William C. Duong, Stephen M. Gordon, Vernon J. Lawhern, Maciej Sliwowski, Vincent Rouanne, and Piotr Tempczyk. 2021 BEETL Competition: Advancing transfer learning for subject independence & heterogenous EEG data sets. In Douwe Kiela, Marco Ciccone, and Barbara Caputo (eds.), *NeurIPS 2021 Competitions and Demonstrations Track, 6-14 December 2021, Online*, volume 176 of *Proceedings of Machine Learning Research*, pp. 205–219. PMLR, 2021.

Laura Weidinger, Maribeth Rauh, Nahema Marchal, Arianna Manzini, Lisa Anne Hendricks, Juan Mateos-Garcia, Stevie Bergman, Jackie Kay, Conor Griffin, Ben Bariach, Iason Gabriel, Verena Rieser, and William S. Isaac. Sociotechnical safety evaluation of Generative AI systems. *ArXiv*, abs/2310.11986, 2023.

Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. Brain–computer interfaces for communication and control. *Clinical neurophysiology*, 113(6):767–791, 2002.

World Medical Association. WMA Declaration of Helsinki: Ethical principles for medical research involving human subjects, 2013. URL `https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/`. First adopted in June 1964, and last amended in October 2013. Accessed: 2024-08-30.

Drishti Yadav, Shilpee Yadav, and Karan Veer. A comprehensive assessment of brain computer interfaces: Recent trends and challenges. *Journal of Neuroscience Methods*, 346:108918, 2020.

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.

Shaoting Zhang and Dimitris Metaxas. On the challenges and perspectives of foundation models for medical image analysis. *Medical Image Analysis*, pp. 102996, 2023.

Shuai Zhang, Lei Sun, Xiuqing Mao, Cuiyun Hu, and Peiyuan Liu. Review on EEG-based authentication technology. *Computational Intelligence and Neuroscience*, 2021(1), January 2021. ISSN 1687-5273. doi: 10.1155/2021/5229576.