Towards Brain-Computer Interface-Based Information Systems: Leveraging Machine Learning for Cognitive Interaction

PhD Thesis

Zenon Lamprou

NeuraSearch Lamboratory
Computer and Information Sciences
University of Strathclyde, Glasgow

May 23, 2025

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Abstract

The field of computational neuroscience has seen significant growth, driven by the development of sophisticated machine learning algorithms. These advancements allow for detailed analysis of brain signals, helping researchers to discover new properties and phenomena within the human brain. Progress in machine learning, especially with the introduction of the transformer architecture and Large Language Models (LLMs), has revolutionized Natural Language Processing (NLP) by achieving unprecedented results. The continuous evolution of these methods highlights the ongoing advancements in NLP technologies and applications.

Beyond NLP, machine learning models like Data2Vec and Wav2Vec2 have broadened possibilities in image and video processing by integrating multiple data modalities. These models improve multi-modal frameworks, enhancing the capability to interpret various inputs, such as combining textual and visual data in question-answering systems. This integration signals a paradigm shift in machine learning, exemplified by advancements like voice-activated assistants and Google Lens, which offer innovative interaction methods. This thesis explores the potential of brain-computer powered interfaces for user interaction through cognitive processes as an alternative way of interacting with a computer system, recognizing the need to address foundational challenges to advance this cutting-edge field. Due to the early stages of this innovative discipline, substantial groundwork is required to identify and systematically resolve the multifaceted challenges intrinsic to the development of such advanced interaction systems, thereby establishing a robust foundation for future advancements in this intriguing domain.

In the Chapter 1 of this thesis, a thoroughly developed introduction is presented.

This section fulfils several critical functions: it outlines the structure of the thesis, providing readers with a coherent roadmap of the thesis's contents and the trajectory of the forthcoming discussion. Additionally, it concisely summarizes the research achievements to date, offering a retrospective overview of the progress attained during the investigation. The chapter also examines the foundational motivation driving the research effort, clarifying the reasoning behind the development of the proposed system. Central to this section is the expression of the core research questions that the thesis aims to explore, which are essential in steering the scholarly inquiry and contributing to the broader academic dialogue.

Within Chapter 2 of the present thesis, a comprehensive literature review has been rigorously executed to encompass a broad spectrum of the most esteemed brain imaging modalities, along with their recent advancements within the multidisciplinary sphere of neuroscience. This extensive analysis also highlights certain contemporary developments in machine learning that are pertinent to the field of neuroscience. Moreover, it presents particular tools, such as Data2Vec, which primarily are not being specifically intended for interaction with neural data, bear potential utility in the conceptualization and design of such a complex system.

In Chapter 3, the proposed system is introduced in detail. This chapter offers an exhaustive description of the system, articulating all of its essential components and anticipating potential challenges that may arise during its development. A schematic high-level design of the system is presented. It is crucial to emphasize that the full implementation of such an ambitious system lies beyond the scope of this thesis. Chapters 4 through 6 provide a thorough investigation into three critical components of this complex system. Each component is meticulously examined to clarify the challenges encountered, the recent progress made, and the subsequent improvements achieved to integrate these components seamlessly into the comprehensive system framework.

In Chapter 4, we employ a rigorous analytical methodology to evaluate the congruence of advanced natural language processing models with empirical neuroscientific data. This evaluation is imperative for the prospective implementation of models capable of interpreting human cognitive processes. Our findings validate that while certain

models do not achieve complete congruence with brain data, they demonstrate a significant level of alignment, thereby affirming the proficiency of current state-of-the-art models in acquiring intricate representations.

Chapter 5 embarks on a comprehensive investigation into the domain of text generation derived from neurological data. This investigation is motivated by two primary considerations. Firstly, text serves as one of the most prevalent means through which individuals interact with computational systems. A variety of mechanisms have been developed to facilitate this interaction with both precision and security. Adhering to the principle of eschewing the duplication of existing mechanisms, it was hypothesized that if a model could be developed to generate text from neural signals, it could be seamlessly integrated with existing systems. The brain's capacity to provide precise lexical terms could, therefore, enhance search engine inputs by mitigating issues related to query disambiguation.

Secondly, within the domain of Neuroscience, the ambition to generate text from brain activity, particularly through non-invasive neuroimaging techniques, has persisted as a longstanding intellectual venture among many researchers. As presented in Chapter 5, the preliminary attempt to actualize this type of generation involved utilizing transformer models to synthesize brain features and subsequently applying a large language model to produce text. Diverse activation functions were employed to optimize the performance of the processing pipeline. This approach was based on the observation that conventional activation functions predominantly assume linearity within the data at some stage of the training process, as demonstrated by their graphical representations.

Nonetheless, empirical studies have indicated that the aforementioned assumption fails to hold in real-world data situations, notably within meteorological datasets. The empirical assessment of various activation functions identified that those based on polynomials, along with polynomial functions featuring adjustable constants fine-tuned during the training phase, emerged as the most effective. These results represent a significant advancement in the neural data-to-text transformation framework, presenting a novel dimension to the fields of computational and neuroscientific research.

In the course of advancing this thesis, particularly in the forthcoming Chapter 6, an exploration is undertaken into the development of a novel brain encoder. This encoder seeks to establish a comprehensive and generalized modelling framework capable of effectively learning and encapsulating general features pertinent to the cognitive processes by which the human brain interprets language. The impetus for this research trajectory stems from the findings of Chapter 5, wherein, despite surpassing existing baseline results, there remained a significant gap in achieving authentic brain-to-text decoding capabilities. The hypothesis informing this investigation posits that the limitation does not reside in the incorporation of specific neural features derived from the transformer encoder model. Chapter 6 further explores how advanced methodologies such as Data2Vec and Wav2Vec2 might be harnessed to formulate such holistic brain 'embeddings.' Previous efforts in this domain have predominantly focused on utilizing generic embeddings for mental state classification, with minimal focus on their potential in the generation domain. Consequently, Chapter 6 articulates and implements a systematic pipeline designed to construct these generic embeddings, systematically applying them to the complex domain of brain-to-text conversion processes, thereby offering a novel perspective and making a significant contribution to the field.

In Chapter 7, we present the design of a novel interface that offers a dual contribution to both the Neuroscience and Machine Learning communities. The primary objective of this thesis was to develop a software system that enables users to interact through brain activity. We have designed and implemented a chatbot interface capable of concurrently capturing brain data while functioning as a conventional chatbot. Furthermore, this chatbot is engineered to be adaptable and highly customizable with millisecond precision, allowing it to serve as a bridge between machine learning and neuroscience, as well as a platform for further neuroscience-focused data collection.

In the final Chapter 8 of this thesis, a comprehensive and detailed synthesis of the research findings is meticulously articulated, with particular emphasis on the systematic addressing of each research question. Additionally, this chapter scrupulously delineates the current limitations that hinder the development of such an innovative system. A thorough and comprehensive report is subsequently presented, providing robust and

practical guidelines for leveraging the insights derived from this research. Such a report is crucial for constructing a framework upon which future scholarly research can build to ultimately achieve the completion and realization of this sophisticated system.

Contents

Abstract **ii**

List of Figures **x**

List of Tables **xiii**

Preface/Acknowledgements **xv**

1 Introduction **2**

1.1 Introduction 2

1.2 Motivation 4

1.3 Thesis statement 5

1.4 Research Questions 7

1.5 Publications 7

1.6 Overall Layout and Outline 8

2 Literature Review **10**

2.1 Introduction 10

2.2 Neuroscience 11

2.2.1 fMRI 11

2.2.2 EEG 17

2.2.3 Intra-cortical microelectrode arrays. 26

2.2.4 Neurolinguistics 27

2.2.5 NeuraSearch 29

2.3 Machine Learning 37

2.3.1 LLMs, Transformers And Different Modalities 37
2.3.2 Activation Functions 50
2.3.3 Information Need and Recommender Systems 55
2.4 Brain-To-Text Decoding 57

3 Brain-Computer Interface Neural Information System 68
3.1 Introduction 68
3.2 Brain recorder 69
3.3 Frame Handler 71
3.4 Word Perceived Model 73
3.5 Brain To Text Decoder 74
3.6 Rationale for choosing the focus area 79
3.7 Chapter Summary 81

4 Development of an fMRI-Based Brain Encoder and Identification of Brain Alignment in Machine Learning Models. 83
4.1 Introduction 83
4.2 Methodology 85
4.2.1 Data 85
4.2.2 Transformer models 86
4.2.3 Experimental Procedure 90
4.3 Results and discussion 100
4.4 Chapter Summary 106

5 Optimizing Brain Decoding using different Activation Functions 108
5.1 Introduction 108
5.2 Methodology 110
5.2.1 Data 111
5.2.2 Model 112
5.2.3 Activation Functions 113
5.2.4 Training process and Evaluation 118
5.2.5 Evaluation Metrics 119

5.3 Results . . . 121
5.4 Chapter Summary . . . 126

6 Development of a Generic Brain-to-Text Decoding Module 128
6.1 Introduction . . . 128
6.2 Methodology . . . 130
6.2.1 Introduction . . . 130
6.2.2 Data . . . 131
6.2.3 LLM substitution . . . 136
6.2.4 CTC integration . . . 138
6.2.5 Wav2Vec2 and Data2vec Implementation . . . 142
6.2.6 Bendr And EEG-Conformer Integration . . . 146
6.3 Results . . . 147
6.4 Chapter Summary . . . 152

7 Design and Development of a Customizable Brain-Powered Chatbot System 154
7.1 Introduction . . . 154
7.2 System architecture . . . 155
7.2.1 User Interface . . . 156
7.2.2 Flexibility and Customization . . . 157
7.2.3 Utilizing Different LLMs . . . 158
7.2.4 Logging Mechanism . . . 159
7.2.5 PyLSL integration . . . 160
7.3 Conclusion and Future Work . . . 161
7.4 Chapter Summary . . . 162

8 Conclusion And Future Work 164
8.1 Introduction . . . 164
8.2 Chapter Outline . . . 164
8.3 Contributions . . . 165
8.3.1 High Level Architecture . . . 165

8.3.2 Are the current state-of-the art NLP models brain aligned 166
8.3.3 Effectively train NLP models on brain data 169
8.3.4 Brain-To-Text Decoder 170
8.3.5 NSChat 172
8.4 Future Work 173

A Optimizing Brain Decoding using different Activation Functions 177

Bibliography 177

List of Figures

3.1 The proposed BCI-NIS architecture is outlined in this Figure, comprising several key components. Initially, the Brain Decoder captures real-time brain data from the user. This data is managed by the Frame Handler, which stores and preprocesses the most recent frames securely. The frames are then analyzed by the first AI layer to ascertain if they form a user-conceived word. Finally, the core component decodes the frames to generate text, which can be sent to a virtual assistant or a search engine. 69

3.2 A high level overview of the 3 main components that are used to construct the Brain-To-Text Decoder. 79

4.1 A brief description of the architecture of every model used in our experimental procedure. 90

4.2 A high level representation of the pipeline used to extract NLP features from the different NLP models. 91

4.3 This figure depicts the procedure involved in loading the natural language processing features until the reconstruction of neural data from the established features. The NLP features are temporally aligned with the neural features, extraneous edges are eliminated, and a ridge regression model is subsequently trained. 94

4.4 This diagram explains using a graphical interfaces the process of evaluating the predictions. 96

4.5 This Figure illustrates how we changed a sample text depending our 4 different implementation scenarios on testing the role of punctuation in text comprehension in the brain. 99

4.6 This figure shows the overall accuracy across all subjects for the four different models we investigated. It is clearly shown that RoBERTa and DistiliBERT outperform the other 2 models and the baseline and are the most brain-aligned models. 101

4.7 On the left hand side are the original results reported in [1]. On the right hand side are the reproduced results we obtained when running our code for BERT. Note that the y-axis has a different minimum and maximum value for the two panels, original and reproduced. 102

4.8 The figures presents the results of the 4 different punctuation scenarios. The immediate observation is that the first 6 layers are more brain aligned than the last 6 and also that as the context length increases the role of punctuation to understand the text meaning is less significant. . 105

5.1 This figure shows the flow used to fine-tune BART alongside a transformer encoder and generate text from brain data. 120

6.1 This figure shows the enhancement from the previous implementation illustrated in Figure 5.1 with the introduction of several state-of-the art LLMs 137

6.2 This figure illustrates how CTC loss was integrated to the pipeline with the hope of learning positional alignment of characters and brain data. For each time step the log probabilities of each character in the vocabulary were calculated and then the CTC loss was calculated between the predicted and actual sentence. 139

6.3 This figure shows the architecture of a Conformer Neural Network as proposed by Gulati et Al. [2] and how Convolution Layer can be integrated with a Multi Headed Attention Layer. 142

6.4 This Figure shows the proposed architecture for a Wave2Vec2 model training regime as proposed by Baevski et Al. [3] 143
6.5 This figure illustrates the proposed architecture for training effectively a Data2Vec model as proposed by Baevski et Al. [4] 143

7.1 Basic Usage of the NSChat System: The user transmits a message and subsequently receives a response, which they then evaluate as pertinent by selecting the thumbs-up icon.. 157
7.2 This Figure showcases the customization available for the NSChat in a set of parameters that can be set by the user. 159
7.3 This figure illustrates that the majority of events in NSChat exhibit an offset of less than 1 millisecond. 161

List of Tables

5.1 Rouge Scores results table 123
5.2 In this table we present our BLEU score results for 1-Gram and 2-Gram BLEU configuration. 124

6.1 List of Model Names Tested as an enhancement to our pipeline. 138
6.2 Performance comparison of different techniques and scenarios using EEG and IMA metrics (BLEU, ROUGE). 148
6.3 P-values for statistical significance comparison between different methods against the Random and Plain baseline. 150

A.1 Rouge Scores results table 178

Preface/Acknowledgements

I am profoundly grateful to all those who have supported me with unwavering dedication throughout this exhilarating and demanding voyage that is my PhD journey. This thesis stands as a testament to the invaluable guidance, unfaltering encouragement, and relentless assistance provided by numerous remarkable individuals.

First and foremost, I extend my deepest thanks to my esteemed supervisors, Dr. Yashar Moshfeghi , whose mentorship and staunch support have been nothing short of transformative. Their unparalleled expertise, boundless patience, and insightful feedback have been the bedrock of shaping my research vision and navigating the intricate labyrinths of my scholarly work. I am especially indebted to him for his profound encouragement during the most challenging times, and his unwavering belief in my potential has fortified my resolve to persevere.

I extend my profound gratitude to my second supervisor, Professor Frank Pollick, whose invaluable insights in the field of Neuroscience have been instrumental. His expertise significantly facilitated my introduction from a background in Computer Science to the Neuroscience domain and was essential for the successful completion of this thesis.

My heartfelt gratitude also reaches out to my esteemed colleagues and cherished friends within the NeuraSearch Lab. The spirit of collaboration and camaraderie fostered in this vibrant group has significantly enriched my research experience, making it both enjoyable and exceedingly fulfilling.

On a deeply personal note, I extend my sincerest appreciation to my family for their unyielding support and steadfast encouragement. To my dear parents, Despo and Christakis, words cannot adequately capture my gratitude for the foundational values of

hard work and perseverance that you have instilled in me. Your unconditional love and countless sacrifices have been my constant source of motivation and inspiration. Your understanding, patience, and unwavering faith in me during this demanding period have meant the absolute world to me.

Thank you all for being integral parts of this significant and transformative chapter in my life. Feel free to enhance any sections or infuse specific details that resonate with your own experiences!

Chapter 0. Preface/Acknowledgements

Chapter 1

Introduction

1.1 Introduction

This thesis addresses a fundamental challenge in current human-computer interaction (HCI) paradigms: the limitations imposed by traditional input methods such as keyboards, mice, voice commands, or touchscreens. These conventional modalities, while functional, restrict the fluidity and intuitiveness of communication with digital systems. To overcome these limitations, this research investigates the development of a Brain-Computer Interface (BCI) system that facilitates interaction and control of digital platforms using only cognitive activity, thereby eliminating the need for physical interaction.

The study presents a comprehensive exploration of a system that allows the user to engage directly with technology via brain signals alone. The primary objective is to establish a seamless interaction mechanism that relies solely on the user's cognitive faculties. By focusing on the integration of Brain-Computer Interfaces (BCIs) with machine learning, particularly in the context of natural language processing (NLP), this research aims to create a novel framework for digital interaction. The research delves into the feasibility of decoding neural signals into coherent text, with the dual goals of enhancing human-computer communication and contributing to foundational scientific understanding. The creation of such a system hinges on the effective combination of brain signal data and machine learning models capable of real-time interpretation and

translation of human thought.

Moreover it outlines a blueprint for the construction of such a system, with a particular emphasis on one core component: the machine learning-based decoding of brain signals into natural text. This focus is selected both because it represents a critical functional challenge and because the complete system exceeds the developmental scope of a single doctoral study.

Based on this focus, the thesis explores how various brain modalities—such as fMRI and EEG—can be combined with contemporary AI tools in the NLP domain. Rather than attempting to redefine existing text-based HCI pipelines, this research proposes a novel input layer: text generation from brain activity. This addition can integrate seamlessly into existing HCI infrastructures while offering an enriched, non-physical modality for interaction.

The use of machine learning models for tasks like mental workload assessment and emotion detection has been investigated in earlier studies [5–7]. Even though scientists have successfully manage to classify such states a long standing achievement for true Brain Computer Interface (BCI) is to be able to generate speech from text [8, 9]. Currently in the domain of NLP ,LLMs excel on generating text amongst other tasks [10,11]. Nevertheless, little research has been done on the connection between the development of LLMs and the brain mechanisms underlying language comprehension and how these models can be utilised to achieve brain to text generation [12].

Moreover large volumes of data are produced by the brain imaging methods, which can be difficult for people to understand. Recent developments in deep learning and artificial intelligence, however, have created models that can efficiently handle these data and identify complex patterns in them [13, 14]. This capacity could have a big impact on how we comprehend the human brain in the future and help solve its many mysteries. Researchers may be able to better understand intricate neural processes by utilizing these cutting-edge AI technologies, which could ultimately result in revolutionary discoveries in cognitive neuroscience and related fields.

Finally this research aims to contribute to our understanding of language cognition and laying the groundwork for more intuitive AI applications. In doing so, it also brings

us closer to a new generation of more adaptive, general-purpose, and real-time BCIs that enable users to interact with digital systems using only their thoughts. Advances in brain data acquisition—such as fMRI, EEG, and MEG—have enabled diverse experiments that identify brain functions ranging from emotional processing [15–17] to semantic comprehension [18, 19]. These techniques provide the foundational data upon which AI models can be trained. As neuroscience uncovers the structural and functional complexities of cognition, machine learning has evolved in parallel, inspired by biological processes to develop models capable of reasoning and language generation. The emergence of LLMs has brought us closer than ever to human-level linguistic performance in machines, including reasoning, coherence, and even passing the Turing Test.

1.2 Motivation

The emerging confluence of neuroscience and artificial intelligence offers unparalleled prospects for advancing computational technologies as well as enhancing our comprehension of human cognition. Furthermore, recent developments in machine learning have facilitated interactions with computer systems in modes that transcend text input, including the utilization of voice commands and images as a search medium.

This thesis undertakes an investigation into a promising area by addressing a notable gap in the prevailing research landscape, specifically the integration of brain modalities with LLMs to examine their potential utility in enabling mechanisms for text generation from brain data and facilitating interaction with computer systems. Furthermore, it devises a comprehensive blueprint delineating the essential components and inherent challenges in constructing such a system. Lastly, it attempts to address certain challenges associated with a pivotal component in the architecture, namely the brain-to-text generator.

Despite significant advances in machine learning applications for tasks such as mental workload assessment and emotion detection, a chasm remains unabridged: the comparative analysis of language processing in both human brains and LLMs. This undertaking constitutes the primary motivation for our research. By establishing cor-

relations between how humans and machines comprehend language, we aspire to bolster the development of AI applications that better mimic human-like understanding, leading to more authentic and nuanced natural language interfaces. The potential for such advancements holds particular promise in the realm of BCIs. As we innovate methods to convert raw brain data into coherent natural text through AI, the resulting Bi-directional communication could dramatically transform the interaction between humans and machines. The intricate conversion of neural signals into language, if achieved, could mark a transformative milestone in empowering individuals with communication impairments, enabling real-time dialogue through BCIs.

Moreover, the methodological advancements anticipated from this work promise substantial contributions to cognitive neuroscience. The expansion of techniques such as fMRI, EEG, and MEG, paired with robust AI models, could elucidate the neural correlates of language processing. By deciphering these patterns, the thesis aspires to lay the groundwork for novel insights into the cognitive processes underlying language comprehension.

In summary, this thesis is wishes to explore the intersection of neuroscience and artificial intelligence, particularly the integration of brain modalities with LLMs to enable text generation from brain data and improve human-computer interactions. It addresses gaps in current research by proposing a system that facilitates bi-directional communication between the brain and computers, potentially benefiting individuals with communication impairments. The work focuses on mapping language processing in humans and LLMs to enhance AI's mimicry of human understanding. The research holds promise for BCIs and aims to advance cognitive neuroscience by employing AI models with neuroimaging technologies like fMRI, EEG, and MEG to better understand language comprehension processes.

1.3 Thesis statement

This thesis explores the novel concept of developing an advanced system designed to enable seamless user interaction exclusively through cognitive functions. The research concentrates chiefly on BCIs and seeks to promote progress in direct mental interaction

with digital platforms, thereby offering a unique user experience.

The primary objective of this thesis is to provide a detailed framework for a system that integrates machine learning models to translate human thoughts into textual data derived from brain signals. This domain was recognized as a critical functional component essential for the system's development, selected because of its broad scope, which surpasses the limits of a single PhD thesis.

In pursuit of this vision, the thesis investigates the intersection of various brain modalities with contemporary AI tools, particularly in the domain of natural language processing. Given the extensive body of research on text-based computer interactions, this thesis contributes a novel dimension by proposing models for text generation, positing their potential utility in enhancing current human-computer interaction frameworks.

Despite significant progress in employing machine learning to assess mental workload and identify emotions, the generation of spoken text from neural inputs continues to be an essential objective for authentic BCIs. NLP has achieved significant progress in text generation, largely facilitated by LLMs. However, there is limited investigation into the interrelationship between these models and the brain's mechanisms for language comprehension.

This thesis seeks to fill this gap by employing neuroimaging technologies such as fMRI, EEG, and MEG. It argues that the development of machine learning models utilizing brain data will elucidate neural processes, thereby offering significant insights into the field of cognitive neuroscience.

The foundational motivation for this thesis illustrates the emerging synergy between neuroscience and AI, providing unparalleled opportunities for the advancement of computational technologies and the deepening of human cognitive understanding. The incorporation of brain modalities with LLMs suggests transformative applications that facilitate brain data-assisted text generation for the enhancement of human-computer interfaces. In conclusion, this thesis seeks to pioneer the direct communication between the brain and computers, specifically addressing individuals with communication impairments while augmenting the field of human-computer interaction by aligning AI

applications with human-like language processing emulation. It aims to contribute significantly to both cognitive neuroscience and AI, establishing foundational frameworks for cutting-edge technological applications and enhancing humanity's cognitive understanding.

1.4 Research Questions

This thesis aims to understand one main question :

1. **To what extent can we design and implement a system that enables users to generate coherent natural language text from brain signals in real time, thereby facilitating interaction with a computer without physical input?**

To achieve this main research question we divided this thesis to address some sub-questions. The answer to these questions should pave the path towards achieving the construction of the system.

1. **Neural-Linguistic Alignment**: How closely do the neural representations of natural language in the brain align with the internal representations of language in large language models (LLMs)?
2. **Feasibility of Brain Data Training** : Can machine learning models, particularly deep learning architectures, be effectively trained on neuroimaging data (e.g., EEG, fMRI) to learn mappings between brain activity and language representations?
3. **Real-Time Brain-to-Text Decoding**: : Is it possible to decode raw brain signals into accurate and contextually appropriate natural language text in real time using current AI techniques?

1.5 Publications

1. Role of Punctuation in Semantic Mapping Between Brain and Transformer Models [20]. - Published in ACAIN 2022

2. On the Role of Activation Functions in EEG-To-Text Decoder [21]. - Published in ACAIN 2024

3. Customizable LLM-Powered Chatbot for Behavioral Science Research [22]. - Submitted in SIGIR 2025

4. On Creating A Brain-To-Text Decoder [23]. - Submitted in Arxiv

1.6 Overall Layout and Outline

This thesis is divided into 8 chapters.

Chapter 1 : Introduction

Chapter 1 aims to provide an overall introduction to the thesis. It describes in general what this thesis is about and outlines also the motivation behind it. It also outlines the thesis statement and the research objectives for this thesis.

Chapter 2 : Literature Review

Chapter 2 presents a concise and thorough literature review around every pillar of this research. It provides a thorough research on 3 different brain imaging techniques (fMRI, EEG, electronic arrays), and how these techniques have been used so far on training machine learning models. It also provides a blueprint on the landscape of machine learning models outlining all the latest advancements in the field. Finally, it outlines the state-of-the-art models used in NLP and how they can be used or have already been used with brain data or in the field of Neuroscience.

Chapter 3 : Core System

Chapter 3 seeks to present a comprehensive implementation of a system enabling user interaction through brain activity. It further aims to construct a conceptual framework of the entire system and to elucidate potential challenges inherent in its development. Lastly, it offers a synopsis of the system components that will be examined in subsequent sections of the thesis.

Chapter 4 : fMRI Encoder

Chapter 4 introduces the fMRI Encoder. The fMRI Encoder is a collection of models that by utilising a novel technique showed to be brain aligned in terms of how they

understand and encode language. We present the methodology to produce such fMRI encoders and we present our results that show the validity of our approach.

Chapter 5 : Activation Functions

Chapter 5 explores the application of various activation functions for improved fitting of brain data and more efficient training of deep learning models on such data. We detail our methodology and outcomes when employing diverse activation functions within the same model architecture, highlighting how our findings differ based on the activation function used.

Chapter 6 : Brain Decoder

Chapter 6 details the implementation of the Brain Decoder, outlining the methodology employed to achieve real-time brain decoding within an open vocabulary framework. It further presents the experimental results and insights gleaned from utilizing the Brain Decoder.

Chapter 7 : NSChat Chapter 7 delineates the implementation of NSChat. This chapter provides a detailed exposition of the system and its present features. It explicates the usage and advantages of this system in comparison to other extant systems. Lastly, it presents a thorough analysis of all the prospective new features anticipated in subsequent versions of the system.

Chapter 8 : Conclusion

Chapter 8 encapsulates the key points of the document, emphasizing the importance of the results, and frequently presents suggestions or implications for further research or application. This chapter concludes the document by connecting back to the introduction, ensuring that the reader grasps the broader message.It also outlines the potential directions for further research, possible improvements or expansions upon the current study, and related questions that remain unanswered. It might identify limitations of the present work and suggest how future studies could address these limitations. Additionally, it could propose applications of the research findings in real-world contexts or explore how emerging technologies might impact the research area.

Chapter 2

Literature Review

2.1 Introduction

This chapter presents an extensive and thorough literature review, forming an integral component of this thesis.It begins with a foundational presentation, of a chapter that meticulously acknowledges and categorizes the quintessential studies associated with each specific brain-imaging technique. This categorization is executed with methodical precision, advancing sequentially through Sections 2.2.1 to 2.2.3. Thereafter, Section 2.2.4 provides an in-depth and detailed exploration of Neurolinguistics. Within this section, significant emphasis is placed on current pioneering research efforts concerning EEG-to-text decoding methodologies, as further articulated in Section 2.4. This section highlights the crucial importance of converting neural signals into coherent and meaningful linguistic forms.

Subsequent to Section 2.2.5, a succinct overview of the nascent field of "NeuraSearch" is presented. NeuraSearch represents an evolving discipline focused on the integration of neuroscience with information retrieval through the application of machine learning techniques. This chapter delivers a comprehensive yet concise examination of the current state of the field, emphasizing the applications and interconnections between machine learning and neuroscience. It seeks to provide insights into the intersection of these domains, showcasing recent developments and potential future research trajectories within NeuraSearch. By investigating the synergies between these disciplines,

this work aims to enhance the understanding of NeuraSearch and its relevance to the advancement of both fields.

Finally, in its concluding segments, this review examines the integration of LLMs within the framework of EEG-to-text decoding systems, as discussed in Section 2.2.4. It further contemplates the profound potential of LLMs to synergize proficiently with various modalities, as outlined in Section 2.3.1. The chapter's meticulously structured organization not only compiles and synthesizes the existing body of literature but also adeptly identifies critical research gaps while offering insightful prospects for future scholarly explorations. Thus, it provides a robust and substantial foundation for the ensuing chapters of this academic thesis.

2.2 Neuroscience

2.2.1 fMRI

fMRI stands as a sophisticated, non-invasive neuroimaging modality designed to quantify and map cerebral activity through the detection of alterations in cerebral blood flow and oxygenation levels, which are inherently linked to neuronal activation dynamics. The principal mechanism underpinning fMRI technology is the blood-oxygen-level dependent (BOLD) contrast, which adeptly leverages the intrinsic magnetic property disparities between oxygenated and deoxygenated haemoglobin molecules. In physiological scenarios where neurons in distinct cerebral regions exhibit activity, there is a heightened consumption of oxygen molecules, necessitating an augmented influx of blood to maintain physiological homeostasis. This physiological response culminates in an increased proportion of oxygenated to deoxygenated blood within neural substrates, manifesting as an observable transformation in the magnetic resonance imaging signal.

Conceived in the early 1990s—a pivotal period in neuroimaging history, as evidenced by foundational work [24]—fMRI has profoundly transformed neuroscientific inquiry by delivering high-precision spatial resolution imaging capabilities, which facilitate the visualization of active cerebral regions during diverse cognitive operations or states of quiescence. Crucially, unlike neuroimaging techniques employing ionizing

radiation, fMRI remains inherently safe for iterative application in human populations. Typical fMRI experimental paradigms encompass the longitudinal acquisition of image sequences, chronicling the haemodynamic cascade that ensues subsequent to neuronal excitation.

Despite the fact that the temporal acuity of fMRI is comparatively constrained vis-à-vis methodologies such as EEG, its capacity to yield granulated spatial maps of brain function compensates for this limitation. fMRI has played an indispensable role in advancing cerebral understanding, illuminating complex cognitive processes, including but not limited to, linguistic processing, memory encoding, attentional mechanisms, and affective modulation, as documented in substantial scholarly repositories [25–27]. In clinical paradigms, it has been instrumental in evaluating neurological function within pathological states and shaping surgical strategies by delineating critical functional zones [28–30]. Accordingly, fMRI is now a quintessential instrument in both experimental and applied contexts, affording profound insights into the intricate interplay of human cognition and neural architecture.

Regarding fMRI experiment design, a prevalent methodology involves juxtaposing brain activity across dual conditional paradigms [31]. Customarily, one paradigm constitutes a rigorously controlled baseline, whereas the alternate paradigm embodies the cognitive process under scrutiny. Illustratively, to elucidate cerebral semantic processing, experimental conditions frequently compare narrative comprehension against word list processing [31]. Nonetheless, divergent research methodologies [32–36] have adopted narrative forms to probe into cognitive semantics in greater depth.

An innovative study conducted by Wehbe et al. (2014) [37] represents a pioneering attempt to analyse the intersection of cognitive neuroscience and AI through the lens of narrative processing, illuminating how the brain assimilates semantic information. This was achieved by having subjects engage in reading textual sentences while their neural activities were meticulously captured using fMRI. By leveraging this approach, the study provides a novel bridge linking cognitive science with AI principles, offering a crucial framework to interpret and potentially enhance state-of-the-art NLP architectures.

The research intricately dissects the operation of four noteworthy NLP models: ELMo [38], Universal Sentence Encoder (USE) [39], Bidirectional Encoder Representations from Transformers (BERT) [40], and Transformer-XL [41]. Each model exemplifies diverse structural paradigms and methodologies for processing linguistic data. By scrutinizing these algorithms in juxtaposition with the neural activities captured from human subjects, this work provides unparalleled insights into the comparative mechanisms between synthetic and biological neural networks.

Hence, this examination permits an in-depth understanding of the parameters influencing word and sequence representation across myriad model facets such as depth of layers, context lengths, and attention modalities. The investigation reveals distinct variations in context-related information representation, elucidating the inherent strengths and deficiencies across examined models. Notably, the interactions between layer complexity and context magnitude, as well as attention typologies in transformer architectures, offer profound insights into the hierarchical organization of language processing within artificial neural networks.

A central hypothesis proposed in this research postulates that aligning NLP architectures with the brain's processing methodologies could substantially augment their linguistic comprehension capabilities. To substantiate this hypothesis, a particular emphasis was placed upon BERT, due to its extraordinary performance in numerous NLP domains. By modifying its architecture to simulate brain processing patterns as revealed by fMRI, a brain-aligned version of BERT was generated. This adapted model was rigorously evaluated through syntactic NLP tasks, demonstrating superior performance over its conventional counterpart. This pivotal finding corroborates the feasibility of brain-inspired architectural adjustments, heralding potential enhancements within machine learning paradigms through insights extracted from cognitive neuroscience.

Furthermore, the broader implications of this research transcend mere performance optimization, establishing a paradigm of model interpretability and biological alignment that opens pathways for AI systems with heightened intelligibility and plausibility relative to human cognitive processes. Such advancements portend the future devel-

opment of NLP models capable of more authentically processing language, leading to substantially refined human-AI interactions.

Moreover, the interdisciplinary synergy propounded in this paper could stimulate groundbreaking progress in both AI and neuroscientific inquiries. For the neuroscience community, contrasting artificial with biological neural mechanisms unveils new vistas in understanding cerebral language comprehension. Concurrently, for AI developers, the invaluable insights garnered from neural imaging exemplify an untapped reservoir for innovating more proficient and streamlined linguistic algorithms.

Thus, the work spearheaded by Toneva et al. is acknowledged as a monumental leap towards crafting more interpretable and biologically congruent NLP frameworks. Demonstrating the fertility of interdisciplinary research that melds cognitive neuroscience with AI, this study embarks on new trajectories for enriching machine learning models whilst advancing our comprehension of both biological and machine-driven language processing. As the NLP landscape continues to evolve, the principles and strategies delineated herein are expected to significantly influence the trajectory of future language understanding systems, propelling the creation of AI that mirrors human cognitive processes in language perception and interpretation.

Another exemplary study that sought to amalgamate neurological signals, more specifically through fMRI, with language models is the investigation by Jain and Huth (2018) [42]. This research constitutes a noteworthy progression in deciphering the mechanisms through which the human brain interprets linguistic information. It addresses a pivotal shortcoming in pre-existing language encoding models for fMRI by integrating contextual information, thereby forging a connection between artificial language structures and the innate processes of human linguistic comprehension. Typified by the conventional methodologies, encoding models for fMRI data employed word embeddings that considered each lexical stimulus in isolation, neglecting the inherent contextual dynamics intrinsic to language comprehension. In contrast, Jain and Huth's methodology confronts this paradigm by utilizing comprehensive contextual representations that originate from Long Short-Term Memory (LSTM) language models.

This innovative approach facilitates a more refined and precise modelling of cerebral

reactions to linguistic stimuli. The methodological innovation propagated by the authors yields substantial enhancements in encoding accuracy across nearly all examined cerebral regions, performing superiorly compared to contemporary word embedding frameworks. These improvements in performance can be ascribed to two fundamental elements: the elevated calibre of word embeddings formulated by the LSTM model and the deliberate integration of contextual nuances. By meticulously modulating the magnitude and calibre of context incorporated within their models, the scholars elucidate the indispensable role of contextuality in predicting cerebral activity during linguistic processing. One of the most noteworthy contributions of this investigation is the intricate delineation of context sensitivity throughout the cortex, providing unprecedented insights into the processing and amalgamation of contextual information by various cerebral regions during the comprehension of language. The outcomes imply a hierarchical stratification of language processing within the brain, with distinct regions exhibiting variable sensitivity to contextual cues.

Furthermore, the study establishes a substantive interconnection between artificial language constructs and human cognitive processes. The abstract representations discerned by LSTM language models reveal strong concordances with those observed within the human brain, positing that these artificial constructs might be successfully encapsulating cardinal elements of human linguistic processing. This correlation engenders new investigative avenues in BCIs and engenders the evolution of more biologically plausible language models. The ramifications of this scholarly work extend well beyond the domain of neuroscience, reverberating into the realms of AI and NLP. By compellingly demonstrating the significance of context in modelling cerebral reactions to language, the study offers critical insights for the enhancement of language models and the development of advanced AI systems adept in replicating human-like linguistic understanding and generation. Jain and Huth's research delineates a substantial leap forward in comprehending the neurological basis of language processing.

By assimilating contextual intelligence into fMRI encoding paradigms, they have not only augmented the precision of predictions regarding cerebral activities but also established an innovative framework for examining the neural substrates underpinning

language comprehension. This pioneering work paves the way for future inquiries probing the complex interplay between synthetic language models and human cognition, potentially catalysing progress in brain-computer interface technologies, advanced language comprehension solutions, and therapeutic approaches for language disorders.

The pioneering research conducted by Caro et al. [43] represents a remarkable advancement in the integration of fMRI with cutting-edge machine learning methodologies. At the heart of this study lies the innovative Brain Language Model (BrainLM), a sophisticated analytical framework purpose-built to facilitate the comprehension and exploration of extensive brain activity datasets obtained from fMRI recordings. By applying self-supervised masked-prediction training strategies, BrainLM is adeptly trained on a substantial dataset encompassing more than 6,700 hours of fMRI data.

This training endows BrainLM with the capacity to execute an array of tasks proficiently, such as refining predictive models for clinical variables and forecasting future neural states, while also performing zero-shot inference to distinguish functional networks and articulate latent, interpretable representations of neuronal dynamics. A salient feature of BrainLM is its adeptness at task-specific adaptive learning, exhibited through fine-tuning processes that enable the model to assimilate additional data for enhanced task performance. This is particularly beneficial in medical arenas, where BrainLM can predict critical clinical metrics from observed neural patterns.

Furthermore, its zero-shot inference capability exemplifies its ability to extend its computational understanding of cerebral dynamics without necessitating data tailored to specific tasks, a trait instrumental for delineating functional networks in the brain and unearthing the interactions of different regions during diverse cognitive activities. The introduction of a groundbreaking prompting mechanism further enhances BrainLM's utility, enabling it to simulate brain responses to hypothetical perturbations in silico. Through this functionality, researchers are empowered to scrutinize speculative scenarios, projecting the impact of alterations in precise brain regions on overarching neural activity schemas.

The capability for simulating these dynamics underpins BrainLM's role as a vital analytical instrument, enriching the scholarly discourse on brain function complexities

and fostering novel approaches to interpreting expansive brain activity data. The profound implications of these findings lie in BrainLM's establishment as a comprehensive, scalable framework for brain activity analysis. This framework not only broadens the horizons of our understanding of neural dynamics but also opens new avenues for practical clinical applications, such as informing treatment strategies through predictive brain state modeling in neurological pathologies.

Moreover, BrainLM's proficiency in generating interpretable neural activity models heralds deeper inquiries into the fundamental processes underpinning cognition. In essence, BrainLM's introduction into the scientific landscape marks a watershed moment in neuroscience and neuroimaging domains. By capitalizing on voluminous fMRI repositories and avant-garde machine learning paradigms, this model provides a holistic apparatus for the investigation of cerebral dynamics. Its adeptness at fine-tuning and zero-shot inference, when coupled with progressive simulation strategies, positions BrainLM as an invaluable asset to both empirical researchers and clinical practitioners in quest of elucidating and deciphering the intricate functionalities of the brain.

Consequently, this body of work lays the foundational stone for future explorations intended to demystify the complex tapestry of human thought and to propel advancements in comprehending neural disorders.

2.2.2 EEG

As elucidated in Section 2.2.1, fMRI has been a predominant modality in the scientific pursuit to map and decode the intricacies of human cognitive processes, particularly in linguistic functions and thought pattern decoding. fMRI's adoption in elucidating insights derived from LLMs underscores its relevance in neuroscience and computational linguistics spheres. Despite its contributions, fMRI is beset with inherent limitations that curtail its utility. Notably, the BOLD signal, a cornerstone of fMRI-based studies, is hampered by a significant temporal delay of approximately 5-6 seconds, thereby impeding real-time data acquisition.

Furthermore, the exorbitant costs associated with fMRI data acquisition present substantial financial barriers, especially given the burgeoning data demands imposed

by the evolution of LLMs, which are now architected with billions of parameters necessitating vast datasets for successful training. Consequently, the scientific community has pivoted towards an earlier, simpler modality: EEG. EEG emerges as an invaluable non-invasive neuroimaging methodology employed to measure and document the brain's electrical activity. By strategically affixing electrodes onto the scalp, EEG facilitates the capture of spontaneous neural electrical discharges, predominantly emanating from cortical neurons. This technical facet allows for the real-time monitoring of brain activity, rendering it indispensable for gaining insights into both cognitive operations and a spectrum of neurological conditions.

EEG's operational modality is particularly instrumental in diagnosing epilepsy and other seizure-related disorders, owing to its proficiency in detecting abnormal electrical discharges, typified by spikes or sharp waves, indicative of seizure activity. Despite the progression in imaging technology advancements, EEG retains critical importance owing to its superior temporal resolution capabilities, with the ability to track brain dynamics on a millisecond scale [44]. The genesis of EEG dates back to the early 20th century, marked by Hans Berger's pioneering recording of the inaugural human EEG in 1924. Since this seminal advancement, EEG technology has undergone significant evolution, culminating in diverse applications that transcend conventional clinical diagnostics. In addition to its role in seizure disorder identification, EEG is harnessed for sleep pattern analysis [45, 46], brain tumour diagnostics [47, 48], coma state assessments [18, 49], Alzheimer's disease symptoms predictions [50] and encephalopathy evaluations [51, 52].

Its application scope extends further into research domains, where it is employed to probe cognitive functions inclusive of attention, memory consolidation, and multimodal EEG decoding [53]. EEG's competency in capturing rapid fluctuations in brain activity renders it particularly adept at investigating the cerebral response to diverse stimuli or task-oriented engagements. One of the salient advantages of EEG lies in its cost-effectiveness and accessibility, which markedly surpasses that of other neuroimaging modalities such as fMRI and Positron Emission Tomography (PET). The portability of EEG equipment, coupled with its adaptability for diverse environmental

settings, augments its suitability for both clinical and extensive research applications. Technological advancements have further engendered the development of dry electrodes and wireless systems, thereby streamlining the apparatus setup process while concurrently enhancing patient comfort. Consequently, EEG sustains its pivotal role in both deciphering brain function and diagnosing neurological afflictions. The continuum of research within this domain continues to unravel EEG's potential as a tool for cognitive augmentation and brain-computer interface innovation [44, 54].

The investigation carried out by Frank, Otten, Galli, and Vigliocco (2015) [55] offers an insightful and robust contribution to the scientific discourse on the cerebral mechanisms underlying linguistic information processing. This meticulous study remarkably bridges the distinct yet interconnected domains of information theory, neurolinguistics, and EEG technology, providing groundbreaking insights into the neural correlates pertaining to the informational content of words during the process of sentence comprehension. The researchers meticulously applied Event-Related Potentials (ERPs) as a tool to methodically examine the brain's responses to varying degrees of information conveyed by individual lexical items within sentences.

Through the sophisticated integration of EEG recordings with computational metrics delineating information content, they furnish compelling empirical data substantiating the intricate relationship between the information-theoretic attributes of language and neural information processing mechanisms. A pivotal central innovation within this study emanates from their employment of surprisal and entropy reduction as methodological, quantitative indices for gauging a word's informational content. Surprisal, functioning as an indicator of the unexpectedness of a word given its contextual surroundings, alongside entropy reduction, encapsulating the degree by which a word mitigates uncertainty regarding subsequent sentence elements, are both derived from sophisticated probabilistic linguistic frameworks. This innovative methodology facilitates a more intricate and nuanced understanding of cerebral information processing compared to conventional linguistic categorizations.

The empirical results from this investigation elucidate that both surprisal and entropy reduction exert modulatory influences on the N400 component of the ERP, a well-

established neural marker indicative of semantic processing. Specifically, higher levels of surprisal and entropy reduction are correlated with a more pronounced negativity in the N400 amplitudes, implying that words encapsulating heightened informational content are potent elicitors of intensified neural responses. Significantly, these effects are discernible even when accounting for conventionally recognized psycholinguistic variables such as word frequency and close probability.

Moreover, the study elucidates distinct temporal dynamics for surprisal and entropy reduction within the ERP waveform. Surprisal effects manifest in earlier temporal windows, signifying immediate lexical-semantic processing, whereas entropy reduction exerts influence over later stages, indicative of extended sentential integration processes. This differentiation highlights the brain's specialized processing strategies for various aspects of lexical information. The scientific implications of this research are extensive and profound. By establishing a tangible linkage between information-theoretic constructs and neural activity, this study provides robust support for probabilistic language comprehension models. It suggests that the cerebral apparatus is exquisitely sensitive to minor variations in word informational content, dynamically revising its predictive framework and attenuating uncertainty throughout sentence unfolding. Furthermore, this pioneering work unlocks avenues for advancing our understanding of language processing disorders. The methodologies employed may be instrumental in studying the sentence-level information processing of individuals with linguistic impairments, potentially contributing to novel diagnostic techniques or therapeutic interventions.

In the expansive arena of cognitive neuroscience, this investigation stands as a quintessential exemplar of the transformative power of interdisciplinary methodologies. By blending insights across information theory, computational linguistics, and electrophysiological profiling, the authors afford a comprehensive, multifaceted account of language processing beyond the limits of any single disciplinary perspective. Finally, the scholarly contributions of Frank et al. denote a significant leap forward in elucidating the neural foundation of language comprehension and extend beyond mere empirical corroboration of information-theoretic models. The integration of computational models with neurophysiological data underscores the broader potential for refined

and precise explorations into the complex landscape of cerebral language processing, thereby establishing a formidable platform for progressing future interdisciplinary research efforts at the nexus of linguistics, cognitive science, and neuroscience.

In the specialized and intricate realm of neurolinguistics and computational cognitive science, the research conducted by Hale, Dyer, Kuncoro, and Brennan (2018) [56] marks a considerable leap forward in our comprehension of the neuromechanisms underlying syntactic structure processing during language comprehension. This trailblazing study adeptly bridges the divisions between computational linguistics and cognitive neuroscience, employing beam search — a sophisticated algorithm extensively utilized in NLP — to systematically analyse EEG data. The authors' pursuit aimed explicitly at explicating the dynamic interplay between incremental parsing decisions and concomitant brain activity.

Through the strategic application of beam search, they were able to model the sequential, stepwise process integral to sentence parsing, thereby correlating it with the temporal neurological response dynamics. Not only does this methodology afford a profound and intricate understanding of syntactic processing as witnessed in the neural matrix, but it also offers a nuanced comprehension that eclipses the explanatory power of prior analytic frameworks.

One of the paramount discoveries of the study is the discernible and substantive correlation identified between beam search-derived surprisal metrics and corresponding EEG responses. This critical association signifies that the human brain engages in syntax-sensitive predictive processing during the grooming of linguistic structures in comprehension tasks. The methodology, by its merit, demonstrated that the beam search approach surpasses the predictive efficacy of conventional n-gram models in elucidating EEG data variance, thereby offering compelling corroboration for the significance of hierarchical syntactic frameworks in sentence processing.

The methodological approach is distinguished by its groundbreaking implementation of beam search within a neurolinguistic research context. By recalibrating this computational technique to encompass brain data analysis, the authors have facilitated exploration avenues that were previously uncharted in the nexus of cognitive

neuroscience with language processing paradigms. This interdisciplinary venture is exemplary of the fertile exchange of ideas and methodologies between AI and cognitive neuroscience, showcasing an innovative synergy.

Moreover, this research furnishes invaluable insights into the temporal dynamics characteristic of syntactic processing. A detailed analysis conducted by the authors elucidates the manifestation of syntactic surprise across various chronological junctures in the EEG signal, presenting an enriched understanding of the temporal progression inherent in language processing mechanisms in the human brain. Such temporal insight is indispensable for decoding the swift and intricate cognitive processes entailed in human language comprehension.

The research outcomes possess implications transcending the immediate landscape of neurolinguistics. By illustrating the remarkable efficacy of computational models in decoding neural data, this study sets the groundwork for developing more sophisticated cognitive domain analyses through the prism of brain function. It underscores the potential of leveraging advanced NLP methodologies to procure deeper insights into human cognitive processing.

Furthermore, the work contributes saliently to the sustained discourse concerning the nature of language processing within the cerebral domain. The success of utilizing the beam search model to predict EEG responses stands in support of theoretical structures that ascribe a central standing to hierarchical syntactic mechanisms in linguistic comprehension. This decisive finding carries significant ramifications for linguistic theory, language acquisition research, and the advancement of language processing constructs within artificial cognition.

In conclusion, the pivotal research landscape charted by Hale et al.'s study represents a cardinal progression in decoding cerebral language processing intricacies. By adeptly applying beam search to EEG analytical frameworks, the authors have not only unearthed new facets of understanding regarding syntactic processing dynamics but have significantly showcased the substantive merit of synthesizing computational methodologies with neuroscientific inquiry. This seminal work establishes a foundational reference for ensuing research investigations that may further illuminate the

symbiotic relationship between linguistics' structural elements and neural activity, potentially advancing towards more integrative models of language processing encompassing human and AI systems alike. It is poised as an essential referential cornerstone for researchers located at the confluence of linguistics, cognitive science, and AI, imparting both methodological innovations alongside insightful theoretical contributions vis-a-vis enhancing the field's trajectory.

In the rapidly transforming domain of BCIs, the limited availability and diverse nature of EEG data present substantial obstacles for proficient model development and optimal performance outcomes. The contemporary investigation by Cui et al. (2023) [57] introduces Neuro-GPT, an innovative foundational model meticulously engineered to surmount these impediments by employing expansive publicly accessible datasets. This scholarly initiative embodies a paramount progression in the deployment of machine learning methodologies to analyse EEG data, specifically within the parameters of motor imagery classification tasks. Neuro-GPT integrates an EEG encoder in conjunction with a Generative Pre-trained Transformer (GPT) architecture, establishing a robust analytical framework adept at accommodating heterogeneous EEG data inputs.

This foundational model undergoes a pre-training phase using a self-supervised learning protocol centred on reconstructing occluded segments of EEG data. This self-supervised paradigm is especially advantageous in contexts characterized by a scarcity of labelled data, permitting the model to extract significant insights from the extant data corpus without necessitating extensive manual annotation procedures.

The efficacy of Neuro-GPT is substantiated through rigorous fine-tuning on a motor imagery classification exercise involving nine participants, evidencing its capability to substantially enhance classification performance relative to models developed de Novo. The findings elucidate that the foundational model not only augments performance in scenarios with limited data availability but also demonstrates adaptability across divergent datasets, effectively addressing the intrinsic variability and complexity inherent in EEG signal patterns. Cui et al. (2023) [57] offer compelling substantiation that the implementation of a foundational model can efficaciously alleviate challenges associated with data paucity and heterogeneity within EEG research domains.

Their results imply that such models represent formidable instruments for advancing BCI applications, facilitating more dependable and precise decoding of neural representations. The authors have opted to make their codebase publicly accessible, thereby promoting transparency and inciting continued scholarly inquiry within the research community. This transparency is consonant with current trends in machine learning scholarship, which accentuate reproducibility and collaborative efforts.

In summation, the advent of Neuro-GPT signifies a substantive leap forward in the realm of EEG-centric BCIs. Through the judicious application of expansive datasets via self-supervised learning mechanisms, this foundational model efficaciously tackles cardinal challenges related to EEG data scarcity and heterogeneity. The auspicious outcomes derived from motor imagery classification tasks highlight its potential applicability across broader disciplines in neurotechnology and rehabilitation science. This initiative paves the way for forthcoming innovations in assistive technologies, with specific pertinence to individuals impacted by motor function impairments.

A comprehensive study aimed at developing generic brain embeddings from EEG data was undertaken by Kostas et al. [58], showcasing an innovative framework intended to substantially enhance the efficiency and effectiveness of BCI systems by utilizing advanced self-supervised learning methodologies. This research meticulously addresses the significant challenges characteristic of conventional BCI systems, which typically rely heavily on annotated datasets that are both restricted in size and circumscribed in diversity. The authors make a compelling argument that prevailing approaches have inadequately exploited the extensive abundance of unlabelled EEG data, a resource that, if harnessed appropriately, could lead to markedly improved generalization of models across a wide variety of contexts, subjects, and tasks. The BENDR (Bert-inspired Neural Data Representations) model derives its theoretical underpinning from successful language modelling strategies, particularly those prevalent in the realm of NLP.

By judiciously incorporating the architectural principles of models akin to Wav2vec 2.0 [3], the researchers encode arbitrary EEG data segments into informed feature vectors identified as BENDR. This pioneering approach facilitates the effective modelling

of raw EEG sequences acquired from diverse hardware configurations and a range of subjects. The architecture of the BENDR framework is bifurcated into two principal segments: an initial phase where raw EEG data undergoes downsampling through a series of convolutional layers, resulting in the generation of BENDR vectors, and a subsequent phase where a transformer encoder transposes these vectors into sequential forms applicable to numerous downstream tasks. This architecture is meticulously fashioned to encapsulate critical features while concurrently maintaining computational efficiency.

A cornerstone of this study is its pronounced emphasis on self-supervised learning, which adeptly enables the model to glean insights from unlabelled data by reconstructing masked portions of the input sequences, thereby significantly augmenting the model's capability to generalize across an array of datasets and tasks devoid of the need for extensive labelled training data. The evaluative component of the study scrutinizes the efficacy of the BENDR model across a spectrum of EEG classification tasks, wherein it demonstrably succeeds in generating resilient representations that surpass the performance metrics of conventional task-specific models. Empirical findings conspicuously illustrate that a singular pre-trained model possesses the adaptability requisite for effective application to novel EEG datasets recorded under variable conditions.

Consequently, this study posits that self-supervised learning methodologies exemplified by BENDR could usher in a substantive transformation within the domain of BCI, facilitating superior utilization of large-scale unlabelled EEG data and, by extension, propelling expansive research initiatives into scalable and adaptive BCI systems designed to confront assorted applications within the spheres of neuroscience and human-computer interaction. In essence, the investigative pursuit by Kostas et al. illuminates the profound potential possessed by self-supervised learning in fortifying BCI systems via the BENDR framework. Through the proficient leverage of vast unlabelled EEG datasets, this methodological approach not only elevates model performance metrics but concurrently lays the groundwork for pioneering advancements in brain-computer interface technology. The insights accrued from this research decidedly augment our

comprehension of the application of sophisticated machine learning techniques to intricate neural data, thereby establishing a robust foundation for the impending exploration within this swiftly advancing domain.

2.2.3 Intra-cortical microelectrode arrays.

The realm of speech neurolinguistics has witnessed significant advancements, particularly in the domain of assistive technologies tailored for individuals grappling with profound speech impairments due to conditions like amyotrophic lateral sclerosis (ALS). A seminal work in this field is presented by [59], which delineates the design and success of a high-performance speech BCI aimed at decoding neural signals associated with speech production. This investigation not only addresses a critical demand for efficacious communication solutions for individuals rendered speechless by paralysis but also sets a new benchmark in neuroprosthetic research.

The study by [59] showcases an extraordinary feat in both decoding accuracy and operational velocity. A reported WER of 9.1% for a constrained vocabulary encompassing 50 words, alongside 23.8% for a significantly broader lexicon of 125,000 words, underscores this achievement. The participant, denoted as T12, achieved a communication velocity of 62 words per minute, a pace nearing natural conversational dynamics. Such an advancement signifies a monumental stride beyond preceding speech neuroprosthetic frameworks, which historically encountered challenges in achieving similar efficiency and precision levels.

A pivotal focus of this study is on dissecting the neural representation of speech within the motor cortex. The findings elucidate that this region is robust in encoding orofacial movements, facilitating effective decoding even amidst the participant's paralysis. This significant insight into the neural machinations underpinning speech production is invaluable for comprehending how BCIs may harness preserved neural activities to enable communication. Furthermore, the employment of Recurrent Neural Networks (RNNs) in predicting phonemes from captured neural signals is noteworthy. By employing sophisticated language models, the authors achieved heightened transcription accuracy. This pioneering methodology not only enables real-time speech

output but also provides a practical conduit for individuals with speech impediments to communicate proficiently. Moreover, the research accentuates the persistence of articulatory codes years post-paralysis, insinuating that intricate details pertinent to phoneme production remain extractable from the motor cortex.

This revelation bears significant implications for future explorations into neural encoding and its potential applications in neurolinguistics. The researchers also deliberate on pivotal design considerations impacting BCI performance, such as vocabulary breadth, electrode density, and training data prerequisites. They advocate the expansion of microelectrodes and the refinement of language models as strategies to bolster system performance and adaptability subsequently. Such insights are imperative for steering future progressions in speech neuroprostheses.

In essence, this research by [59] epitomizes a palpable leap in neurolinguistics, evidencing that high-performance speech BCIs can viably restore swift communication functionalities to individuals with severe speech impairments. Although challenges persist concerning system resilience and long-term applicability, this study offers a promising perspective for imminent innovations dedicated to ameliorating the quality of life for those confronting communication hindrances due to paralysis. These revelations not only enhance the scholarly grasp of speech but also underpin future inquiries aimed at advancing assistive technologies for individuals impeded by speech ailments.

2.2.4 Neurolinguistics

Neurolinguistics stands as a quintessential interdisciplinary field that meticulously explores the intricate relationship between language and the brain. Its core focus lies in understanding how neural mechanisms intricately govern the comprehension, production, and acquisition of language. This field synthesizes insights from an array of disciplines including linguistics, neuroscience, psychology, and cognitive science, to delve into how various brain structures are intricately involved in language processing. Researchers dedicated to neurolinguistics employ a vast array of methodologies, prominent among which are neuroimaging techniques such as fMRI and EEG, aimed at observing brain activity during language-related tasks. This nascent field traces its lin-

eage to seminal studies on individuals presenting with language impairments, notably aphasia, and has progressively expanded to embrace a comprehensive understanding of how the brain organizes, processes, and utilizes linguistic information.

The significance of neurolinguistics transcends mere theoretical inquiry, venturing into impactful practical implications particularly within the realms of education, healthcare, and AI. By discerning the neural underpinnings of language acquisition and associated disorders, neurolinguistics are equipped to devise more effective, individualized teaching strategies that cater to diverse learning needs and enhance diagnostic tools for language-related conditions. Moreover, insights gleaned from neurolinguistics critically inform the burgeoning development of NLP systems within the domain of AI, significantly enhancing their capabilities to comprehend and generate human language. As research endeavours continue to push the frontiers in this arena, neurolinguistics holds a pivotal role in bridging the dichotomy between linguistic theory and neurological function. This ultimately contributes significantly to an enriched understanding of human communication.

From a neuroscience perspective, the field of neurolinguistics remains largely underexplored with respect to the role of punctuation in semantic interpretation and its cognitive processing within text. Despite being a relatively emergent domain, neurolinguistics has witnessed numerous notable efforts aimed at advancing scholarly understanding and probing unanswered questions within its expansive purview.

In relation to the pivotal role of punctuation in the syntactical analysis of language, certain scholarly inquiries have striveded to concentrate on the formulation of multidimensional features that are intricately linked to the syntactical components of language [60]. Subsequently, employing these derived features, scholars attempted to accurately model the syntactical representation of punctuation marks within textual materials. Notably, there exists considerable variability within the literature regarding the treatment of punctuation. For instance, utilizing identical fMRI data yet divergent research aims, one investigation involved preprocessing the text encountered by participants inclusive of punctuation [61], whereas another excluded it entirely [62].

Additionally, distinct research efforts [63] have sought to demonstrate that the amal-

gamation of words can engender more intricate meanings, and have attempted to identify which cerebral regions are accountable for representing such meanings. Within their preprocessing phase, they elected to excise punctuation from text prior to presentation to participants. This evidentiality research further elucidates the absence of consensus regarding the semantic and syntactic processing of punctuation by the brain, and whether its inclusion in a corpus is imperative for achieving optimal outcomes. Moreover, there is a discernible paucity of investigative studies that scrutinize or enhance NLP models via brain recordings, as posited by Toneva and Wehbe [1].

Although there exists research into cognition that evaluates whether word embeddings encapsulate relevant semantics [64], additional research [65] has sought to create novel embeddings that coincide with brain recordings to determine if these embeddings exhibit superior alignment with behavioural measures of semantics. By leveraging cutting-edge models to ascertain how their representations are congruent with brain activity, we can significantly contribute to the expansion of this vital and burgeoning area of research. Through this process, we have the potential to elucidate which specific training choices may enhance alignment with neural data by examining how various training decisions can enhance or diminish this alignment.

2.2.5 NeuraSearch

In recent years, a significant volume of interdisciplinary research has been conducted to identify the potential applications of neuroscience in advancing the field of information retrieval. This emerging area of study has been collectively referred to as NeuraSearch [66], as evidenced by various studies [67–70]. Moshfeghi et al. [71–75] conducted pioneering works to establish the NeuraSearch field by introducing neuroscience methods to improve information retrieval systems through implicit relevance feedback from users' emotional and physiological responses and in turn satisfying searchers' information need. Traditional systems rely on explicit feedback, like clicks or ratings, which may not fully capture user engagement. This study utilizes affective signals such as emotions derived from facial expressions and physiological metrics like heart rate, combined with user interaction patterns. The researchers employed a multi-modal

approach, collecting data on participants' physiological responses during information retrieval tasks. This method allowed them to identify implicit cues that signal user relevance and interest, often overlooked in standard systems. Experimental results showed that incorporating these emotional and physiological features significantly improved the precision of implicit feedback mechanisms, leading to better predictions of user preferences.

The findings suggest that the use of neurophysiological signals can help information retrieval systems adapt more effectively to users, offering personalized recommendations without explicit input. This not only enhances user experience but also aids in developing intuitive search interfaces. The research lays the foundation for further exploration of affective computing integration into information retrieval to create more responsive and user-centric technologies, with potential applications across sectors focused on user engagement.

NeuraSearch embodies the intersection of neuroscience and information retrieval, investigating novel approaches to improve the ways in which information is accessed and utilized [66]. The domain of NeuraSearch applications encompasses a diverse array of investigations across several pivotal areas. A principal area of inquiry pertains to comprehending and fulfilling the information requirements of users, as extensively documented in numerous academic studies [76].Moshfeghi et al. (2016) [73] conducted a seminal investigation into the concept of information need. This study explores the neural correlations of information need utilizing fMRI. The researchers assessed brain activity in 24 participants engaged in a Question Answering (Q/A) task, uncovering a distributed network of brain regions implicated in information retrieval processes. Their results demonstrate distinct patterns of brain activity contingent upon whether participants possessed prior knowledge of the answers or were required to seek them out, thus offering insights into the neurological underpinnings of information needs. This research highlights the multifaceted nature of information needs and signifies a substantial advancement in enhancing user satisfaction in information retrieval systems by elucidating the cognitive processes involved.

In addition Moshfeghi et al. (2019) [74], the authors further their prior research

by employing EEG to forecast users' information needs in real-time during search activities. This investigation underscores the capacity of EEG data to offer immediate insights into users' cognitive states and their assessments of the relevance of search results. By examining brain signals while participants engaged in question and answer tasks, the researchers discerned distinct EEG patterns associated with various phases of information need recognition. The paper advocates for the integration of neurophysiological measures into information retrieval systems to formulate more adaptive and responsive user experiences, thereby accentuating the potential of EEG to augment user profiling and enhance system interactions.

Finally Michalkova et al. (2024) [77] make a significant contribution to this field with their work titled "Query Augmentation with Brain Signals," which investigates the integration of brain signals to enhance search query augmentation in order to optimize relevance and user satisfaction. The authors introduce a framework that synergistically combines conventional query augmentation strategies with insights gleaned from users' neurophysiological responses during search activities. Through the application of EEG data, they illustrate that the incorporation of brain signal information can yield more personalized and contextually appropriate search outcomes. This research underscores the practical advantages of employing neuroimaging techniques to refine information retrieval processes, positing that an understanding of users' cognitive states can greatly improve their interactions with digital systems. This comprehensive text provides a cohesive overview of each cited work, highlighting their contributions to advancing the understanding of user information needs within the scope of NeuraSearch applications.

Another critical application is situated in the formulation of search queries, contributing to the creation of more precise and user-centered searches [78]. Moreover, the user search process itself has been conceptualized as a navigation through varying search states, offering significant insights into search behaviors [79]. A crucial facet of NeuraSearch research involves assessing the potential for utilizing brain activation to enhance relevance feedback [71, 80–82].

Pinkosova et al. [83] examine the neurodegenerative mechanisms underlying Huntington's disease (HD), with a particular emphasis on the dysfunctions within cortical

and basal ganglia circuits. The authors underscore that HD is marked by pronounced neurodegeneration, primarily affecting the striatum and neocortex, which culminates in motor and cognitive impairments. Their discussion includes the loss of GABAergic medium spiny neurons within the striatum and the degeneration of cortical pyramidal neurons as key contributors to these symptoms. Notably, the study highlights that alterations in neuronal function may precede cell death, indicating that initial changes in neural circuits could underlie the early stages of the disease. These findings highlight the imperative for future research into therapeutic strategies that target these neural circuits to impede the progression of HD.

In a subsequent study [84] that extends their prior research, new insights into the structural and functional alterations in cortical circuits associated with HD are provided. This paper systematically reviews recent advancements in the comprehension of how these alterations impact cognitive abilities and motor control in individuals with HD. The authors examine findings from various studies, incorporating imaging and electrophysiological data, which indicate that progressive cortical atrophy and modified neuronal excitability are principal characteristics of HD pathology. Furthermore, potential interventions are explored that might restore normal circuit functionality or mitigate deficits, emphasising the significance of comprehending these mechanisms for the development of effective treatments.

Finally in a final study [85] The authors examine the impact of extended physiological stimulation on blood-brain barrier (BBB) permeability and its association with cortical plasticity. Their findings indicate that such stimulation results in heightened BBB permeability mediated by AMPA receptor signaling, a mechanism essential for synaptic potentiation. This study offers new insights into the role of BBB modulation in facilitating synaptic alterations during learning and memory processes. The authors propose that a deeper understanding of these interactions may guide the development of therapeutic strategies to improve cognitive function in disorders marked by compromised cortical plasticity.

Moreover NeuraSearch concentrates on how neurological responses can yield more refined and nuanced feedback relative to conventional methods. Additionally, an area

of focus is the thorough satisfaction of user information needs through the application of innovative neuroscience approaches [86, 87]. In addition, NeuraSearch emphasizes the detection of mental workload, employing sophisticated techniques to gain a deeper understanding of cognitive load during information retrieval tasks.

Kingphai et al. (2021) [88] provide a comprehensive review of methodologies employed in the evaluation of mental workload through EEG signals. The authors examine how variations in experimental designs and analytical techniques can affect the estimation of workload and underscore the significance of selecting suitable features for precise assessment. Their study accentuates the potential of EEG as a non-invasive instrument for monitoring cognitive load across diverse contexts, such as human-computer interaction and occupational health.

Furthermore, Kingphai et al. (2021) [89] conducted an examination of diverse methodologies for the classification of mental states utilizing EEG data, encompassing preprocessing methods, feature extraction techniques, and classification algorithms. The authors critically analyse the strengths and limitations inherent in existing studies, highlighting the necessity for standardized protocols to augment reproducibility and comparability across research endeavours. They conclude that while EEG demonstrates considerable promise for implementation in brain-computer interfaces and affective computing, it necessitates further methodological refinement to achieve its full potential.

Kingphai et al. (2023) [90] have investigated the problem of identifying the most pertinent EEG channels for the purpose of effective emotion recognition. They introduce an ensemble learning methodology that integrates various feature selection techniques to enhance classification accuracy. The findings suggest that this approach substantially improves emotion recognition performance in comparison to conventional single-channel methods. This research advances the domain of affective computing by offering insights into the optimization of EEG data for real-time emotion detection applications.

McGuire et al. (2023) [91] conducted an investigation into the encoding mechanisms of the anterior cingulate cortex (ACC) regarding sequential action strategies influenced

by recent behavioral choices. The authors elucidate that the dynamics of the ACC are modulated by a summary statistic of previous actions rather than being predominantly driven by the prevalence of rewards. Their experimental findings indicate that the ensemble activity within the ACC monitors both global and local contexts during task performance, positing that the ACC is integral to processes of statistical learning and decision-making. This study significantly advances our comprehension of ACC function within cognitive neuroscience and its relevance to behavioral adaptability.

Last but not least in the pursuit of advancing these applications, several neuroimaging techniques have been integrated into NeuraSearch studies. Prominent among these is MEG, which has been highlighted in scholarly reports for its efficacy in real-time cerebral monitoring [92]. Additionally, fMRI is a widely utilized method, extensively documented for its proficiency in observing neural activity over extended durations and with high spatial resolution [72, 73, 75, 86, 87, 93].

Lamprou et al. (2022) [94] employed fMRI data to examine the influence of punctuation on semantic comprehension in both human brains and transformer models utilized in NLP. The authors emphasize that contemporary neural networks, notably those engineered for NLP, do not adhere to explicit linguistic regulations. Rather, they propose that these models might acquire generic linguistic patterns through the process of training. To investigate this proposition, the study adopts an experimental methodology that leverages human brain recordings to ascertain the potential for establishing a correspondence between cerebral activity and neural network representations.

In their research, the authors evaluate four advanced NLP models to determine which model most closely aligns with human semantic processing. They perform experiments wherein punctuation is systematically removed from text across four distinct scenarios to assess its impact on semantic comprehension. The findings indicate that the RoBERTa model exhibits the highest congruence with brain activity, surpassing BERT in terms of accuracy. Importantly, the results suggest that the removal of punctuation can improve the performance of BERT, highlighting the significant role of punctuation in the semantic interpretation of text by both humans and models. This research contributes to the ongoing investigation of the interplay between linguistic features and

cognitive processing within the realms of AI and neuroscience.

Moreover, EEG has been frequently employed, esteemed for its capability to capture rapid neural responses [78, 89, 90, 95–98]. An examination of the advantages of these techniques and their specific suitability for NeuraSearch is provided in the ensuing sections.

Allegretti et al. (2015) [80] investigated the temporal and neural correlates of relevance judgments in the context of information retrieval tasks through the utilization of EEG. The study seeks to elucidate the mechanisms by which users evaluate the relevance of information during their interaction with digital content. In this research, the authors conducted experiments in which participants were presented with various informational items and tasked with assessing their relevance. By capturing EEG signals throughout these tasks, the researchers sought to pinpoint specific brain activity patterns correlating with the process of relevance judgment formation. The findings reveal that distinct neural markers, particularly within Event-Related Potential (ERP) components, are associated with the timing of these judgments. Furthermore, the study underscores the influence of individuals' self-assessed knowledge on their relevance evaluations, suggesting that confidence in one's comprehension can alter cognitive processing during such assessments.

This research significantly augments the comprehension of cognitive mechanisms involved in the processes of information retrieval and relevance evaluation. Through the correlation of EEG data with distinct decision-making instances, the study advances our understanding of user interactions with information systems and underscores the necessity of integrating cognitive elements in the design of more efficacious retrieval systems. This work establishes a foundational basis for subsequent inquiries into the convergence of neuroscience and information retrieval, potentially facilitating enhanced user experiences within digital contexts.

Jacucci et al. (2019) [96] have introduced an innovative methodology to enhance Information Retrieval (IR) systems by integrating implicit relevance feedback obtained from neurophysiological measures, specifically EEG and eye-tracking. The authors contend that conventional IR systems are predominantly dependent on explicit user signals,

such as clicks and queries, which may inadequately capture the nuanced information needs of users. By leveraging neurophysiological signals, the study seeks to offer a more sophisticated comprehension of user intent and relevance assessments in real-time, thus augmenting the effectiveness of information retrieval processes.

The present study introduces a comprehensively integrated IR system that effectively processes implicit relevance feedback derived from brain activity and eye-tracking data in an online context. An evaluative experiment, conducted with 16 participants, demonstrated that the system is capable of computing neurophysiology-based relevance feedback with performance exceeding chance levels in complex data domains. The authors meticulously describe their methodology, which involves the training of a user-specific classifier designed to predict relevance from EEG signals and eye movements observed during keyword fixation. This classifier functions in two distinct phases: a calibration phase that serves to collect labelled data and an online phase during which real-time relevance predictions are executed. The findings indicate that the integration of neurophysiological feedback into interactive intent modelling significantly enhances the precision of relevance assessments. This advancement paves the way for the development of more adaptive and user-centred IR systems. Moreover, the research underscores the potential of neuroadaptive IR systems to exploit implicit feedback mechanisms without interfering with the user experience, thereby presenting promising implications for subsequent research in this field.

In conclusion, the NeuraSearch field investigates the interdisciplinary domain that amalgamates neuroscience with information retrieval to optimize user interactions with digital systems. Prominent research contributions within this area include leveraging neurophysiological signals to provide implicit relevance feedback, thereby enhancing user experience absent explicit input. Different research attempted the employment of emotional and physiological signals, such as EEG data, to tailor and modify information retrieval processes, thereby markedly enhancing the accuracy of predicting user preferences. Moreover extending these findings by employing brain signals for query augmentation, underscoring the potential of neuroimaging in improving search results. Furthermore, the review underscores the significance of understanding informa-

tion needs through neural activities, as illustrated by examining brain activity patterns during cognitive tasks utilizing fMRI and EEG.

Research also examines the neurological foundations of cognitive states influencing information requirements, mental workload, and search behaviour. For example, some research investigations concentrate on evaluating methods for mental workload, demonstrating the effectiveness of EEG in assessing cognitive load, while others explore the function of the anterior cingulate cortex in decision-making processes. Furthermore, advancements in methodologies utilizing fMRI and EEG highlight the continuous integration of neuroimaging techniques in observing cognitive processing and relevance judgment in information retrieval systems.

2.3 Machine Learning

2.3.1 LLMs, Transformers And Different Modalities

One notable recent advancement in the field of machine learning is the introduction of LLMs, which leverage the Transformer architecture. As these models have evolved, they have increased in size, contributing to their growing complexity. Notably, these models exhibit multimodal capabilities, functioning as generic encoders that integrate diverse modalities such as text and images within a single model framework. This capability for cross-modal integration, decoding, and generation presents new opportunities for utilizing LLMs and the Transformer architecture to decode EEG data and generate text.

The Transformer architecture, as proposed in [99], revolves around the attention mechanism, which overcomes the limitations associated with traditional sequence transduction models reliant on RNNs and Convolutional Neural Networks (CNNs). By employing a pure attention-based model, the Transformer eliminates the necessity for recurrence and convolutions. This architectural choice facilitates enhanced parallelization during training, significantly reducing the time required for model training on extensive datasets. Consequently, the Transformer architecture has emerged as a cornerstone for many state-of-the-art NLP systems, excelling in applications such as machine transla-

tion, text summarization, and question answering.

A fundamental innovation of the Transformer model lies in its employment of self-attention mechanisms, allowing the model to evaluate the relative importance of different words within a sentence. This mechanism functions by transforming input sequences into three critical components: queries (Q), keys (K), and values (V). The dot product computation between queries and keys derives attention scores, permitting the model to concentrate on pertinent segments of the input when generating outputs. Additionally, the integration of multi-head attention, whereby numerous attention mechanisms operate concurrently, empowers the model to discern various intra-data relationships. This feature enhances the model's ability to comprehend context and semantics, thereby augmenting its performance in tackling complex language tasks.

The authors carried out extensive experimental evaluations, substantiating their method by publishing results that demonstrate the Transformer's superiority over existing models across standard machine translation benchmarks. For example, a BLEU score of 28.4 was achieved on the WMT 2014 English-to-German translation task, and a new state-of-the-art score of 41.8 was recorded for English-to-French translation. Such outcomes underscore not only the Transformer architecture's effectiveness but also its training efficiency relative to antecedent models, which necessitated greater computational resources and prolonged training durations.

Moreover, the Transformer's applicability extends past machine translation, as it effectively generalizes to other tasks like English constituency parsing. Beyond architectural innovations, the paper "Attention Is All You Need" delves into practical ramifications for future research within NLP and adjacent domains. The authors posit that attention mechanisms can transcend language processing, finding application in diverse tasks like image recognition, where discerning relationships among elements is essential. This versatility has spurred further advancements, including the development of Vision Transformers (ViTs) [100], which have adapted self-attention principles for image classification tasks by treating image patches analogous to text words. Accordingly, transformers are poised to replace CNNs as the prevailing architecture in computer vision.

In summation, "Attention Is All You Need" has significantly influenced machine learning and AI research. Introducing a streamlined yet potent architecture predicated solely on attention mechanisms, the authors have pioneered progress in NLP and computer vision alike. The Transformer model's ability to efficiently process sequences while capturing intricate dependencies positions it as a foundational framework for developing expansive language models and multichannel AI systems. By 2024, this seminal paper has accrued over 100,000 citations, underscoring its pivotal role in shaping contemporary AI methodologies and applications.

One of the pioneering implementations of the attention mechanism is exemplified through the development of BERT [40]. This transformative model constitutes a significant leap forward in the domain of NLP, as it harnesses the self-attention mechanism inherent to the transformer architecture to perform bidirectional training. Such capacity allows it to process textual data by simultaneously considering contexts from both preceding and subsequent segments, thereby deviating from the earlier models which predominantly processed information in a linear, unidirectional fashion, working from either the left-to-right or vice versa. The multifaceted bidirectional approach as evidenced by the authors, markedly enhances the model's interpretative capabilities regarding linguistic contexts, thereby yielding improved outcomes across a spectrum of NLP tasks.

BERT's preparatory training involves large-scale text corpora and is based on two distinct training objectives: Masked Language Model (MLM) and Next Sentence Prediction (NSP). Within the MLM framework, a portion of the input tokens are subjected to random masking, compelling BERT to anticipate these obscured tokens by leveraging the contextual clues surrounding them. This method plays a critical role in fostering highly detailed contextual embeddings for lexical items. Concurrently, NSP requires the model to ascertain the likelihood of one sentence succeeding another in a continuous context, thereby fine-tuning its grasp of sentence interrelations. Through conditioning on bidirectional contexts during the training phase, BERT effectively achieves profound insights into the subtleties of human language, establishing its application across a diverse array of contexts. The authors note that BERT redefines standards by setting

new state-of-the-art results across eleven NLP tasks, which include but are not limited to question-answering, sentiment analysis, and natural language inference. For instance, BERT accomplishes an unprecedented GLUE score of 80.5%, thereby surpassing previous benchmark records by notable margins. The model's capacity for fine-tuning necessitates minimal customization for task-specific objectives, enabling it to seamlessly transition across various application domains while maintaining exemplary performance.

This notable flexibility makes it an attractive solution for researchers and developers aiming to construct advanced NLP applications without extensive architectural overhaul. Beyond its outstanding performance metrics, BERT has incited considerable scholarly interest, giving rise to the field known as "BERTology," dedicated to exploring the model's internal mechanisms and representations. Scholars are delving into how BERT, through its attention mechanisms, encapsulates linguistic properties and relationships, and how these insights can enhance subsequent models. Moreover, the inception of compact variants like DistilBERT addresses computational resource concerns, facilitating deployment in more constrained environments without sacrificing substantial efficacy. The advent of BERT epitomizes a pivotal milestone in NLP by showcasing the dynamic potential of bidirectional training and transformer-based architectures.

BERT's proficiency in generating nuanced, contextualized representations has set exceptional benchmarks for efficacy across varied NLP tasks and continues to inspire ongoing research into the comprehension and refinement of language models. As BERT's influence perpetuates advancements in NLP, it establishes a foundational platform for future breakthroughs in AI-enhanced language comprehension systems and applications that span across diverse industry sectors.

In their initial application, transformer models and LLMs were primarily deployed within the NLP field, becoming instrumental in various tasks like sentiment analysis [18, 101, 102], word classification [103–105], text generation [106–108], and information retrieval [109–111]. Recent notable advancements in the area of information retrieval involve the use of Retrieval-Augmented Generation (RAG) models [112]. According to

Lewis et al., this model presents a groundbreaking framework tailored to bolster LLM capabilities, specifically in knowledge-intensive NLP tasks. Traditional LLMs, though powerful, exhibit limitations in effectively accessing and utilizing factual knowledge, which hampers their performance in precision-demanding tasks such as information retrieval.

The proposed RAG framework addresses these issues by integrating parametric and non-parametric memory, thus enabling models to dynamically access external knowledge sources. This innovative approach not only enhances the generated responses' accuracy but also tackles challenges related to knowledge provenance and model updates. The RAG framework operates through two primary phases: retrieval and generation. In the retrieval phase, a neural retriever interacts with an external knowledge base, like a dense vector index of Wikipedia, to collect pertinent data corresponding to the user's query. This information subsequently facilitates the generative phase, where a pre-trained Sequence-to-Sequence (seq2seq) model produces context-aware responses. The RAG models are evaluated against multiple knowledge-intensive tasks, including open-domain question answering, demonstrating superior performance over both traditional parametric-only seq2seq models and task-specific retrieve-and-extract architectures.

One of RAG's significant advantages lies in its ability to generate language that is not only specific and diverse but also factual, compared to existing methodologies. By anchoring responses in current external information, RAG produces answers that are both relevant and verifiable, an attribute critical in domains such as medicine and technology. The authors also note that RAG minimizes the need for extensive retraining of foundational LLMs, presenting a cost-effective strategy for organizations intending to enhance their NLP capabilities without substantial computational expense.

Furthermore, the study delves into the broader consequences of merging retrieval systems with generative models. By establishing a dynamic connection between LLMs and external knowledge repositories, RAG promotes more informed decision-making processes within AI applications. This synergy allows for real-time updates to the model's knowledge, circumventing the need for complete retraining, thereby overcoming one of the significant limitations of traditional LLMs: their static nature with respect

to training data. Consequently, RAG introduces a promising pathway for crafting more adaptable and reliable AI systems capable of addressing complex queries across diverse fields.

The advent of RAG signifies a substantial progression in the NLP sector. By adeptly integrating retrieval with generation methodologies, RAG not only elevates LLM performance on knowledge-demanding tasks but also pushes the frontier in efficiency and precision. The framework establishes novel benchmarks for question answering and information retrieval, while concurrently unlocking potential areas for future investigations in multimodal learning and AI-driven applications. As corporations increasingly pursue the deployment of intelligent systems proficient at delivering precise and context-driven insights, RAG stands out as a formidable instrument to meet these objectives amidst a perpetually shifting data and knowledge ecosystem.

The domain of NLP has seen remarkable advancements due to the exceptional capabilities of transformer models, as illustrated in prior examples. These models have shown remarkable prowess in NLP tasks, owing to their intricate architecture. However, the inherent complexity of transformers has driven researchers to explore their potential beyond the conventional realm of NLP. The research conducted by Alexey Dosovitskiy et al. [100] investigates the novel application of transformer architectures within the sphere of computer vision.

The authors have introduced the Vision Transformer (ViT) model, an innovative paradigm that conceptualizes image patches with the same theoretical approach as words in a sentence. By segmenting an image into uniform patches and embedding these segments, the ViT utilizes the transformer architecture to process them, thus enabling it to perform image classification tasks with efficacy that rivals traditional CNNs. This shift in methodology illustrates the transformers' capability of handling visual data without dependence on CNNs, which have historically been predominant in computer vision. The ViT model operationalizes its function by transforming an image into a sequence of flattened patches, further embedding these entities in high-dimensionality contexts.

By analogizing each patch to a token, similar to NLP paradigms, the ViT em-

ploys the attention mechanism to reassess the inter-patch relationships that contribute to a comprehensive understanding of imagery. This not only enhances model performance but also showcases transformative superiority by encapsulating long-range dependencies across the image's multiple aspects. The authors further emphasize the greater scalability and adaptability harnessed through transformers compared to the traditional CNNs, which primarily emphasize localized features. Their experimental endeavours involve the pre-training of the ViT model with expansive datasets, with subsequent testing against benchmarks such as ImageNet and CIFAR-100. Encouragingly, the ViT's outcomes exhibit competitiveness relative to sophisticated CNNs, while also demonstrating an economy of computational resources during the training phase. The implications of these findings suggest that transformers, when trained on extensive data sets, possess the ability to surpass conventional CNN architectures in image classification tasks. This potential indicates a pivotal shift towards redefining computer vision methodologies and prompts further inquiry into utilizing transformers beyond established paradigms. In addition, the authors delve into possibilities facilitated by the application of transformers on a large scale for image recognition. By asserting that pure transformer architectures can operate independently from CNNs and still secure outstanding results, they question prevailing assumptions in the field of computer vision.

This exploration illustrates the immense advantages in terms of performance and suggests that integrating transformers can broaden the scope of numerous vision-centric applications, including object detection and pixel-level segmentation tasks. The inherent versatility of transformers thus encourages novel research trajectories that surpass preceding conventions, advancing AI-driven visual recognition technology. Ultimately, this study offers persuasive evidence supporting the integration of transformer architectures within computer vision disciplines. By verifying the viability of pure transformer models in undertaking image classification duties, the research underscores that such models not only match but potentially exceed the robustness of existing CNNs, offering notable benefits in terms of scalability and versatility. This progression lays crucial groundwork for transforming contemporary approaches to computer vision, urging a

re-evaluation of traditional methodologies to exploit the capacities of groundbreaking transformer models in varied application domains.

The utilization of transformer architectures extends beyond textual applications and enters the realm of Visual Question Answering (VQA) as discussed in foundational works such as [113]. VQA constitutes a sophisticated pursuit within AI, focusing on the provision of answers to open-ended questions predicated on the visual data obtained from images. The core aim of VQA is to empower computational systems to comprehend visual input and produce responses in a natural language format. To achieve this, it necessitates a synergistic integration of computer vision techniques and NLP methodologies, demanding that models simultaneously interpret the visual components of an image along with the linguistic elements of the posed question. VQA architectures are meticulously crafted to generate responses that are not solely accurate but are also contextually pertinent, broadening their utility across several sectors, such as educational frameworks, assistive technologies for individuals with visual impairments, and the refinement of image retrieval methodologies.

A distinctive hallmark of VQA is its engagement with free-form, unconstrained queries that may address particular facets of an image, including but not restricted to background nuances and contextual information. Contrasting the conventional image captioning tasks that typically yield generic summary descriptions, VQA mandates a profound comprehension of the image's intricacies to furnish precise responses. For illustration, when confronted with a query such as "What is the color of the car in this image?" the system is tasked with not only identifying the presence of the car but also delivering an accurate assessment of its colour. Such intricacies render VQA an intellectually demanding yet fruitful discipline, propelling the boundaries of AI's contemporary capabilities.

The progression of VQA has catalysed the compilation of specialized data repositories to support both training and rigorous evaluation undertakings. For instance, a prominent dataset encompasses roughly 265,000 images coupled with an excess of 1 million textual inquiries and 11 million substantiated answers. These datasets are instrumental for the critical benchmarking of VQA systems, providing a basis for com-

parative analyses among divergent methodologies and facilitating the identification of potential enhancements. They commonly include a multiplicity of plausible responses per inquiry to foster resilient evaluative standards and stimulate progress in model precision. VQA's applicability spans a myriad of real-world scenarios. Notably, it can serve as a valuable tool for individuals with visual impairments by offering articulate image descriptions or fulfilling inquiries regarding their environments. In pedagogical contexts, VQA systems stand to revolutionize museum visitations by enabling direct interaction inquiries about the exhibits on display. Moreover, VQA bolsters image retrieval systems by allowing users to locate specific visuals based on descriptive prompts. Its functional reach into multimedia is further augmented by the capacity to retrieve video segments predicated on visual content parameters.

In conclusion, VQA epitomizes a prominent synergy between computer vision and NLP, tasking models with the understanding and analytical reasoning of visual information in conjunction with textual interrogatives. As research in this domain evolves, advancements in VQA are anticipated to significantly bolster AI's interactive capabilities with users, leading to increased accessibility and enriched user engagement across diverse applications. The ongoing evolution of advanced models and expansive datasets will be pivotal in defining the prospective contours of visual intelligence technologies.

The applications of transformers transcend merely processing textual data, as illustrated in the examples provided, by showing their utility across various modalities and their capacity to unify different data types through the use of generic transformer designs. Baevski et al. [4] have introduced an innovative method for creating embeddings applicable to any modality. Their approach, Data2Vec, establishes a comprehensive framework for self-supervised learning applicable to a multitude of modalities, namely speech, vision, and text. The authors elucidate that, despite self-supervised learning being a universal concept, the algorithms and objectives employed have hitherto been tailored to specific domains. Data2Vec seeks to bridge this by employing a consistent learning strategy that forecasts latent representations derived from complete data inputs via masked views, all while utilizing a self-distillation technique situated within a conventional transformer framework. A key innovation of Data2Vec lies in its emphasis

on producing contextualized latent representations instead of targets tied to specific modalities, such as words, visual tokens, or speech elements. This strategic choice allows the model to encapsulate exhaustive insights from the entire data input, thereby elevating its capability to decode complex data structures. The efficacy of this framework is corroborated by extensive experimentation on significant benchmarks relevant to various domains: speech recognition (LibriSpeech), image categorization (ImageNet), and natural language comprehension (GLUE). In these evaluations, Data2Vec matches or exceeds state-of-the-art performance compared to existing methodologies, solidifying its adaptability and effectiveness across modalities.

The architectural details of Data2Vec, as outlined by the authors, integrate a standard transformer model which is further augmented using modality-specific encoding approaches. Specifically, it employs a multi-layer 1-D CNN for processing speech data, adopts a ViT schema for interpreting image inputs, and encodes text through sub-word units. This versatility permits Data2Vec to retain high performance across diverse data types without sacrificing operational efficiency. Data2Vec's capacity for effective pre-training on extensive datasets underscores its prowess, demonstrating remarkable performance improvements even in contexts with limited resources.

Moreover, the paper discusses the broader implications of a unified self-supervised learning architecture for advancing research. By streamlining the training procedure for multiple modalities, Data2Vec not only mitigates the complexity involved in crafting separate models but also enhances generalization capabilities. This innovation lays the groundwork for more integrated AI systems capable of comprehending and processing heterogeneous information sources, ultimately leading to significant advancements in multi-modal applications.

In summary, the advent of Data2Vec signifies a substantial progression toward realizing generalized self-supervised learning techniques applicable across speech, vision, and language disciplines. By harnessing a unified architecture and learning objective, this framework achieves notable advancements on principal benchmarks, promoting straightforward model development and deployment. The work of Baevski et al. establishes a fundamental basis for future endeavours in multi-modal learning and highlights

the potential for evolving sophisticated AI systems that are competent in understanding and interacting with the world through myriad information forms.

Another significant study introduces a comprehensive model for processing various modalities [114], presenting an innovative MultiModal Large Language Model (MM-LLM) capable of generating and interpreting content across different media formats, including text, images, videos, and audio. This model, known as NExT-GPT, developed by the NExT++ research group at the National University of Singapore, addresses a crucial limitation in existing MM-LLMs, which typically prioritize multimodal input comprehension over output generation in different modalities. The authors describe an integrated system that associates a LLM with multimodal adaptors and diffusion decoders, enabling efficient processing and generation across diverse formats.

Built with a robust architecture of established encoders and decoders, NExT-GPT requires minimal parameter adjustments—approximately 1%—in certain projection layers. This approach not only facilitates cost-effective training but also allows for scalability to include additional modalities as required. The model introduces a novel technique called Modality-Switching Instruction Tuning (MosIT), supported by a meticulously curated dataset that enhances cross-modal semantic interpretation and content creation capabilities. Through these technological advancements, NExT-GPT aims to develop an AI system capable of universal modality representation, thereby enhancing human-like interaction capabilities within AI frameworks. NExT-GPT's architecture is composed of three main components: multimodal encoding, LLM processing focused on semantic comprehension, and multimodal content generation. At the encoding stage, high-performance encoders process varied modal inputs, transforming them into language-representative formats understandable by the LLM. Following this, the LLM engages in semantic reasoning, generating outputs based on user input, which include modality-specific tokens that guide the decoding layers to create the appropriate content. This systematic methodology allows NExT-GPT to undertake complex tasks necessitating concurrent understanding of multiple modalities.

The performance of NExT-GPT is promising across several multimodal applications, demonstrating its versatile handling of inputs and outputs among various modal-

ities. For example, users can input an image to receive a descriptive text or a correlated video response. Such capabilities have significant implications for user interaction, potentially broadening application possibilities within educational, entertainment, and content creation domains. In conclusion, NExT-GPT signifies a crucial stride forward in the evolution of multimodal AI systems. By fostering seamless interaction between diverse content types—text, images, videos, and audio—it addresses pressing limitations of current MM-LLMs. With its efficient training strategies combined with an innovative design, NExT-GPT emerges as a pertinent model for driving future development in human-like AI interactions. As ongoing research in this domain progresses, NExT-GPT lays the groundwork for further advancements in multimodal AI technology capabilities and applications.

Transformers have emerged as a versatile framework capable of adapting across various modalities, facilitating advancements in numerous applications. In our research, we leveraged the robust adaptability of transformers to develop an innovative brain encoder. This encoder aims to provide learnable features that facilitate the EEG-to-text decoding process, addressing a significant challenge in the current scientific literature where the exploration of such applications remains limited. Brain signals exhibit complexity similar to wave data like speech but are recorded at much higher frequencies. Given the extensive research in the speech-to-text domain, we sought to incorporate proven techniques from this field to enhance our understanding and interpretation of brain data.

The Conformer Architecture, introduced by Anmol Gulati et al. [2], stands out as the current state-of-the-art model for speech-to-text decoding tasks. It represents a groundbreaking architecture designed to improve Automatic Speech Recognition (ASR) systems through the integration of CNNs with transformers. The Conformer seeks to combine the strengths of both approaches to more accurately capture local and global dependencies within audio sequences, effectively addressing the limitations associated with models that rely solely on either CNNs or transformers. By synergizing the parallel processing capabilities of transformers with the local feature extraction strengths of convolutional layers, the Conformer achieves an equilibrium that significantly enhances

performance in speech recognition tasks.

The architecture of the Conformer incorporates several innovative components, namely Multi-Head Self-Attention (MHSA) layers, convolutional modules, and feed-forward networks (FFNs). The authors introduce a distinctive configuration that integrates two macaron-style FFNs, featuring half-step residual connections as a structural framework enclosing the attention and convolution modules. This design choice ensures the model sustains high accuracy while optimizing parameter efficiency. Exhaustive experimentation using the LibriSpeech benchmark has demonstrated the Conformer's superior performance, achieving a WER of 2.1% on the test-clean set and 4.3% on the test-other set, even without the addition of an external language model. Remarkably, when utilized in conjunction with an external language model, these WERs improve to 1.9% and 3.9%, respectively. A notable advantage of the Conformer architecture is its effective handling of varied input lengths, rendering it well-suited for real-world ASR applications where audio sequences often exhibit substantial differences in duration. The inclusion of convolutional layers permits efficient capture of local context, while the attention mechanism ensures the retention of long-range dependencies within input data. This dual capability is crucial for accurate comprehension of speech, which is characterized by intricate patterns and variability in pronunciation, intonation, and rhythm.

Despite its impressive performance, the Conformer architecture does encounter challenges related to computational efficiency, primarily due to the inherent complexity of the attention mechanism. Such complexity has the potential to be a bottleneck during both training and inference phases, potentially restricting the model's deployment in extensive ASR systems. To mitigate this issue, ongoing research attempts focus on optimizing the Conformer model further. These initiatives include exploring pruning, quantization, and more efficient attention mechanisms aimed at reducing computational overhead without compromising accuracy.

The introduction of the Conformer model marks a substantial advancement in the realm of ASR technology. By effectively merging convolutional and transformer architecture, it sets new performance benchmarks while maintaining parameter efficiency.

The Conformer's prowess in capturing both local and global dependencies underscores its potential as a formidable tool for ASR applications across diverse domains. Continued research into refining this architecture addresses its computational challenges and promises further improvements in speech recognition systems, broadening their application in real-world settings. This work not only contributes significantly to ASR technology but also envisions future innovations in multi-modal learning and AI-driven communication systems. In the ensuing sections, we will elucidate our employment of the Conformer Architecture in executing EEG-to-text decoding, demonstrating its applicability beyond traditional speech recognition contexts.

2.3.2 Activation Functions

Activation functions are mathematical functions used in artificial neural networks to determine the output of a neuron based on its input. They play a crucial role in enabling neural networks to learn complex patterns and relationships within data by introducing non-linearities into the model. Without activation functions, the output of a neural network would be a simple linear transformation of the input, severely limiting its ability to perform tasks such as image recognition, language processing, and other complex computations. By applying an activation function, the network can effectively "decide" whether to activate a neuron based on the weighted sum of its inputs, allowing it to capture intricate patterns in the data.

There are several types of activation functions, including linear, sigmoid, hyperbolic tangent (Tanh) [115], and Rectified Linear Unit (ReLU) [116], each with its own characteristics and applications. For example, while the sigmoid function outputs values between 0 and 1, making it suitable for binary classification tasks, ReLU is often preferred in hidden layers due to its ability to mitigate issues related to vanishing gradients. The choice of activation function can significantly impact the performance and convergence of a neural network during training, making it essential for practitioners to select the appropriate function based on the specific requirements of their models and tasks.

Goyal et al. [117] highlighted the existing research gap in devising activation func-

tions for neural networks. They also addressed the pivotal role of activation functions in the performance and learning dynamics of neural networks. Activation functions are essential components that introduce non-linearity into the network, enabling it to model complex relationships within data. The authors argue that the choice of activation function significantly influences the network's ability to learn and generalize, and they propose a framework for understanding and optimizing these functions.

The authors begin by reviewing traditional activation functions such as sigmoid and hyperbolic tangent (Tanh) [115], which have been widely used in early neural network architectures. While these functions helped in introducing non-linearity, they also posed challenges, particularly with issues like vanishing gradients, which hindered the training of deep networks. The paper discusses how these limitations led to the development of newer functions like ReLU [116] and its variants, which have become popular due to their simplicity and effectiveness in mitigating vanishing gradient problems.

One of the key contributions of this paper is the introduction of a systematic approach to learning activation functions. The authors propose a method where activation functions can be treated as parameters that can be optimized during training. This paradigm shift allows for adaptive learning of activation functions based on the specific characteristics of the data and tasks at hand. By integrating this approach into existing neural network architectures, models can potentially achieve better performance by tailoring their activation mechanisms to the nuances of the input data.

The paper also explores various experimental results demonstrating how learned activation functions can outperform traditional fixed activation functions across different datasets and tasks. The authors present empirical evidence showing that their proposed method leads to improved convergence rates and overall model robustness. They emphasize that this new approach not only enhances performance but also provides deeper insights into the inner workings of neural networks, facilitating a better understanding of how different configurations affect learning dynamics.

Furthermore, the authors discuss potential applications of learned activation functions in various domains, including computer vision, NLP, and reinforcement learning. They highlight how adaptive activation functions can lead to more efficient training

processes and improved generalization capabilities in complex models.

Summing Goyal et al. presented a comprehensive framework for optimizing activation functions within neural networks. By treating these functions as learnable parameters, the authors open up new avenues for research and application in deep learning. This work not only enhances our understanding of neural network behaviour but also lays the groundwork for future innovations in model design and optimization strategies that leverage adaptive mechanisms for improved performance across diverse tasks.

To enhance the claim made by Goyal et al., Bilonoh et al. [118] proposed another category of activation functions, as they aimed to go beyond the universally used activation functions to speed up convergence and improve accuracy. Their proposed function is termed the tunable activation function, which allows for the modification of its parameters, as well as the between neuron connection weights during the training process. They presented a novel approach to enhancing the performance of deep learning models through the introduction of tunable activation functions. Traditional activation functions, such as ReLU [116] and sigmoid, are fixed and may not be optimal for all tasks or datasets. This research addresses this limitation by proposing a framework that allows activation functions to be learned and adapted during the training process, thereby improving the model's ability to capture complex patterns in data.

The authors begin by discussing the importance of activation functions in neural networks, emphasizing their role in introducing non-linearity and enabling the network to learn intricate relationships. They highlight common challenges associated with standard activation functions, including issues like the vanishing gradient problem and the "dying ReLU" phenomenon, where neurons become inactive and fail to contribute to learning. By allowing activation functions to be tunable, the proposed method aims to mitigate these issues and enhance overall model performance.

Tunable activation functions can be adjusted according to the characteristics of the input data. This adaptability allows the neural network to optimize its performance based on the specific features and patterns present in the dataset. The authors employ a parameterized approach that enables each neuron to adapt its activation function

dynamically during training. This adaptability is shown to lead to improved convergence rates and better generalization on various tasks compared to static activation functions.

To give more perspective of the differences of a tunable activation function typically the equation of a polynomial function is written as follows:

$f(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0,$

The authors proposed a different polynomial equation that replaces the constant values with adjuted parameters :

$$a_i = \begin{cases} \alpha_{i1} z_i + \alpha_{i2} z_i^2 + \alpha_{i3} z_i^3, & \text{if } z_i \geq 0 \\ \beta_{i1} z_i + \beta_{i2} z_i^2 + \beta_{i3} z_i^3, & \text{otherwise} \end{cases},$$

Through extensive experiments on benchmark datasets, the authors demonstrate that their tunable activation functions outperform traditional fixed functions in terms of accuracy and robustness. They provide empirical evidence supporting their claims, showcasing how different configurations of tunable functions can lead to superior performance across multiple deep learning architectures.

Additionally, the paper discusses practical implications for implementing tunable activation functions in real-world applications. The authors suggest that this approach can significantly enhance model flexibility, making it easier for practitioners to achieve optimal performance without extensive manual tuning of hyper-parameters.

To conclude Bilonoh et al. offered a compelling advancement in neural network design by introducing an innovative framework for adaptive activation functions. This research not only addresses existing limitations in traditional activation methods but also opens new avenues for improving deep learning models' efficiency and effectiveness across diverse applications. The findings underscore the importance of flexibility in model architecture, suggesting that future research should continue exploring adaptive mechanisms to enhance neural network capabilities.

Wang et al. [119], also highlighted the importance of different activation functions in machine learning by utilising a different approach then the 2 research investigations mentioned previously They introduced a novel approach to enhancing neural network performance through the use of polynomial activation functions. Traditional activation functions, such as ReLU and sigmoid, have limitations that can hinder the learning

capabilities of deep networks, especially in complex tasks like precipitation forecasting. The authors propose using polynomial functions of order two or higher as activation mechanisms, which can better approximate continuous real-valued functions within specific intervals. This flexibility allows the model to capture intricate non-linear relationships inherent in meteorological data.

The authors detail the mathematical formulation of polynomial activation functions and discuss their advantages over conventional methods. One significant benefit is the ability to learn non-linearities that may not be monotonic, enabling the network to adapt more effectively to the underlying data distribution. However, they acknowledge challenges such as potential exploding gradients associated with higher-order polynomials. To mitigate these issues, the authors introduce techniques like dynamic input scaling, output scaling, and a lower learning rate specifically for polynomial weights. These strategies help stabilize training and improve convergence rates.

Their research presents empirical results from experiments conducted on three public datasets related to precipitation forecasting. The findings demonstrate that networks utilizing polynomial activation functions can achieve performance levels comparable to or exceeding those of state-of-the-art activation functions. The authors emphasize that their approach allows each layer in the network to discover its preferred nonlinearity during training, enhancing the model's overall adaptability and effectiveness.

To sum up, this research contributes valuable insights into the design of activation functions in deep learning models, particularly for complex tasks such as precipitation forecasting. By leveraging polynomial activations, the authors provide a promising alternative that addresses some limitations of traditional activation functions while offering enhanced flexibility and performance in neural network architectures. This work not only opens new avenues for future research but also highlights the importance of exploring diverse activation mechanisms to improve model training and prediction accuracy in various applications.

2.3.3 Information Need and Recommender Systems

As extensively examined in Chapter 1.1, the principal aim of this comprehensive thesis is to conduct an in-depth investigation into the opportunity of the development of a sophisticated computer system that the users can navigate using a brain powered machine as the input. This exploration is grounded on the foundational framework presented in Section 3.1, which outlines two primary and extensive categories of user interaction that the system intends to address: an advanced search engine and a dynamic chatbot system.

The principal objective of both systems is to address the user's informational needs. Fundamentally, the core emphasis is placed on satisfying this inherent need for information possessed by users. A chatbot system functions as an interactive mechanism wherein users engage in dialogue to acquire answers to their inquiries by interacting with a virtual agent. This interaction directly attends to the conversational dimension of information seeking [120–122]. Conversely, a search engine is designed to fulfil a comparable role, albeit without engaging in a dialogue-oriented exchange. Instead, it delivers information through a more direct retrieval methodology to satisfy the information need [123–125]. Although these two systems differ in their modes of interaction, the essence of their functioning is grounded in addressing the user's pursuit of information. The effective delivery of pertinent information, whether through a chatbot or a search engine, is essential. These systems aim to satisfy the fundamental informational desire, an intrinsic component of user interaction. Consequently, both systems, notwithstanding their divergent methodologies, converge on the shared objective of fulfilling the information needs of users, albeit through divergent mediums and methodologies.

Within the historical framework of IR systems, notwithstanding the compelling necessity to integrate these two systems—each exemplifying state-of-the-art methodologies in information retrieval—a discernible trend emerged wherein users predominantly preferred alternative solutions for acquiring information. This tendency is unequivocally evidenced by the sustained engagement with and reliance on RSS feeds, which have been recognized as a pragmatic and effective method for accessing information [126]. With the progression of technological advancements, the development of increasingly

personalized information retrieval systems became apparent, particularly through the advent of recommender systems.

From an IR perspective, recommender systems aim to optimize user experience by pre-emptively anticipating and meeting users' information needs, subsequently providing a customized array of recommendations. The scope of recommender systems is extensive, with their applications spanning various sectors, such as movie recommendation systems [127], and systems intended for the recommendation of contextual photo tags [128]. Therefore, recommender systems signify a notable progression in the field, proficiently addressing user requirements ahead of their explicit articulation.

In order to enable these systems to fulfill their intended purposes of delivering high-quality recommendations and relevant information to users, thus addressing their informational needs, a comprehensive array of feedback data has been methodically collected over a prolonged duration [129]. It is posited that a proficient recommender system should adeptly integrate a diverse array of methodologies and various feedback inputs to provide valuable suggestions and improve the overall search experience [130]. The compiled data encompasses a broad spectrum of elements, including facial expressions [131], emotional and semantic-based characteristics [132], as well as physiological and behavioral attributes [133].

In recent developments, despite significant advancements in these technological frameworks, it has been recognized that cognitive insights can be derived from the human brain [66, 73–75, 77, 93, 134]. This crucial discovery has initiated a paradigm shift, leading researchers to move away from traditional methodologies such as pseudo-relevance feedback [135], and gravitate towards leveraging implicit neural feedback originating from the user's cognitive processes to determine relevance and enhance information retrieval in complex tasks such as passage retrieval [136]. Among the research activities focused on neural relevance are the extraction of cerebral features using video topological significance [137], the analysis of the transitional dynamics occurring between neural states during search [138], and the investigation of cognitive saturation [139]. Through this perspective, the exploration and implementation of neuropsychological feedback mechanisms have the potential to transform the understanding

of relevance and retrieval processes, symbolizing a shift from outdated paradigms to neuroscience-based approaches.

The application of implicit feedback data, sourced from brain activity, within recommender systems has been the subject of extensive scholarly investigation over the past several decades, as previously noted. In contexts such as chatbots and search engines, recommender systems aim to fulfill the information requirements of users, not only by recommending pertinent content but also by improving the quality of responses available to users' queries. Nevertheless, the incorporation of neural data into these systems has largely remained unexplored. There exists a research gap concerning the potential enhancement of interactions with these systems through the utilization of neural data. Additionally, there is substantial, unexplored potential in employing such systems to gather a more comprehensive range of implicit feedback, influenced by users' brain activity throughout their interaction with the systems.

2.4 Brain-To-Text Decoding

In recent years, there has been a significant body of research aimed at understanding the decoding of language within the human brain, specifically focusing on how language can be processed when individuals either read text or listen to spoken language. This domain has been explored in adequately by a range of studies but still not a definite implementation has been made available to openly decode human thoughts and inner speech.

In their pioneering study, one of the initial forays into the domain of brain-to-text decoding was conducted by [1]. Their seminal research, titled 'Interpreting and Improving Natural-Language Processing (in Machines) with NLP (in the Brain)', meticulously investigates the nexus of cognitive neuroscience and the burgeoning field of AI. The research underscores a particular emphasis on the transformative insights that can be gleaned from human brain functions related to language processing, with the express intent of using these insights to innovate and refine machine learning algorithms in NLP.

The distinguished authors postulate that deciphering the neural substrates and

mechanisms pivotal to human language comprehension holds the potential to furnish AI researchers with robust frameworks that can inform and elevate the efficacy of algorithms currently deployed in NLP. The investigative endeavour embarks by presenting a comprehensive overview of the present landscape of deep learning architectures, which have notably championed significant advancements and efficacy in a spectrum of NLP tasks encompassing, but not limited to, text generation, automatic translation, and nuanced sentiment analysis.

Despite these advancements, it is acknowledged that such models often navigate challenges related to interpretability and do not yet authentically emulate the intricate and nuanced modes of human linguistic understanding. To approach this limitation, the authors propose a methodical juxtaposition of cerebral activity data elicited during linguistic tasks and the output of various state-of-the-art NLP models. The methodological framework utilized incorporates advanced neuroimaging techniques, specifically fMRI alongside MEG, to map and record the neurophysiological responses of participants as they cognitively engage with language stimuli. The accrued empirical data was rigorously analysed to pinpoint distinct neural circuits engaged at various stages of linguistic processing.

The results illuminate a fascinating finding: certain advanced deep learning models, especially those employing transformer-based architectures, demonstrate neural activation patterns bearing a significant resemblance to those observed in the human brain during analogous language processing tasks. Two pivotal insights emerge from this comparative analysis: foremost is the assertion that the degree of congruence between artificial and cerebral representations is predominantly influenced by the models' proficiency in contextual word prediction. This infers that enriching contextual paradigms within NLP algorithms could stimulate more brain-analogous processing schemas. Furthermore, the research delineates specific cortical zones instrumental for processing perceptual, lexical, and compositional language constructs, thereby charting a path for the creation of refined NLP systems that intricately simulate human cognitive pathways.

Moreover, the treatise deliberates on tangible implications for the enhancement of NLP algorithms through the assimilation of cognitive science-derived principles. Mod-

els architected to mirror hierarchical and contextual linguistics akin to human cerebral activities promise to be not only more efficient but also possess enhanced interpretability.

The authors ambitiously recommend that ensuing research pursuits should be oriented towards fine-tuning these models to augment their analytical prowess across diverse linguistic tasks, all the while ensuring resonance with human cognitive paradigms. In summation, this comprehensive paper presents a profoundly persuasive argument advocating for the strategic application of cognitive neuroscience insights to propel advancements in NLP technologies. By marrying machine learning innovations with human cognitive processing frameworks, scholars can architect NLP systems that are not only more robust and interpretable but also more accurately mirror the rich complexities inherent in human communication. Such a cross-disciplinary approach extends its value beyond mere machine capability enhancement, enriching the broader understanding of language as an essential element of cognition, thus fortifying the bridge between AI and human linguistic faculties.

In the expansive and rapidly evolving field of machine learning, it is a widely accepted fact that the quality and accessibility of data serve as foundational elements for the conception and refinement of models that are both robust and accurate. The process of gathering and utilizing fMRI data is notably expensive and inherently lacks real-time capabilities, which poses certain limitations for its application in timely decision-making processes.

In contrast, EEG data, particularly within the interdisciplinary domains of cognitive neuroscience and Neuroinformatics, offers real-time data collection capabilities and is relatively more economical to acquire. Despite these advantages, EEG data brings forth unique challenges attributable to its inherently complex signal structure, as well as the resource-intensive nature of its acquisition and the sophisticated analysis required for meaningful interpretation .

The Zurich Cognitive Language Processing Corpus (ZuCo) [140, 141] represents a pivotal contribution to this field by amalgamating high-density EEG recordings with eye-tracking data for an enriched understanding of cognitive language processing. The

ZuCo corpus is meticulously curated to elucidate the intricate neural and behavioural dynamics involved in language comprehension, thereby offering an invaluable asset for researchers at the intersection of NLP and cognitive neuroscience.

Specifically, the dataset encompasses comprehensive recordings from 12 neurotypical English-speaking adults engaged in an array of reading tasks, with a cumulative duration of 4 to 6 hours per participant. These tasks yielded a substantial corpus of data, encompassing 21,629 discrete words embedded in 1,107 sentences, alongside a total of 154,173 eye-tracking fixations. This extensive dataset not only permits the in-depth analysis of reading behaviour but also provides a granular view of cognitive load and processing strategies employed during naturalistic sentence comprehension. ZuCo is meticulously partitioned into structured tasks, including two paradigms of typical reading alongside a specialized task-focused reading activity. This pertains to the systematic exploration of diverse facets of cognitive and linguistic processing. The symbiotic integration of EEG data, representing neurophysiological activities, with eye-tracking metrics, illustrating oculomotor behaviour and visual engagement, amplifies the potential to decipher the reciprocal interactions between cognitive mechanisms and linguistic constructs.

Advancing research in NLP methodologies, the authors of ZuCo underscore its profound utility in enhancing machine learning paradigms for sophisticated applications such as entity recognition, relational extraction, and sentiment analysis. The dataset's EEG and eye-tracking components provide essential inputs for the refinement of algorithms aimed at emulating human linguistic processing capabilities with higher fidelity. Moreover, the corpus extends its significance beyond machine learning applications by serving as a foundational dataset for probing core inquiries into human reading processes and language cognition. By elucidating the confluence of brain activity patterns and ocular dynamics during text interpretation, ZuCo offers critical insights into the cognitive architecture driving language comprehension. This dual data synthesis not only enriches theoretical frameworks but also has potential translational implications in developing cognitive and linguistic interventions for individuals with reading and language impairments.

The ZuCo has introduced an advanced iteration, designated as ZuCo-2.0, subsequently producing an enriched dataset. In comparison to its predecessor, ZuCo-1.0, the ZuCo-2.0 dataset incorporates comprehensive psycholinguistic data from a cohort of eighteen subjects. This expanded dataset embodies a substantial linguistic resource, presenting 15,138 distinct words articulated across 739 exemplified sentences [141]. The meticulous construction of these datasets involved the integration of sophisticated eye-tracking metrics, as methodically detailed within the GECO corpus [142], thereby offering a dual-layered analysis encompassing both lexical and sentential dimensions.

Notably, the dataset leverages critical eye-tracking features, such as 'first fixation duration,' representing the initial temporal period during which participants focus their visual attention on a specific word, in conjunction with 'total reading time,' which aggregates the entirety of fixation periods observed on the target word. These metrics not only provide nuanced insights into cognitive processing but also serve as a fundamental basis for advancing psycholinguistic research.

In summary, the ZuCo epitomizes a landmark advancement in the empirical study of cognitive language processing through its systemic integration of EEG and eye-tracking modalities. This dataset provides a pivotal resource for the dual exploration of neural mechanisms and behavioural paradigms underpinning reading and comprehension, furthering both theoretical insights and practical applications in cognitive neuroscience and NLP.

The utilization of the Zurich Cognitive Language Processing Dataset, commonly referred to as ZuCo, represents a groundbreaking development in the realm of neuroscience-assisted NLP. One of the first large-scale applications, spearheaded by Hollestein et al. [143], sought to investigate the utility of EEG data in enhancing the capabilities of NLP tasks. Historically, human behavioural data, such as eye-tracking from reading tasks, predominantly served to elucidate underlying cognitive processes. However, this pioneering research diverges by exploring the deployment of neural signals, inherent to language processing, to augment machine learning models within the NLP domain.

At the heart of the study is an innovative multi-modal machine learning architecture that harmonizes both textual data and EEG features. This integrative approach is de-

signed to pinpoint and analyse EEG signals' elements that predominantly bolster NLP performance. A pivotal discovery by the authors is the superior efficacy of EEG signals partitioned into discrete frequency bands over general broadband signals. This finding posits that particular frequency domains within EEG data encapsulate more pertinent information essential for comprehending language processing mechanisms compared to others.

The research rigorously examines two principal NLP tasks, namely sentiment classification and relation detection. In the domain of sentiment analysis, involving both binary and ternary classification tasks, the incorporation of EEG-derived data significantly amplifies performance metrics compared to baseline models devoid of neural inputs. Conversely, the intricacy of relation detection tasks necessitates reliance on advanced word embeddings. The specific superiority of contextualized embeddings, with BERT proving particularly beneficial, underlines the complexities inherent in integrating neurophysiological features in sophisticated NLP tasks. This calls for a continued inquiry into optimizing EEG data application in more complex NLP functionalities.

Complementary modalities, such as eye-tracking, are juxtaposed with EEG data to provide a holistic assessment of cognitive load experienced during language processing. Whereas eye-tracking furnishes an indirect gauge of cognitive exertion, EEG can supply immediate insights into neurological activities allied with linguistic comprehension. This differentiation accentuates the merit of synergistically employing these methodologies to attain a comprehensive apprehension of human language processing.

The authors elucidate their methodical framework, entailing an exhaustive evaluation of diverse EEG features and their consequential impacts on selected NLP tasks. By examining varying neural network models adept at concurrently processing both textual and cerebral data streams, the investigation champions the importance of customizing feature extraction to cater to EEG's distinct signal properties.

In summary, the discussed investigation epitomizes a pivotal advancement in the integration of neurophysiological data with NLP architectures. Through demonstrating EEG's augmentative role in machine learning approaches oriented towards language processing, the work not only unveils newer trajectories for developing more nuanced

and interpretable NLP systems but also sets a precedent for future inquiry. This trajectory encompasses probing the intricate relationship between neural activities and language understanding, heralding advancements in assistive communication technologies for individuals with linguistic impediments, and expanding our comprehension of human cognitive processes.

Another EEG-to-Text research was conducted by Herff et al. [144] epitomizes a pioneering advancement in the domain of BCIs by presenting the unprecedented ability to decode continuous spoken language directly from neural signals acquired through intracranial Electrocorticography (ECoG). This research addresses a formidable challenge in the field: the translation of neural activity into coherent and fluid text, thereby bridging the communicative gap between human cognitive processes and computational devices. By building upon previous efforts that have been limited to recognizing isolated components of speech such as phonemes or discrete words, the authors achieve a breakthrough that extends these capabilities to decode entire phrases with remarkable proficiency.

The innovative Brain-To-Text system rooted in this study adopts a refined model for the integration of ASR techniques with neural signal analysis, focusing initially on the sub-processes related to single phoneme recognition. Herff et al. [144] successfully transition from this micro-level analysis to a macro-level demonstration of decoding continuously spoken phrases, a feat previously unattainable due to the intricate nature of speech-related neural encoding. By addressing this complexity, the research significantly advances our technical and theoretical understanding of speech signal processing from a neural perspective.

The empirical results underscore the efficacy of the Brain-To-Text system, which achieves commendable WERs as minimal as 25% and phone error rates under 50%. These figures represent a substantial stride in the accurate decoding of continuous speech from neural activity, indicating potential practical applications. In elucidating the cortical regions with rich neural correlates to phonemic information, the study enhances our comprehension of the spatial and functional mapping of speech production within the brain, thereby contributing to the broader corpus of neurocognitive

linguistics.

Methodologically, the research framework involved the development of neural phone models that were rigorously trained on ECoG data obtained during specific speech tasks. Through meticulous analysis, the researchers were able to draw a nexus between distinct neural firing patterns and their corresponding phonetic and linguistic outputs. This methodological innovation not only accentuates the translational potential of BCIs but also demystifies the underlying neural mechanisms of speech, providing seminal insights that fuse neuroscience with advanced computational linguistics.

The broader implications of this research are transformative, heralding a new epoch in the development of assistive technologies prioritizing human-machine communication grounded in the neural representation of imagined speech. By potentially revolutionizing the modalities through which individuals with speech impairments or neurological conditions interact with external environments, this research holds the promise for substantially improving life quality through improved intuitive and personalized assistive solutions.

In reflection, the Brain-To-Text system signifies an epochal advancement in the intersection of cognitive neuroscience and machine intelligence. By successfully decoding neural signals into spoken phrases, this body of work opens expansive avenues for realizing sophisticated BCIs that seamlessly convert natural human thought into machine-readable language. This pursuit not merely heightens current technological capabilities but significantly enriches our scientific understanding of human cognitive processes related to speech production, thereby contributing to an interdisciplinary dialogue between neuroscience, AI, and linguistics.

Furthermore in the domain of neuroscience and AI the last study that touches the domain of EEG-to-Text decoding, is the cutting-edge study by Wang et al. [145] breaks new ground in the realm of EEG-to-text decoding systems. This research presents an innovative paradigm for translating spoken language from EEG signals while concurrently processing sentiment classification, all without the burdensome necessity of substantial annotated training datasets. The overarching goal of this research is to advance communication technologies, particularly for individuals experiencing speech

impairments, by directly converting neural activity into textual output while capturing the latent emotional undertones of such communications.

The research commences by elaborating on the constraints inherent in conventional EEG-based methodologies for speech decoding, which typically depend upon a narrow lexicon and extensive datasets for training. The researchers introduce an adaptive open-vocabulary framework that transcends the limitations of a predefined vocabulary by facilitating the decoding of any spoken phrase. This adaptability is realized through sophisticated machine learning paradigms that harness the capabilities of expansive language models designed to comprehend and generate text based on nuanced contextual signals.

To operationalize this open-vocabulary paradigm, the authors employ a meticulously structured two-phase process: Initially, they extract critical features from EEG signals that correlate with brain activities indicative of speech articulation; subsequently, these extracted features are integrated with a pre-trained linguistic model to yield textual representations. This bifurcated approach not only augments the fidelity of decoding but also bolsters the system's capacity to adapt to diverse speech idiosyncrasies and situational contexts.

Beyond merely facilitating speech decoding, the study ventures into the realm of zero-shot sentiment classification. This enables the model to infer the emotive tone of the decoded text, effectively classifying sentiments as positive, negative, or neutral, without necessitating explicit training on predefined sentiment labels. Such a capability is invaluable, particularly in contexts where labelled sentiment data is sparse or entirely absent. The authors empirically demonstrate how their methodology effectively utilizes the inherent contextual nuances of decrypted text to perform sentiment classification.

The empirical results delineated within the study reveal that the proposed framework substantiates marked improvements in decoding precision and sentiment classification efficacy relative to pre-existing methodologies. The facility to decipher phrases with an open vocabulary from neural signals signifies a pivotal evolution in BCI technologies, heralding a potential shift towards more intuitive human-machine interactions.

Moreover, the implications of this research resonate beyond assistive communica-

tion technologies; they imply burgeoning applications across disciplines such as mental health surveillance, where decoding an individual's emotional state via neural activity could yield profound insights into their psychological welfare.

Ultimately, this seminal paper illustrates a forward-looking trajectory in neurotechnology by synergizing open-vocabulary EEG decoding with advanced sentiment classification techniques sans training datasets. The developments elucidated not only expand the functional repertoire of BCIs but also deepen our understanding of the neural encoding of language and emotion. This cross-disciplinary strategy paves the way for novel research inquiries and practical implementations spanning both the neuroscientific and AI fronts.

Extensive initiatives have been undertaken in the domain of EEG-to-text and brain-to-text decoding, each confronting intrinsic limitations reflective of the complexity inherent in neural signal translation. Our research initiative is poised to significantly elevate the operational effectiveness of the leading EEG-to-text conversion framework by strategically revising its architectural design. This endeavour entails a shift from the traditional reliance on extracted EEG features towards a profound emphasis on the utilization of unprocessed, or raw, EEG signals. Empirical evidence underscores the superiority of such raw data in enhancing classification performance, particularly when merged with the state-of-the-art deep learning mechanisms or advanced transformer-based technologies [146,147]. This empirical foundation not only confirms the efficacy of raw signals but also mitigates the exigencies associated with exhaustive preprocessing.

Preprocessing of EEG data typically demands substantial computational resources and significant manual effort, a burden substantially alleviated through the adoption of unrefined data. By circumventing the preprocessing phase, computational overheads are appreciably reduced, aligning with the overarching objective of creating a more efficient and streamlined workflow for EEG data analysis. Thus, our approach is not merely revolutionary in the context of data processing but also in resource management, showcasing a shift towards sustainability in computational practices. Through the strategic alignment of raw EEG data with pioneering advancements in the realms of neuroscience and AI, specifically machine learning paradigms, we envisage approaching

the conceptualization of a functional, real-time brain decoding apparatus. This ambition carries profound implications for the future of neurotechnology, suggesting the potential to bridge the existing gap between abstract neural signal interpretation and tangible, real-world applications.

Chapter 3

Brain-Computer Interface Neural Information System

3.1 Introduction

The principal objective of this thesis is to present a comprehensive architectural framework enabling user interaction exclusively through BCI technologies. This interaction paradigm obviates the need for traditional input mechanisms by utilizing advanced neuroscientific methodologies for direct command.This section offers a thorough examination and systematic breakdown of each key component required for the implementation of a Brain-Computer Interface Neural Information System (BCI-NIS). Each component is dissected independently, elucidating their distinct contributions and roles within the comprehensive system architecture. In addition, a thorough examination of the sub-modules enabling the functionality of each component is provided, explaining the sophisticated frameworks that support their operations. Particular emphasis is placed on explicating the justification for incorporating each component BCI-NIS, emphasizing their fundamental operational responsibilities and projected outcomes. Additionally, this chapter offers an assessment of potential constraints and technical challenges that may arise during the development and integration of these components and their linked sub-modules. By presenting these elements, this analysis highlights the complexities involved and outlines the tactical approaches to address issues faced

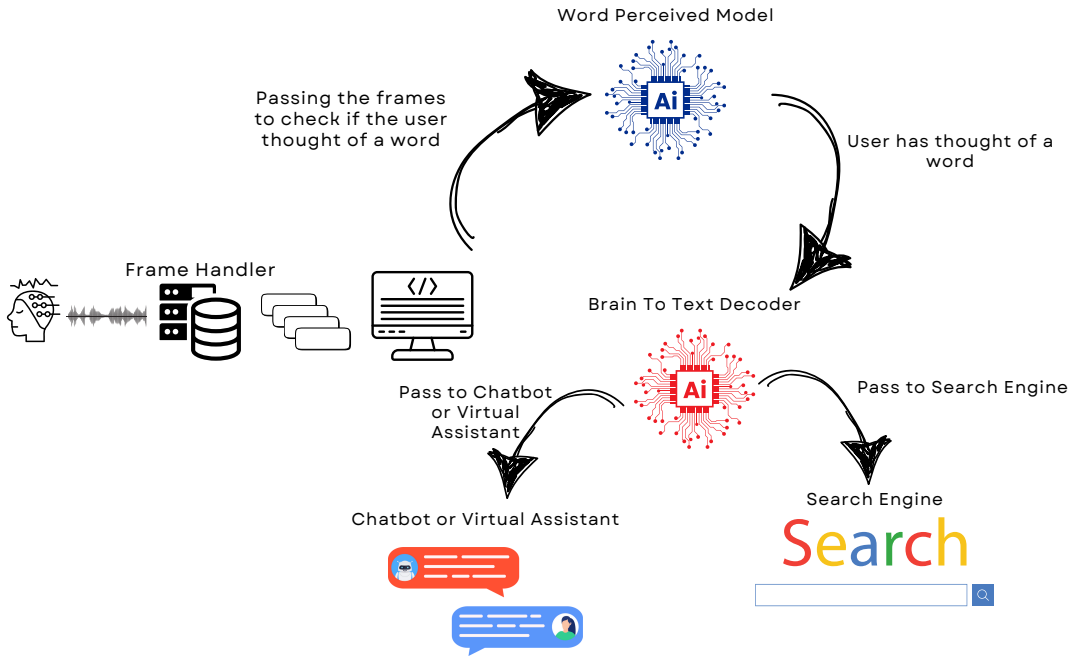


Figure 3.1: The proposed BCI-NIS architecture is outlined in this Figure, comprising several key components. Initially, the Brain Decoder captures real-time brain data from the user. This data is managed by the Frame Handler, which stores and preprocesses the most recent frames securely. The frames are then analyzed by the first AI layer to ascertain if they form a user-conceived word. Finally, the core component decodes the frames to generate text, which can be sent to a virtual assistant or a search engine.

in implementing a fully integrated brain-interactive system, thereby paving the way for future advancements and enhancements in the field of brain-computer interfacing. An accompanying diagram is presented in Figure 3.1 where you can a see a complete diagram with all the high level components and how they interact with each other and a description of the flow of the system.

3.2 Brain recorder

The *Brain Recorder* serves as the foundational and pivotal element of BCI-NIS. Its primary function is to provide a crucial interface for the incorporation of neural data into

the comprehensive BCI-NIS system architecture. It is of utmost importance to maintain the BCI-NIS system's portability; thus, non-invasive brain imaging techniques are identified as optimal solutions. These techniques facilitate the efficient transfer of neural data, thereby enhancing the BCI-NIS system's portability and practicality.The implementation of an integrated brain recorder within the system is significant because, in the field of neuroscience, it is uncommon to find or implement modules that are capable of programmatically recording EEG data and facilitating its transfer to other components of a computer system. More frequently, external software, such as BrainVision, is utilized to record neural activity, with the recorded files later being analyzed through computational methodologies.

An efficacious strategy for the Brain Recorder is to focus on the acquisition and documentation of EEG data from the user. When compared to other traditional brain imaging modalities, EEG technology provides unmatched benefits concerning portability. The intrinsic properties of EEG enable it to surpass the constraints posed by more cumbersome and less mobile alternatives, thus aligning seamlessly with the objectives of BCI-NIS system portability.

Moreover, EEG data offers unparalleled real-time surveillance of neural activities, a characteristic that is critical to the BCI-NIS system's efficacy and responsiveness. This capability of real-time data capture guarantees that the BCI-NIS system obtains prompt and precise information regarding brain activities, thereby augmenting its capacity to process and react expeditiously.

Upon successful acquisition of EEG-based neural data, it is subsequently directed to the ensuing BCI-NIS system component, referred to as the Frame Handler. The Frame Handler meticulously processes the input data, ensuring its seamless integration into the BCI-NIS system, thereby enabling further analytical procedures or actions predicated on the recorded neural activations.

Having presented the functionality of the Brain Recorder now the difficulties and challenges of constructing this component are presented.

The primary challenge delineated in this research pertains to the extant difficulties associated with the portability of BCI-NIS systems that utilize EEG data in practical

applications. Although EEG technology is inherently non-invasive, an array of advanced EEG devices has been identified as notably burdensome to convey due to their inherent complexity. Furthermore, these BCI-NIS systems often necessitate supplementary external apparatus, such as amplifiers, for their effective operation, thereby magnifying the challenge of portability. To address this challenge, insights may be drawn from the seminal work by Kingphai and Moshfeghi, as cited in [148]. Their study, elaborated in Section 2.2.5, reveals the potential for training a deep learning model in an offline context with data from a highly channel-rich EEG apparatus. This model, once trained, can be fine-tuned or employed for inference with data from sources possessing significantly fewer channels, essentially comprising subsets of the original comprehensive electrode data.

Although this specific methodology has not yet been empirically implemented in BCI-NIS, it is underpinned by a robust theoretical framework that suggests its potential effectiveness. This strategy could potentially enable the development of a compact, portable EEG system, utilizing, for instance, merely four EEG channels as the input for the device, while leveraging backend models trained on more extensive datasets. The application of such a solution has the potential to significantly enhance the portability and usability of EEG systems across a variety of practical contexts, marking a considerable advancement in the field.

3.3 Frame Handler

In the section designated as *Frame Handler*, the discourse addresses a pivotal element of the BCI-NIS system architecture tasked with the management and processing of sequential data frames. The role of the Frame Handler transcends mere data organization; it is integral to the BCI-NIS system's capability to manage real-time data processing requirements efficiently. As data frames are received in succession, the Frame Handler expeditiously arranges them into a coherent structure, thereby enabling subsequent modules to process the information both effectively and accurately.

Fundamentally, the Frame Handler serves not simply as a supplementary component of the BCI-NIS system but as a critical element that facilitates the maintenance of

high performance and reliability. Its function in coordinating the flow of data frames constitutes the core of the BCI-NIS system's data management strategy, underscoring its essential contribution to the overall operation and efficiency of the BCI-NIS system architecture.

A key difficulty encountered by the Frame Handler lies in the careful maintenance of a complete history encompassing all recordings obtained from the Brain Recorder during one session. This task is critically important because the AI models integral to the BCI-NIS system require an extensive dataset of recordings, rather than just one, to perform efficient inference procedures. Various studies have conclusively demonstrated that utilizing an ensemble of recordings, as opposed to an individual recording, substantially enhances outcomes [149, 150]. This approach primarily facilitates significant noise reduction in neural data. Given that EEG has been selected as the brain imaging technique for this BCI-NIS system, a solitary EEG recording could be potentially encumbered with noise. Conversely, a collection of recordings not only provides augmented variability for the model but also ensures considerable noise reduction [151].

Additionally, a strong machine learning model is often defined by its generalization skill. Introducing diverse recordings greatly improves the generalization aspect of the data fed into the model [152]. In conclusion, the aggregation of several EEG recordings addresses the challenge of temporal stability. EEG data inherently exhibit temporal fluctuations in brain activity patterns. Capturing these changes over time, multiple recordings yield a more stable depiction of brain functions. This comprehensive depiction embodies both the transient and persistent attributes of neural activity [153].

Given the stipulated requirements for the Frame Handler, a notable challenge presents itself. Taking into consideration that modern EEG headsets can achieve a sampling rate between approximately 256Hz and 500Hz, preliminary computations indicate that this would result in approximately 15000 to 30000 recordings per minute. This represents a considerable quantity of recordings, thereby necessitating the management of a substantial volume of data by the Frame Handler.

Hence, an additional compelling justification arises for the implementation of a specialized component interposed between the Artificial Intelligence models and the Brain

Recorder. This intermediary is designed to proficiently manage the incoming data stream. An effective solution entails the deployment of cutting-edge big data management libraries alongside sophisticated algorithms. These technologies would proficiently facilitate data retrieval while preserving the sequential integrity of the recordings. Consequently, this strategic measure augments the BCI-NIS system's overall capacity to process large datasets with accuracy and efficiency.

3.4 Word Perceived Model

In this crucial segment of the chapter, the detailed mechanisms of the *Word Perceived Model* are thoroughly explicated. This element is crucial to the overall success of the BCI-NIS system due to its indispensable capability to accurately identify instances when a particular word or phrase has been authentically conceived by the user. The implementation of this module is of significant importance due to the continuous activity of the human brain, which invariably generates various forms of activation, such as blinking. The Word Perceived Model provides a crucial mechanism for distinguishing those moments when the system user genuinely conceives words intended for inclusion and engagement with the system, from those periods of idleness or physical movement. This functionality is imperative, as it mitigates the introduction of erroneous samples of misclassified 'words' into the system.

The BCI-NIS system achieves this by distinguishing intentional mental activities from less pertinent states, such as when the user is in a passive resting condition or engaged in unrelated tasks. This differentiation is critical as it enables the BCI-NIS system to accurately map and respond to the user's dynamic cognitive processes without erroneously interpreting unrelated neural signals that may occur during periods of inactivity or distractions. Consequently, the success of this model is contingent upon its proficiency in identifying and isolating genuine word perception among a multitude of potential mental states, a characteristic that substantially enhances the efficacy and reliability of the broader BCI-NIS system within which it functions. This is a particularly important functionality as it will limit the times the Brain-To-Text Decoder is going to be used, since our brain is producing activity 24/7 but not all of those

activations indicate thinking of a particular word or correspond to a text.

To develop this capability and ensure the efficacy of a Word Perceived Model, numerous challenges must be confronted. A primary challenge lies in the effective training of this model. During the initial design and research phase, a suitable dataset that could supply a proper training regimen to achieve this functionality was not identified. It is crucial to document as a potential enhancement of this thesis that there may be a need to curate a dataset endowed with the essential characteristics to train such a model, or alternatively, to identify one that becomes available subsequent to the authorship of this chapter. Addressing these aspects will be imperative for the advancement and refinement of the model's development process.

In the domain of advancing a Word Perceived Model, a considerable obstacle is encountered during the model's training phase. Despite the extensive body of literature demonstrating the efficacy of machine learning models in predicting mental states, this particular investigation was unable to identify a pre-existing model that meets the specific criteria outlined for this module. While one might initially perceive this task as being relatively straightforward, our analysis suggests that training such an advanced model would inevitably reveal further, as yet unexplored, challenges. Consequently, embarking upon the training of this model represents a significant and unresolved challenge, highlighting the complexity and novelty inherent to this pursuit. This underscores the essential need to address various uncharted challenges associated with model training, representing hurdles yet to be overcome in the quest for a truly effective Word Perceived Model.

3.5 Brain To Text Decoder

The final component within the proposed BCI-NIS system architecture is the Brain-To-Text Decoder, which plays an indispensable role in translating neural signals into textual format. The user interaction with the computer system is facilitated through either a search engine or a chatbot component. Both of these modalities primarily require textual information as input. Successful interaction necessitates the translation of user thoughts or neural signals into text. This translation process facilitates interaction

through textual output with virtual assistants or chatbots, in addition to various search engine BCI-NIS systems. The advancement of this decoder constitutes a critical aspect of this research, enabling the development of comprehensive interaction models with broad applicability across digital platforms. By decoding brain waves, this approach bridges the communication divide between human cognitive processes and digital textual interaction BCI-NIS systems, thereby enhancing usability and functionality.

This research strives to concentrate substantially on the Brain-To-Text Decoder, attributed as the most technically complex and innovative element of the architecture. The decoder adeptly interprets intricate patterns of brain activity and converts them into coherent text, subsequently enabling functionality across a wide array of applications, including virtual conversational agents and sophisticated online search mechanisms. Through meticulous analysis, an enhanced understanding of the submodules comprising this component can be attained, utilizing insights from our preliminary research findings and theoretical projections. This component of the thesis necessitates comprehensive analytical exploration, positioning itself as a vanguard in brain-wave interpretation and laying the groundwork for future progress in interactive AI development.

Moreover, the Brain-To-Text Decoder exemplifies the convergence of neuroscience and artificial intelligence, representing an area promising extensive exploration and advancement. By deciphering and converting neural activities into textual data, this component holds the potential to revolutionize interactions with digital BCI-NIS systems, enhancing their intuitiveness and integration with human cognitive processes. Consequently, it not only presents significant opportunities for further academic inquiry but also offers practical applications that could fundamentally alter how users engage with technology. Based on our preliminary investigation, we establish a foundational comprehension of the submodules involved, which will inform the subsequent phases of this research, ensuring that the Brain-To-Text Decoder is comprehensively examined and developed to its utmost potential.

As illustrated in Figure 3.2, the overarching framework is systematically categorized into three principal subcomponents. Among these, the foremost is the Brain

Encoder. The Brain Encoder's primary function is to develop a versatile deep learning architecture by employing advanced methodologies such as Wav2Vec2 or Data2Vec. As described in Section 2.3.1, these methodologies have recently demonstrated an exceptional capacity to formulate models that generate embeddings applicable to multiple modalities. These embeddings provide a comprehensive feature set for each item across the modalities. The significance of adopting a model that internalizes universal brain embeddings lies in its capability to render the entire Brain-to-Text decoder architecture autonomous from the specifics of the EEG cap, such as the sampling rate or the number of channels currently utilized by the user. Consequently, this results in the provision of information to the model that is both generic and task-agnostic, thus enhancing its versatility and adaptability across various applications.

The second primary subcomponent within the architectural framework is designated as the Brain Feature Decoder. This component is pivotal in acquiring and interpreting representations derived from the generic features managed by the Brain Encoder, subsequently translating them into natural language text. During the initial phase of research, it was ascertained that the mechanisms and processes pertinent to speech-to-text conversion exhibit significant commonalities with tasks associated with brain-to-text translation. Consequently, critical modules have been identified as vital for the successful construction and operation of this subcomponent. Among these critical modules are the implementations of LogSoftmax and Connectionist Temporal Classification (CTC) Loss. These modules are employed to efficiently facilitate the learning and translation of representations from the Brain Encoder into coherent natural language text, thereby ensuring a robust and accurate decoding process within the BCI-NIS system.

In the final stage of this research, we present the ultimate component, which involves the integration of LLMs as a fundamental subcomponent. This integration is crucial for the re-evaluation and refinement of the textual output produced by the Brain Feature Decoder. Similar to speech-to-text technology, where homophones may create ambiguity despite their phonetic resemblance, such words can assume different grammatical functions depending on the contextual framework in which they are employed. The implementation of LLMs plays a crucial role as a protective mechanism,

re-assessing the generated text to ensure nuanced contextual interpretations that yield a sequence precisely aligned with the user's original intent. Through this meticulous process, LLMs significantly enhance the reliability and contextual fidelity of the text, thus bridging the gap between intended cognitive expressions and their concrete textual representation the same way this is done in speech-to-text task [103, 154, 155].

The pursuit of developing the current module exemplifies a trailblazing approach within this domain, accompanied by a plethora of anticipated and unanticipated challenges. Our preliminary research efforts have effectively highlighted several critical obstacles that necessitate resolution. As we advanced through the implementation phase, integrating our three principal tasks, a comprehensive array of challenges was discerned. These challenges were meticulously documented, and strategic measures were employed to surmount them. The project's novelty mandates an adaptive and innovative problem-solving approach, as each developmental stage unveiled distinct difficulties requiring targeted solutions. This iterative process of challenge identification and resolution is crucial to enhancing our understanding and optimizing the module's efficacy.

The preliminary domain presenting potential challenges pertains to the training data associated with this model. In the subsequent sections, it becomes evident that a fundamental issue highlighted is the variability in sequence lengths observed across different EEG recordings. The recording of each individual sentence within the training dataset exhibits variability in the number of recordings associated with that particular sentence. Furthermore, it is essential to acknowledge that the volume of data exceeds that of traditional speech-to-text datasets, merely for comparative purposes. Consequently, this indicates that the architecture of conventional speech-to-text models may face difficulties when attempting to accommodate both the extensive data and the model itself within a single machine. This could result in complications, as the computational resources required could surpass the capabilities of a solitary machine, thereby necessitating alternative solutions or adaptations of existing models to effectively and efficiently manage the increased complexity and size of the data.

Subsequent to the initial investigation, a significant challenge of this research is

the construction of a robust model capable of generating text from brain data. Contrasting with other issues in Neuroscience where machine learning techniques have been effectively utilized, this particular problem is characterized by the absence of a well-established framework or guidelines. At the time of this thesis composition, there existed no precedence or documented methodologies in the Neuroscience literature that clearly delineated successful strategies for translating brain signals into coherent text. Recognizing this shortfall, the research community initiated a collaborative effort to identify machine learning approaches from other disciplines that possess similar data structures or requirements. They collectively examined the viability of adapting these approaches for application to brain data. The intricate nature of brain function contributes to the complexity of brain data, which is frequently affected by noise or does not possess real-time attributes, thus making analysis more challenging. This intricacy necessitates more bespoke approaches, as conventional, out-of-the-box solutions frequently prove insufficient or ineffective. Consequently, these conventional models necessitate significant modifications to fulfil the demands imposed by brain data complexities. Hence, researchers are compelled to meticulously refine and tailor extant machine learning techniques to accurately and effectively interpret neural signals within the inherent constraints and limitations of brain data.

Finally, a principal challenge in EEG research is the considerable difficulty in achieving uniformity in the configuration of EEG datasets, even when the datasets are successfully acquired. Each dataset is frequently characterized by unique parameters and settings, which are largely dictated by the specific experimental design and the technical specifications of the equipment employed during data acquisition. This lack of uniformity presents a significant obstacle to the standardization of EEG datasets. Moreover, this variation is exacerbated when attempting to establish a common protocol for data collection across various EEG machines. The absence of standardized procedures, coupled with the diversity in equipment specifications, renders the attainment of consistency in data acquisition protocols exceedingly challenging. Consequently, significant variability is introduced into the datasets, which impedes the generalization of algorithms, such as the Brain-To-Text decoder, across diverse studies or experimen-

tal settings. This variability severely impacts the feasibility of developing a versatile brain-powered BCI-NIS system. Without a common protocol and standardized configuration, the successful realization of such BCI-NIS, applicable to a wide array of EEG data sources, becomes increasingly improbable.

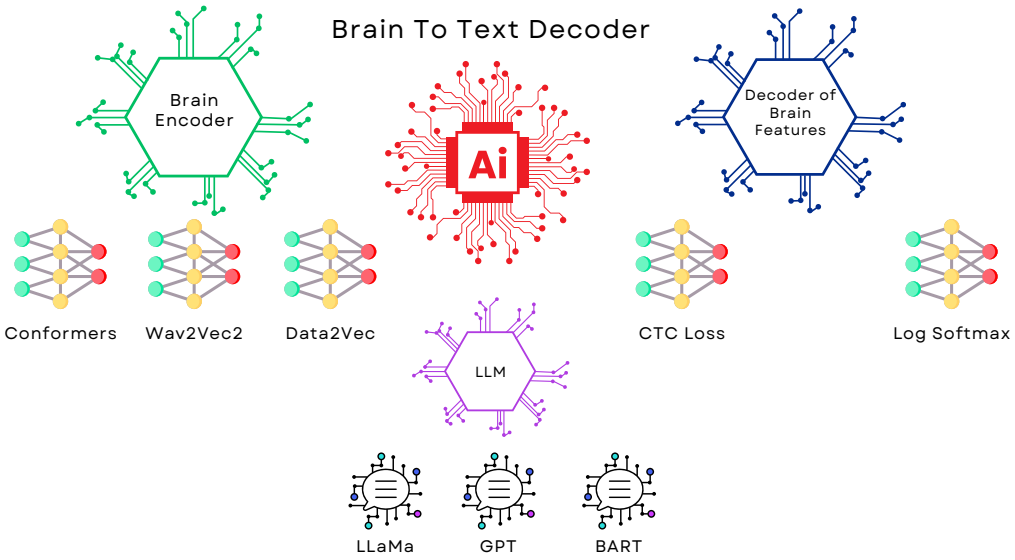


Figure 3.2: A high level overview of the 3 main components that are used to construct the Brain-To-Text Decoder.

3.6 Rationale for choosing the focus area

This thesis aims to conduct a thorough analysis of all submodules, principally aimed at offering deeper understanding into the development of the Brain-To-Text decoder component. The rationale for focusing on this element arises from multiple considerations. Primarily, the difficulties faced in the Brain Recorder and Frame Handler modules are intrinsically associated with software-related challenges. These challenges can generally be addressed by investigating and implementing well-established methodologies from

existing literature, thus providing solutions to these distinct issues.

Although these modules are indispensable for the comprehensive operation of the entire BCI-NIS system, they do not constitute the core of the model. Rather, the central functionalities reside within the AI systems. Consequently, emphasis on the advancement of these AI systems is of utmost importance, as their development is anticipated to reveal a plethora of sub-challenges. Moreover, the evolution of AI systems in this framework is not only critical but is also projected to catalyse significant progress and address emerging challenges throughout the course of this research initiative.

The selection of this specific component was predominantly influenced by the presence of multiple datasets that amalgamate the analysis of human text reading with EEG data. The utilization of these pre-existing datasets facilitates the research process by obviating the necessity to develop a novel dataset, thereby enabling a more streamlined and efficient research methodology. Notably, existing datasets such as ZuCo-1.0 and ZuCo-2.0 were identified as suitable alternatives, necessitating a concentration on this framework instead of delving into the Word Perceived Model. This determination was guided by the datasets' robustness and pertinence in relation to the research objectives.

Furthermore, the ultimate justification for focusing efforts on this module lies in its potential implications across diverse research domains, with a particular emphasis on IR. The component is meticulously designed to interact with virtual assistants and search engines, where a critical element of user engagement involves information retrieval. This offers a distinct opportunity to examine the efficacy of information retrieval processes and the incorporation of neural data into the IR domain. Evaluating the efficacy of information retrieval, while investigating the potential applications of EEG data in augmenting IR systems, represents a significant pathway for interdisciplinary research.

In Chapters 4 through 6, this thesis embarks on a thorough exploration to examine the alignment between the representations generated by state-of-the-art natural NLP models and the manner in which the human brain processes natural language. This section is dedicated to investigating this intellectual synergy in detail. Moreover, the thesis attempts to provide a comprehensive framework aimed at identifying and

analysing the various aspects involved in training deep learning models. The focus is cast more sharply on transformer models, particularly in the context of their application to brain data. Through these investigations, the thesis seeks to elucidate how such models might be refined and adapted for optimal performance in this domain. Finally, it attempts to construct a versatile encoder-decoder model with the capability to generate coherent text from input brain data. This initiative is the culmination of efforts to bridge neuroscience and artificial intelligence, contributing to the broader understanding of cross-disciplinary methodologies in automating text generation from neural processes. These chapters collectively aim to advance the frontier of research that lies at the intersection of machine learning and cognitive neuroscience.

3.7 Chapter Summary

In summary, this chapter presents a comprehensive architectural framework designed to enable user interfaces solely through BCIs, effectively eliminating the need for traditional input mechanisms by leveraging cutting-edge neuroscience techniques. It offers a detailed breakdown and analysis of the critical elements necessary to develop such a system, highlighting potential constraints and technical hurdles. The thesis emphasizes the importance of non-invasive approaches, particularly EEG, to maintain the system's portability. Managing the acquisition, processing, and handling of EEG data is crucial, as is tackling the portability concerns related to current EEG systems.

The Brain Recorder is primarily dedicated to the acquisition of EEG data, with notable advantages in terms of its portability and capabilities for real-time data processing. However, the existing systems are often cumbersome, presenting a need for innovative strategies to achieve more compact designs. The Frame Handler is responsible for the organization of EEG data in preparation for subsequent analysis, addressing challenges such as noise reduction and temporal stability through the utilization of multiple recordings.

The Brain-To-Text Decoder facilitates the conversion of neural signals into textual form, thereby integrating cognitive processes with digital interaction systems. This component employs advanced deep learning methodologies and LLMs to enhance the

precision of textual outputs, addressing issues related to data heterogeneity and the intricacy of model design. The thesis examines the adaptation of speech-to-text methodologies for application to neural data, highlighting the lack of established frameworks for such conversions.

The thesis represents a groundbreaking venture in the integration of neuroscience and artificial intelligence, concentrating on addressing the challenges associated with EEG data standardization and the development of robust models for the interpretation of neural signals. This work seeks to promote interdisciplinary research, particularly in domains such as information retrieval, and exhibits potential for substantial practical applications.

Chapter 4

Development of an fMRI-Based Brain Encoder and Identification of Brain Alignment in Machine Learning Models.

4.1 Introduction

The field of NLP has undergone significant advancements over recent decades, resulting in the creation of advanced deep learning models. These models, including transformer-based architectures, have established new standards across various NLP tasks due to their ability to discern intricate patterns within textual data. Notwithstanding their achievements, the mechanisms by which these models internally encode and represent linguistic subtleties remain obscure. This opacity is chiefly attributable to the models' inherent complexity and large-scale characteristics, which pose challenges to traditional interpretability frameworks.

Traditionally, linguistic theory posits that the understanding and production of language are directed by explicit grammatical and syntactic rules. In contrast, contemporary deep learning models challenge this premise by acquiring statistical representations of language that surpass explicit rule-based methodologies. These models utilize

extensive corpora that encompass a myriad of linguistic phenomena, allowing them to internalize a broad spectrum of language characteristics in a generalized fashion. Subsequently, these acquired representations can be fine-tuned to enhance performance on particular downstream tasks, demonstrating superior effectiveness relative to models focused exclusively on isolated NLP tasks [40, 156, 157].

As a result, a considerable body of research is directed towards elucidating the latent representations of these models in order to gain a clearer understanding of their decision-making processes. This pursuit has led to the development of myriad innovative methodologies aimed at dissecting model behaviour within regulated linguistic contexts [158–160]. By amalgamating insights from these efforts, the NLP community endeavours to align the predictive capabilities of the models with transparency, thereby promoting progress in both theoretical linguistics and practical applications within AI.

Previous investigations have sought to offer a more theoretical analysis of the representational capabilities of word embeddings [161–163]. Toneva and Whebe [1], aiming to clarify the internal representations of four NLP models, introduced a novel methodology utilizing brain recordings obtained through functional magnetic fMRI and MEG. They proposed that the alignment between cerebral data and the features extracted by a model could enable the learning of a mapping from cerebral to model representations, thereby enhancing the understanding of how these models encode linguistic information. This process is identified as the alignment of neural network representations with cerebral activity.

BERT was among the models examined by [1]. Since its inception [40], BERT has garnered considerable academic interest, with extensive efforts aimed at augmenting its performance, addressing particular inadequacies, or refining it to achieve domain-specific expertise [164–170]. Toneva and Whebe [1] successfully developed a mapping of representations from BERT, ELMo [156], USE [171], and T-XL [172] to neural data. However, this mapping was not validated against more advanced versions of BERT. In our research, we adopt the same innovative methodology outlined by [1] to analyse four advanced derivatives of the BERT model: RoBERTa [168], DistiliBERT [167], ELECTRA [170], and ALBERT [169].

Moreover, this innovative approach acts as a proof of concept to demonstrate the correspondence between cognitive and artificial neural systems. The primary objective was to examine the potential impact of omitting various punctuation symbols on the outcomes of the study. This investigation seeks to improve the congruence between neural networks and cognitive processes. Previous research on punctuation, including the study by Moore et al. [173], has highlighted its influence on reading behaviour. While punctuation assists readers in facilitating text navigation, its semantic role and cognitive processing mechanisms remain relatively under-explored. Substantial research investigates the semantic and syntactic processing of textual information within the brain [60–62].

Furthermore, investigations like those by Acunzo et al. [63] have examined particular neural areas associated with specific linguistic functions, such as the processing of intricate lexical semantics. Despite the diversity in the aims of these studies, they typically utilize an fMRI experimental framework in which participants read sets of sentences. Within experimental designs, a determination must be made concerning whether to include or exclude punctuation in the pre-processed text shown to participants; however, a consensus regarding this methodological decision remains elusive.

4.2 Methodology

4.2.1 Data

The fMRI dataset utilized in this study was initially acquired as a component of previously published research conducted by Wehbe et al. [34]. This dataset has been made available in the public domain for academic use, with the relevant data and original code accessible through the following repository.[1].

The experimental design of the original investigation comprised the participation of eight human subjects, each of whom engaged in a continuous reading of Chapter 9 of 'Harry Potter and the Sorcerer's Stone' [174]. Data acquisition was executed in four distinct sessions, wherein each lexical item from the text was presented for a duration

[1] https://github.com/mtoneva/brain$_l anguage_n lp$

of 0.5 seconds on a screen. Correspondingly, cerebral activity, quantified via the BOLD signal, was recorded every 2 seconds, corresponding to a single Time Repetition (TR). This temporal arrangement ensured that each brain volume encapsulated information pertaining to four consecutive words, as preserved in their original narrative sequence within the literary text. The dataset available for analysis, having been preprocessed and smoothed by the original investigators, is publicly accessible in its refined form. Although the initial experimental study also incorporated MEG data, this research venture exclusively utilizes the fMRI data, as they are the only dataset extant for public access.

It should be emphasized that the alignment between words and brain recordings is established within the dataset. Consequently, during the training phase, we can accurately associate specific words with each TR, thereby ensuring precise training on the correct words and confirming the consistency of the alignment. Furthermore, the dataset comprises a total of 5176 words, encompassing repeated instances, with a subset of 1840 unique words. Functional images were acquired utilizing a Siemens Verio 3.0T scanner at the Scientific Imaging and Brain Imaging Center at Carnegie Mellon University. The employed protocol was a T2*-weighted echo planar imaging (EPI) pulse sequence characterized by a repetition time (TR) of 2 seconds, an echo time (TE) of 29 milliseconds, and a flip angle of 79 degrees, covering 36 slices and generating inline graphic voxels.

4.2.2 Transformer models

In our effort to find a more advanced transformer model that might produce better brain alignment we tested four different models against BERT, which we used as our baseline. Each model was selected based on its new characteristics compared to BERT.

- **RoBERTa** constitutes a significant advancement in the realm of transformer-based NLP models, subsequent to the initial introduction of the BERT model. As highlighted by the foundational research of [168], the effective training of neural networks necessitates substantial computational resources, with a critical reliance on precisely determined hyper-parameter configurations. Through an exhaustive

comparative examination of the methodologies utilized in the two primary tasks for which BERT was originally pre-trained, namely MLM and NSP, RoBERTa has been established as a superior iteration of BERT, exhibiting improved performance metrics across these tasks. The primary rationale for incorporating RoBERTa into our study is to conduct an in-depth investigation into its internal representational structures with a focus on evaluating the potential impacts of differing hyper-parameter selections in NLP experiments. This inquiry seeks to determine whether such parametric adjustments might result in representations that more closely resemble the methods by which the human brain processes and contextualizes information. Understanding these elements is essential to narrowing the divide between artificial model representations and genuine cognitive processes.

- **DistiliBERT** Transformers have revolutionized the field of NLP in recent years, providing state-of-the-art results across a variety of tasks. However, these models have inherently large architectures, which poses challenges in terms of computational resources required for both pre-training and fine-tuning processes. Among popular Transformer-based models, BERT is noted for its effectiveness, yet its significant size renders it computationally expensive. As an innovative solution to address these computational limitations, [167] proposed DistilBERT—a distilled version of the original BERT model.

  DistilBERT has been meticulously designed to contain approximately 40% fewer parameters, while concurrently attaining a 60% decrease in training duration, without considerable compromise in performance. Notably, it preserves up to 97% of the original BERT model's capacity for language comprehension. This significant diminution in both model size and training time is a direct result of the distillation process, which leverages the concept of knowledge distillation whereby a smaller student model assimilates the behaviour of a larger teacher model.

  The undertaking by [167] to develop a condensed model underscores the poten-

tial of DistilBERT not only to reduce computational demands but also to be applicable in real-time environments. DistilBERT's streamlined architecture is particularly noteworthy for experimental evaluation, as it presents a significant opportunity to assess whether its internal language representations maintain the same integrity and depth as those of the larger BERT model, despite its reduced size and enhanced computational speed. This feature is particularly relevant for applications that require swift yet precise language processing capabilities, rendering DistilBERT an appealing subject for further empirical study in the domain of cutting-edge NLP applications.

- **ALBERT** ALBERT, as initially introduced by [169], is designed to comprehensively tackle the critical challenge presented by the escalating scale of modern NLP models, which often encounter memory limitations when deployed on GPUs or TPUs. To address this important concern, the authors proposed an elegant two-fold parameter reduction strategy. The first aspect of this innovative approach is termed 'factorized embedding parameterization,' where the authors postulated that by reducing the dimensionality of the embedding space before integrating it into the model's hidden layers, the embedding space could be condensed without negatively impacting model performance. This sophisticated technique illustrates an advanced comprehension of embedding space dynamics and their optimization. The second element of ALBERT's parameter reduction strategy is 'cross-layer parameter sharing.' This mechanism entails the consistent application of parameters across all layers of the ALBERT architecture, thereby enhancing computational efficiency while preserving the model's representational capabilities.

  In the seminal study [1], the model's features are systematically extracted on a per-layer basis, enabling distinct evaluation and prediction utilizing neurological datasets. This layer-by-layer analysis is crucial, providing detailed insights into the model's ability to map cerebral data. We hypothesize that should ALBERT exhibit superior accuracy in brain data prediction tasks, it may suggest the brain's utilization of a uniform and consistent weighting mechanism across its hierarchical layers.

- **ELECTRA** ELECTRA, a pioneering model devised by [170], was purposefully engineered to redefine the core principles underpinning the MLM task, which serves as a fundamental pre-training paradigm of the original BERT architecture. Awareness of the extensive computational demands inherent in current NLP frameworks for effective training prompted the researchers to propose a transformative alternative to the traditional MLM methodology.

  Instead of utilizing the [MASK] token to conceal particular tokens within the input sequence, which subsequently requires the model to predict likely alternatives, they proposed an innovative methodology wherein these tokens are replaced by solutions derived from a succinct yet effective generation network. The primary learning model subsequently undergoes pre-training through the evaluation of the authenticity of these generated tokens, categorizing them as either correct or incorrect within a binary classification framework.

  The authors advance a hypothesis suggesting that this re-conceptualized methodology produces superior contextual representations compared to those generated by BERT. This hypothesis positions the ELECTRA model as an intriguing candidate for assessing its effectiveness in correlation with neurological data, thus facilitating an inquiry into whether its internal representational framework more accurately reflects the mechanisms through which the human brain processes contextual information.

  Studies exploring the alignment between machine learning models and cognitive neuroscience are pivotal in advancing our understanding of both artificial and biological systems. Analysing the performance of ELECTRA in the context of brain data prediction may enable future research to illuminate possible convergences between AI and human cognition, thereby broadening the discourse on representation learning and its relevance across diverse interdisciplinary domains.

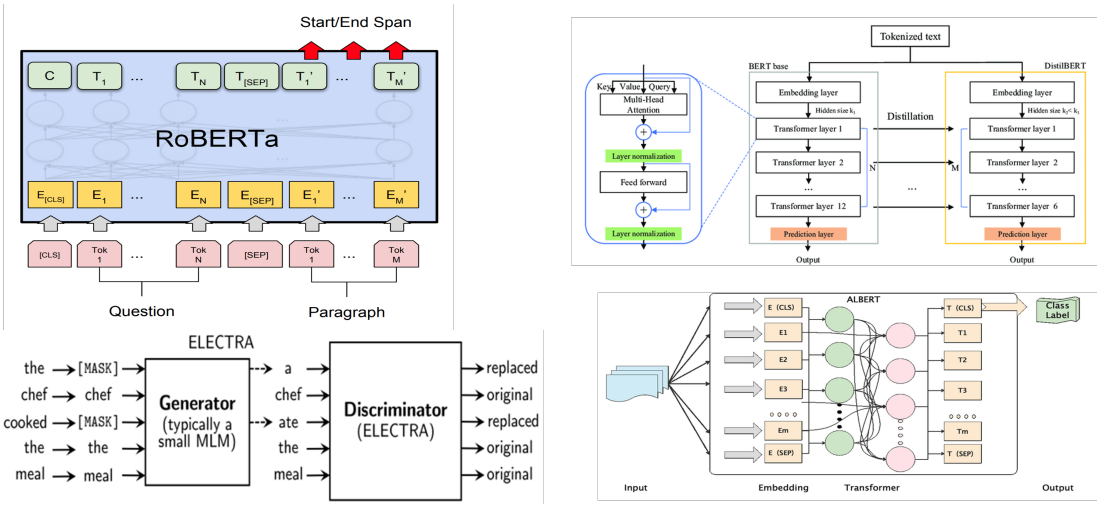


Figure 4.1: A brief description of the architecture of every model used in our experimental procedure.

4.2.3 Experimental Procedure

The methodological framework of this study is delineated into three distinct phases. In the initial phase, features were extracted from the specified transformer models using a range of sequence lengths to achieve a thorough representation of features. Subsequently, these extracted features served as inputs for training a ridge regression model to reconstruct the brain fMRI recordings from the transformer features, which was subsequently employed to infer brain data. In the final phase, the effectiveness of the predictive model was rigorously assessed using searchlight classification, a technique that evaluates classification performance across various brain regions. The following sections provide a comprehensive exposition of each phase. To ensure transparency and reproducibility, the code underpinning this analysis has been made publicly accessible [2].

Extracting features from models

In the context of feature extraction from models, the preliminary stage of the experimental protocol entailed the methodical retrieval of features from a variety of sequences of predetermined lengths denoted by S. The foundational research referenced in [1]

[2] https://github.com/NeuraSearch/Brain-Transformer-Mapping-Punctuation.git

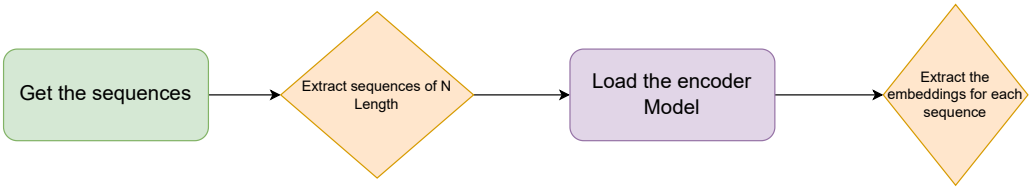


Figure 4.2: A high level representation of the pipeline used to extract NLP features from the different NLP models.

utilized sequence lengths specifically of 4, 5, 10, 15, 20, 25, 30, 35, and 40. Our experimental design adhered to this paradigm to preserve consistency with established methodologies. The sequences in question were carefully compiled from Chapter 9 of the renowned literary work, Harry Potter and the Sorcerer's Stone. A comprehensive dictionary was constructed to contain entries corresponding to each layer within the model architecture, including the embedding layer, where applicable. Initially, representations were extracted from the embeddings layer for each individual word within the chapter. Subsequently, the analysis advanced to the extraction of the first sequence of length S, followed by the acquisition of layer-wise representations for said sequence, repeated S times to ensure methodological rigour. Upon completion of each sequence iteration, a two-dimensional matrix was generated, characterized by dimensions S x L, where S represents the sequence length, L denotes the number of layers present in each model.

Furthermore, the representation extraction for each sequence of length S was facilitated through a sliding window methodology analogous to that applied to the initial sequence. The deliberate extraction of the first sequence re-iteratively S times was purposefully implemented to achieve parity between the quantity of sequences and the total number of words constituting the dataset.

Distinct scripts were developed and executed for each specific model selected for analysis. To facilitate access to pre-trained model checkpoints, the Hugging Face library was utilized, enabling the seamless downloading of these pre-trained checkpoints. This choice of library integration was instrumental in ensuring the reproducibility and scalability of the experimental procedures.

Making predictions using the fMRI recordings and the extracted features

During the preliminary phase of model development, we prioritized the creation of a rigorous computational framework designed to precisely map the complex relationship between cerebral activities and their respective neural network representations. In pursuit of this objective, a ridge regression model was meticulously developed as the fundamental element of this methodological approach. This decision was guided by an extensive evaluation of previous seminal studies [34, 35, 37, 175, 176], which successfully utilized linear functions with ridge penalties, thereby highlighting their efficacy in modelling brain-to-network mappings. To evaluate how closely NLP models mirror human cognitive processes, researchers use encoding models that predict neural responses based on the model's internal representations. Cognitive neuroscience underpins this strategy, utilizing encoding models to link stimulus features to brain activity patterns [177]. Within this paradigm, a linear model is developed to associate word or sentence embeddings from NLP systems (such as RoBERTa or ALBERT) with brain data collected during human linguistic processing. The extent to which the model's representations can predict brain activity in a linear fashion indicates representational alignment [178]. This is based on the premise that analogous cognitive processes yield similar representational structures, permitting a linear relationship between them. This technique has been employed to pinpoint which aspects or layers of NLP models most effectively correspond with the brain regions essential to language processing, offering insights into both model interpretability and human psychological functions [1].

Though ridge regression may appear simplistic, its prior validation as an efficacious tool for this application [1] offers substantial rationale for its employment. However, we recognize the opportunity for future research to integrate more advanced models, potentially achieving superior mapping efficiency. In this study, the ridge regression model was instrumental in producing predictions, leveraging brain data and features derived from the prior analytical phase.

Prediction generation was executed across multiple dimensions: individual layers, distinct subjects, and diverse models, all within the meticulously structured environment of a cross-validation framework. Specifically, a 4-fold cross-validation procedure

was utilized, in accordance with the methodologies delineated by Toneva and Wehbe [1]. The fMRI datasets were scrupulously organized into four discrete runs per validation fold, with each run being employed as a test dataset for the respective fold.

The analytical procedure began with the loading of extracted features, which were subjected to dimensionality reduction into a ten-dimensional space utilizing Principal Component Analysis (PCA). Subsequent stages entailed the meticulous temporal alignment of fMRI data with features derived from the model, establishing a robust mapping framework for the training of the ridge regression model. Initially, 1351 images per subject were recorded; however, to alleviate edge effects, this was reduced to 1211 images.

The alignment procedure required the determination of the temporal resolution associated with each word presentation to the user. This mapping was facilitated by the equivalence in length between sequences and words. The index of the sequence was directly aligned with the TR index. Upon completing this mapping, we concatenated representations from preceding time points (t-4, t-3, t-2, t-1) with the current time point t, as a methodological improvement aimed at enhancing model representation and predictive accuracy [1].

Following alignment, the analytical procedure culminated in a set of 1351 temporally anchored values. Subsequent adjustments involved the removal of edges from each run, ultimately condensing the dataset to 1211 time-stamped sequences. Upon the completion of preprocessing, the training of the ridge regression model commenced. The predictive weight for each voxel was meticulously computed by exploring an extensive range of lambda values within the ridge regression framework, specifically $\lambda = 10^x$ where $-9 \leq x \leq 9$. The lambda value that minimized error for each voxel was crucial in constructing the model's final weight matrix. During each iteration, the error was quantified using the R2 error metric to ascertain the disparity between the synthesized brain recordings and the actual ground truth data at specified temporal points. This error was subsequently employed as the training loss within the ridge regression framework. The resultant model, developed through these comprehensive processes, was subsequently applied to test data, generating predictions ready for further exami-

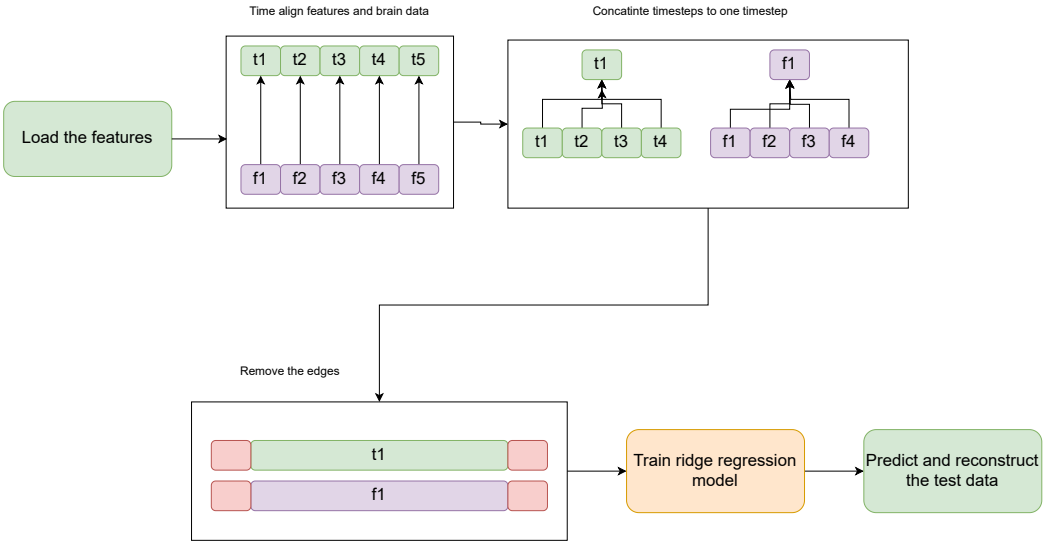


Figure 4.3: This figure depicts the procedure involved in loading the natural language processing features until the reconstruction of neural data from the established features. The NLP features are temporally aligned with the neural features, extraneous edges are eliminated, and a ridge regression model is subsequently trained.

nation.

Evaluating the predictions

The primary aim within our methodological framework was to conduct a rigorous evaluation of the predictive efficacy of the ridge regression model, which had been meticulously trained using neural data to map representations derived from the brain to those derived from neural networks. To ensure a robust evaluation, we implemented and refined the advanced searchlight classification algorithm as previously elaborated by Toneva and Wehbe (2019) [1]. This approach entailed a binary classification task conducted over spatially adjacent voxels for each individual subject's dataset. Consequently, the comprehensive analysis of voxel-wise predictive performance was anchored in pre-computed spatial neighbourhoods, a strategic element provided by the foundational research. These neighbourhoods facilitate a localized examination of voxel interactions and bolster the reliability of predictive accuracy evaluations.

The evaluation process employed the fMRI data recorded from each subject, whereby

a stochastic method was implemented by randomly sampling a segment of 20 TRs, considered to be indicative of correct predictions. Simultaneously, from the predictions generated by the neural model, an equivalently indexed segment of 20 TRs was extracted to provide a standard for accurate prediction comparison. Furthermore, to introduce a contrasting element, another random selection of a 20 TRs segment was drawn from the forecasted data and designated to signify incorrect predictions.

Upon delineation of these parallel data segments, the evaluation progressed with the computation of Euclidean distances within the framework of predefined voxel neighbourhoods. For each voxel, the analysis entailed examining distances between the accurately predicted segment and the predicted segment, subsequently contrasting these against distances associated with the inaccurately predicted segment. When the Euclidean metric demonstrated a closer proximity of the predicted sequence to the correct segment than to the incorrectly labelled one, the prediction for that voxel was recorded as accurate.

This evaluative process was conducted in an iterative manner, executing 1000 unique trial applications per voxel, thereby producing an extensive evaluative dataset. Following this, a statistical synthesis of the accuracy metric was undertaken, involving the computation of the average accuracy across all voxels for each cross-validation fold pertinent to each individual subject. This meticulous procedure not only ensured the reliability of the evaluation but also furnished a refined comprehension of the ridge regression model's prediction accuracy, with spatial detail at the voxel level, across the entire cohort of subjects.

Removing punctuation

To rigorously assess the semantic processing of punctuation symbols by the human brain, we employed a methodology similar to that previously described, with a slight alteration in its initial stage. The study involved the implementation of four distinct scenarios, each aimed at the removal of punctuation characters before the sequences were introduced into computational models for analysis. Following their removal, the feature extraction process was conducted in a manner analogous to that previously

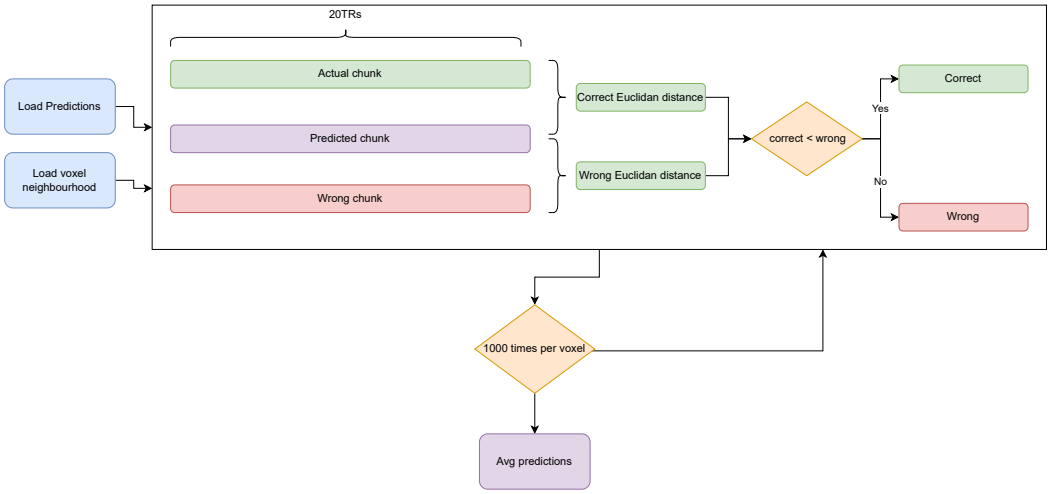


Figure 4.4: This diagram explains using a graphical interfaces the process of evaluating the predictions.

detailed, albeit applied to the altered text sequences. After the extraction of these sequences, the subsequent stages of the methodology remained unchanged, adhering to the previously established procedure. Our investigation focused on the following four scenarios:

1. The initial scenario encompasses the replacement of the fixation symbol "+" with the token "[UNK]", a process subsequently referred to as 'Removing Fixation'. This procedure was meticulously implemented to rigorously evaluate the impact of the fixation symbol's textual presence within input sequences on the fidelity and alignment of the resultant feature set with human brain processing. By substituting the fixation symbol with the "[UNK]" token, this alternative symbol functions to denote an unknown token to the computational model, thus potentially altering the model's interpretative framework. It is further hypothesized that the substitution impacts the processes of semantic extraction and subsequent cognitive mapping, thereby enhancing our understanding of symbolic influence in computational linguistics. Consequently, this scenario highlights the significance of unknown symbols within textual processing, offering insights into the complexities introduced by symbolic variances and their implications for neuro-computational modelling. As a result, the methodological refinement introduced

by this scenario not only underscores the potential disruption in traditional feature interpretations but also suggests broader implications for the fidelity and neuro-semantic coherence of computational models. This scenario was labelled "Removing Fixation" in our experimental results.

2. Within the esteemed realm of cognitive linguistics and neurocomputational modelling, we outline a secondary experimental condition termed 'Padding fixation,' which meticulously examines the semiotic function of fixation symbols within textual sequences. This scenario is of critical importance in our methodological framework as it investigates the effects of replacing the traditional fixation symbol '+', widely employed in sequence alignment processes, with the '[PAD]' token—a methodical placeholder utilized within computational paradigms to denote non-essential data points. The rationale for this substitution lies in its ability to systematically eliminate the interpretive significance of the fixation symbol, thus enabling a clearer evaluation of its underlying influence on the procedural effectiveness of feature extraction.

   Our theoretical framework suggests that the incorporation of the '[PAD]' token induces a significant transformation in the perceptual structure of computational models, thereby reconstituting the algorithmic pathway through which textual sequences are processed and cognitively represented. By pre-emptively rendering the fixation symbol computationally irrelevant, we assert that this strategic obfuscation reveals the inherent adaptability of human-computer interaction systems, particularly concerning brain-aligned feature representation. Consequently, the conceptualization of 'Padding fixation' as a methodological innovation facilitates a sophisticated examination of symbolic retrieval mechanisms, elucidating the foundational neurosemantic alignment that characterizes adaptive information processing within the realm of human linguistic undertaking. This scenario is termed 'Padding Fixation' in our experimental outcomes.

3. In the specified scenario, the substitution of special characters such as double hyphens "–", ellipsis "...", and em dash "—" with the "PAD" token (padding

all instances) is systematically executed to homogenize the text extracted from a selected chapter of the Harry Potter series. This investigative methodology facilitates a comprehensive assessment of the occurrence and prevalence of these particular non-alphanumeric sequences within the corpus. The principal aim is to determine the potential implications of these characters on the efficacy and accuracy of the model's operational functionalities, specifically related to its use in brain-network mapping tasks. By detailing the transformative process of character substitution and *-*/*/- the consequent text modifications, the study aims to augment the comprehension of the model's resilience in managing textual variability. Such comprehension is essential for identifying the impact of non-standard character components on computational analyses involved in intricate network mapping, thereby offering significant insights to the domain. Consequently, this experimental framework serves as a critical evaluation of character-induced variations in model performance metrics when applied to advanced neuro-informatics studies. We have designated this scenario as "Padding All" in our experimental results.

4. In the pursuit of extending the preliminary investigative scenario pertaining to character substitution, an exhaustive modification has been executed to encompass additional non-alphanumeric and punctuation elements within the analytical framework. Specifically, this modification involves the substitution of selected special characters, namely the double hyphen "–", ellipsis "...", em dash "—", alongside the period "." and question mark "?" with the uniform "PAD" token. The rationale underlying this expansion arises from the imperative to determine the implications associated with the excision of these ubiquitous punctuation marks. These marks embody essential components of sentence structure and syntactic demarcation within the corpus, crucial to maintaining textual coherence and continuity.

   Through a systematic expansion of the range of target characters to include previously unexamined elements such as periods and question marks, this study seeks to rigorously assess the impact of character standardization on the fidelity and

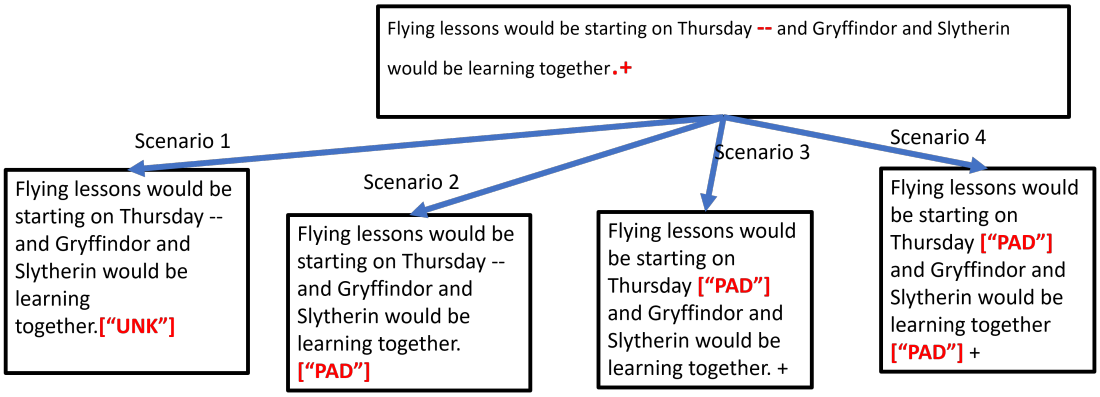


Figure 4.5: This Figure illustrates how we changed a sample text depending our 4 different implementation scenarios on testing the role of punctuation in text comprehension in the brain.

precision of sequence mapping algorithms. Considering that these punctuation marks function as essential syntactic and semantic signals within the corpus, their intentional removal and replacement with the "PAD" token offers a sophisticated network perspective on how their absence might affect elucidation and interpretative accuracy in computational text mappings.

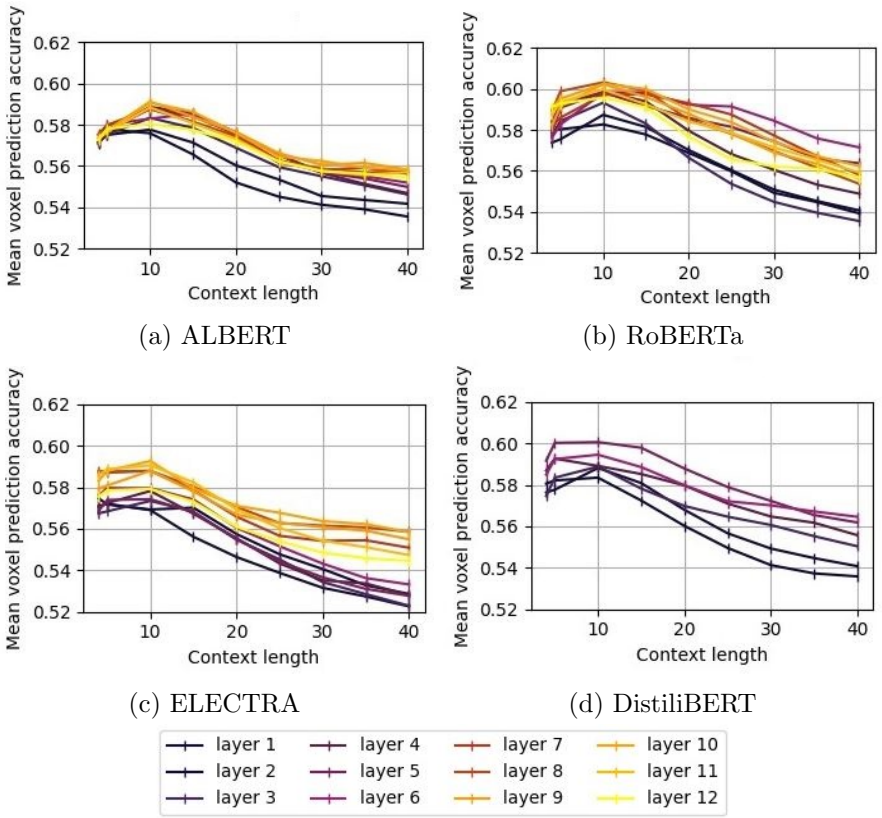
This approach invariably facilitates a deeper understanding of the fundamental principles governing character substitution methodologies in neural linguistic tasks and is positioned to provide substantial insights pertinent to enhancing accuracy rates in sequence mapping. Such insights are crucial for the advancement of methodologies within the field of neuro-informatics, wherein even the subtlest character modifications can result in significant variations in model performance. Therefore, this adjusted experimental framework represents a deliberate and comprehensive investigation into character selection strategies, further reinforcing its potential as a foundational research endeavour within the sphere of PhD-level academic exploration. We referred to this scenario as "Padding Everything" in our experimental results.

4.3 Results and discussion

Comparing the models: We have conducted a thorough analysis of the performance metrics of BERT in comparison to the range of models outlined in Section 4.2.2. In this section, we articulate the differences among these neural network architectures and examine the potential implications these empirical results might have in understanding the neural mechanisms that underpin human cognitive functions.

A consistent pattern observed across the diverse range of neural network models, including BERT, is the marked decrease in the predictive accuracy of the ridge regression model as sequence length increases. This phenomenon highlights the intrinsic limitations these architectures face in sequence processing. All the empirical findings elaborated upon in this discussion are systematically depicted in Figure 4.6, offering a visual representation of the data outcomes discussed.

In addition, to ascertain the validity of the results for each model, we conducted a statistical review to evaluate statistical significance. We further ensured that each model exceeds the random baseline of 50%. This benchmark was selected due to the nature of our evaluation task, which involves a searchlight classification within a binary framework, necessitating the surpassing of a 50% random baseline. The results outlined in the subsequent subsection represent the mean accuracy over 30 iterations for each model, with an average standard deviation(STD) of ~ 0.01 for the models and ~ 0.02 for the punctuation scenarios , thereby ensuring consistency across trials. More over to ensure statistical significance we conducted pairwise t-tests for all the models against the random and BERT baselines and also for our handling punctuation methods against the same baselines. Our results showed that an initial one-way ANOVA revealed a significant overall effect with a p-value of $p \leq 0.05$. When dissecting the results for the individual pairs regarding the models all the pairs had a statistical significance with a p-value of $p \leq 0.05$ except distilibert against the BERT baseline that had a p-value 0.15. Finally the same process was conducted for the remaining methods when removing punctuation from the text. Our results again showed a p-value of $p \leq 0.05$ signifying statistical significance against our 2 baselines.



(a) ALBERT (b) RoBERTa

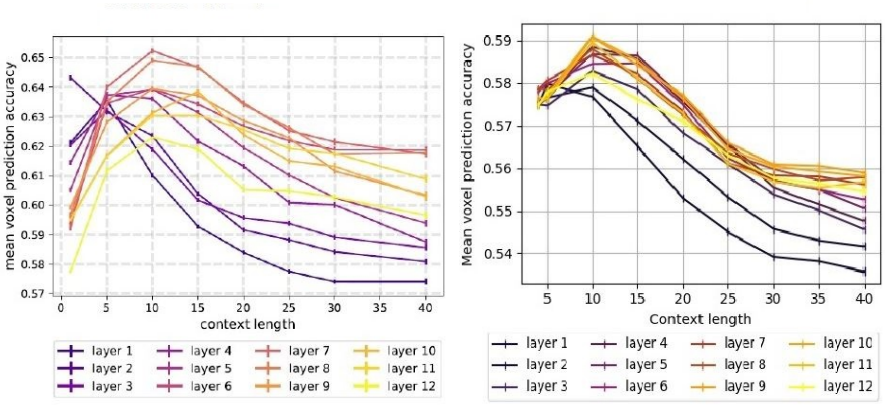
(c) ELECTRA (d) DistiliBERT

Figure 4.6: This figure shows the overall accuracy across all subjects for the four different models we investigated. It is clearly shown that RoBERTa and DistiliBERT outperform the other 2 models and the baseline and are the most brain-aligned models.

BERT

Adhering closely to the experimental procedure detailed in Section 4.2.3, and utilizing the same dataset, we were unable to perfectly replicate the original findings reported by [1]. Consequently, our reproduced baseline, intended as a comparative standard with other models, was established using our independently developed code. This baseline is presented in Figure 4.7, alongside the original findings. We surmise that the discrepancies in results can be attributed to three main factors: firstly, the precise methodological details of the averaging process across subjects and multiple folds for each neural network layer were insufficiently explained in the original publication. Another complicating aspect pertains to the application of PCA, specifically, the stochastic state employed in this analysis was not specified, potentially introducing an additional

variable affecting our ability to replicate the exact outcomes. Lastly, our computational infrastructure differs from that of the original authors, which could further contribute to the outcome differences. Nevertheless, we maintain confidence in the accuracy of our results. A detailed analysis of the outcomes, as depicted in Figure 4.7, which simultaneously presents the findings from [1] and our reproduced data, reveals that although the qualitative trajectory or contour of the data remains consistent, noticeable quantifiable disparities exist in the precise values. Noteworthy consistencies include the observation that our reconstructed results reach a peak at a sequence length of ten, aligning with the peak identified in the original study. Additionally, in agreement with the original outcomes, the observed accuracy declines as the sequence length increases. This comprehensive analysis highlights the inherent complexity and potential sources of variation in replicating experimental outcomes within the realm of computational modelling.



(a) BERT Original Results (b) BERT Reproduced Results

Figure 4.7: On the left hand side are the original results reported in [1]. On the right hand side are the reproduced results we obtained when running our code for BERT. Note that the y-axis has a different minimum and maximum value for the two panels, original and reproduced.

ALBERT

The primary model under consideration in this research is the ALBERT model, an acronym for A Lite BERT. ALBERT is distinguished by its unique architectural de-

sign, which entails the utilization of a shared representation across all network layers. This architectural strategy was rigorously investigated to ascertain its impact on model performance, particularly in relation to BERT, from which ALBERT is derived. Notwithstanding the architectural divergence, detailed analysis revealed that there are no appreciable differences in accuracy levels between ALBERT and BERT. This finding suggests that the alterations implemented in ALBERT were unsuccessful in enhancing the congruence of its latent representations with neural representations as efficaciously as BERT. Moreover, the implementation of shared weights across ALBERT's various layers, which is posited to emulate a cognitive mechanism akin to the shared 'weights' model in the human brain, did not result in improved alignment between patterns of brain activity and neural network activations. This phenomenon underscores a critical area for future inquiry and verification within the disciplines of cognitive neuroscience and artificial intelligence: the question of whether the brain employs a comparable strategy of shared representations across its neural architecture remains an unresolved issue meriting further empirical study.

RoBERTa

In the investigation of transformer-based architectures, a comprehensive examination was undertaken on the RoBERTa model within this research. The observations identified a consistent pattern within the graphical data associated with RoBERTa that mirrors the trends originally observed in BERT. Empirical analyses of the models indicate that RoBERTa demonstrates marginally superior accuracy across almost every layer in comparison to its predecessor, BERT. This finding is particularly prominent at the peak of the model's performance, where RoBERTa surpasses BERT. This enhancement can be attributed to the strategic selection and adjustment of hyper-parameters which, when meticulously calibrated and optimized, result in improved internal representations. These representations demonstrate a closer resemblance to cognitive structures analogous to those found in human neural architectures. Such insights highlight the potential benefits of optimizing training parameters to enhance model efficacy and cognitive similarity. Consequently, the methodological decision to adjust hyper-parameters

plays a crucial role in advancing the parallelism between artificial neural models and human cognitive processes.

DistiliBERT

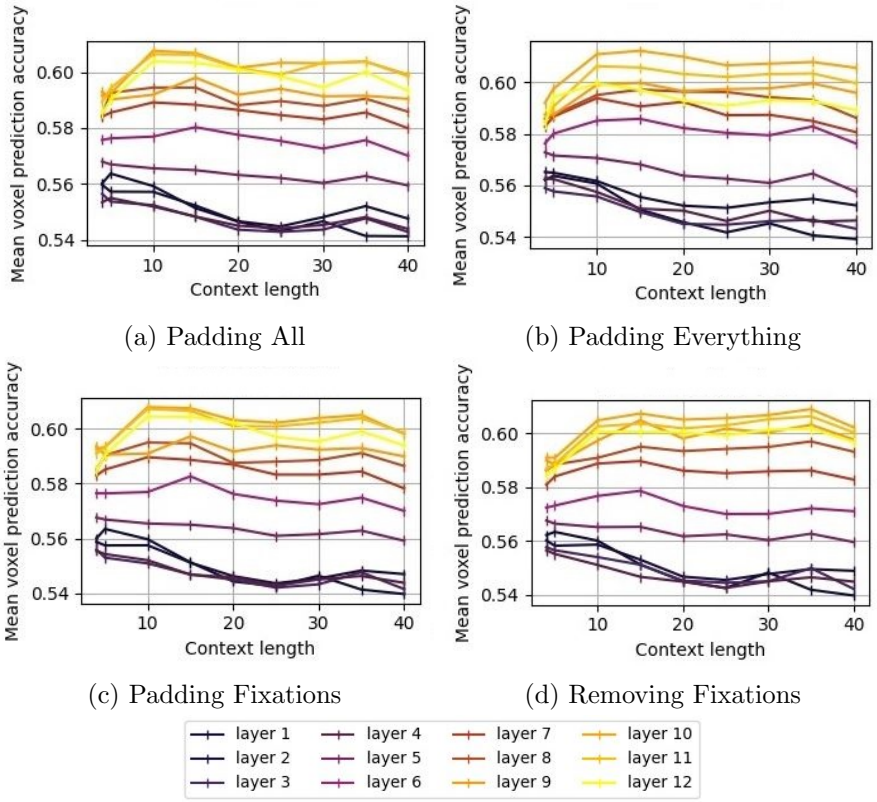
An in-depth analysis conducted with DistilBERT reveals that, notwithstanding its reduced number of layers and a smaller architectural footprint in comparison to the original BERT model, the performance metrics of DistilBERT's layers demonstrate superior capabilities across all evaluated dimensions. This enhancement in performance, although not exceedingly significant, underscores DistilBERT's ability to effectively capture and represent nuanced semantic information at a level equivalent to, or even exceeding, that of BERT. This phenomenon can be attributed primarily to the robust inner representations that DistilBERT develops, attesting to its competence in semantic comprehension. Furthermore, it is noteworthy that the neural correspondence, often referred to as 'brain alignment,' present within DistilBERT's layers is consistently on par with that observed in the more extensive and heavier BERT model, thus highlighting the efficacy of DistilBERT's design in maintaining high levels of neural representational alignment.

ELECTRA

Upon examining the findings related to ELECTRA, it becomes evident that ELECTRA's performance does not surpass the baseline results achieved with BERT. Despite the consistency of our observations across both graphical depictions, ELECTRA's inability to outpace BERT is noteworthy. This consistency indicates that its training strategies did not yield a level of alignment with cognitive processes comparable to that of BERT. From these observations, one can infer that the methodology employed during ELECTRA's training phase has not produced a model architecture that is as closely aligned with human neural patterns as BERT's architecture. This insight suggests a potential limitation within ELECTRA's training regimen, which inadequately mimics the brain's architecture and functioning compared to BERT.

Results with Removing Punctuation: After using the corpus without any modi-

fications, we wanted to modify the corpus by removing punctuation symbols. In doing so, we wanted to see what this might suggest of how the brain semantically processes punctuation symbols. In a comprehensive investigation into the effects of punctuation



(a) Padding All

(b) Padding Everything

(c) Padding Fixations

(d) Removing Fixations

Figure 4.8: The figures presents the results of the 4 different punctuation scenarios. The immediate observation is that the first 6 layers are more brain aligned than the last 6 and also that as the context length increases the role of punctuation to understand the text meaning is less significant.

modification on neural language models, it was observed that alterations in the usage of punctuation influence model accuracy, particularly in specific layers. This study examined four distinct scenarios of punctuation alteration and disclosed a notable trend: improvements in accuracy were confined to layers 7 through 12. Although these improvements were not universally significant, with the highest observed enhancement in performance amounting to approximately 1.5%, they nonetheless highlight an important pattern. Of particular interest was the observation that treating punctuation as padding tokens alleviated the typical accuracy loss associated with increased sequence

lengths. This finding implies that punctuation, when employed as padding, minimally contributes to the degradation of model performance as sentence length increases. Furthermore, the study identified a distinct functional division at layer 6, suggesting it serves as a demarcation between the model's initial and terminal layers.

These findings are consistent with previous research conducted by Toneva et al. (2019) [1], which hypothesized a discrepancy in neural alignment across the layers of BERT, with layers 1 through 6 exhibiting reduced alignment in contrast to layers 7 through 12. This antecedent study suggested that the removal of the attention mechanism from the first six layers could potentially augment the model's representational efficacy.

Drawing from these observations and the assertions made by Toneva et al. (2019) [1], it is reasonable to propose the hypothesis that human cognitive processing does not heavily depend on punctuation for semantic comprehension. Furthermore, as sequences are extended, thereby conveying additional information, the model's performance remains largely unaffected. This resilience underscores the limited role of punctuation in semantic interpretation. Collectively, these findings enhance the understanding of how language models engage with punctuation and inform hypotheses about the cognitive mechanisms underlying human language processing.

4.4 Chapter Summary

In this chapter, a comprehensive examination was conducted on four distinct transformer models: ELECTRA, RoBERTa, ALBERT, and DistilBERT. The principal objective was to determine which of these models, in comparison to a baseline defined by BERT, is capable of producing language representations that more closely resemble those of the human brain. This similarity is quantitatively evaluated through alignment with fMRI brain data. The results of our investigation reveal that RoBERTa and DistilBERT exhibit the highest levels of alignment, outperforming the BERT baseline and the random baseline.

Moreover, our study encompassed an analysis across four distinct scenarios to assess the impact of punctuation symbols on semantic processing. Building on the foun-

dational research of Toneva and Wehbe (2019) [1], which introduced a methodology for translating neural network features into brain representations, we aimed to refine this feature mapping process by systematically omitting punctuation symbols from our dataset. Our experiments indicate that the removal of punctuation symbols consistently improved model performance across all scenarios evaluated and keeping a higher score from the random baseline. Furthermore, it was observed that with increased sequence length, the decline in accuracy when predicting brain data was substantially less pronounced than previously reported.

These empirical findings provide evidence in favour of the hypothesis that the human brain may only minimally rely on punctuation marks to interpret the semantic meaning of sentences. Additionally, they imply that as the length of textual context increases, the requirement for punctuation symbols correspondingly decreases.

We remain confident that the direction of our research can be further enhanced by expanding the scope to include a broader array of transformer models in the quest to identify language processing models with closer alignment to brain activity. Furthermore, the innovative experimental framework introduced by Toneva and Wehbe (2019) [1] provides a strong basis for assessing the neural alignment of novel and emerging models. Moreover, we propose a deeper exploration of the cerebral mechanisms involved in the semantic processing of punctuation, employing various models, as a component of the larger effort to decipher the complex capabilities of the human brain in language comprehension and processing.

Chapter 5

Optimizing Brain Decoding using different Activation Functions

5.1 Introduction

In contemporary neuroscience, the exploration of the human brain has witnessed remarkable progress. Scholars across various disciplines have successfully classified mental states ranging from fundamental sensory experiences, such as the perception of pain, to more complex cognitive phenomena, including the evaluation of the significance of various stimuli [84]. This progress is indicative of the advanced methodologies employed in cognitive and psychological research. One of the pioneering advancements in this scientific domain is the decoding of human thought processes and inner speech. This endeavour is primarily driven by the urgent requirement to restore communicative functionalities in individuals afflicted with neurological disabilities. Multiple studies have demonstrated the feasibility of deciphering human thought using modalities such as fMRI [1] or EEG [145] data.

This pioneering research culminated in the development of 'brain embeddings,' through which computational models are designed to analyze neural signals and establish connections with linguistic features. Nevertheless, the application of fMRI data in real-time Brain-to-Text translation systems faces significant challenges due to two primary factors. Firstly, the acquisition of fMRI data incurs substantial financial costs.

Secondly, fMRI's reliance on the BOLD signal results in a time lag, rendering real-time application infeasible. Given these significant limitations, there has been a gradual shift in neuroscientific research towards employing EEG for real-time Brain-to-Text conversion. The advancement and accessibility of extensive datasets, such as those from ZuCo-1.0 and ZuCo-2.0 [140, 141], have considerably enhanced the capabilities of the research community in this domain. EEG data from these sources have been successfully utilized in previous studies to either decipher human cognition or to develop comprehensive EEG embeddings applicable in downstream analytical tasks, such as the classification of EEG-derived mental states [58, 145, 179].

This transition marks a critical shift towards more viable and utilitarian applications in neural decoding technologies. In the groundbreaking work by Wang and Ji [145], an EEG-based text decoder was constructed, yielding promising initial outcomes. However, there remains considerable potential to improve the effectiveness of brain-to-text translations. Hence, in this section of the doctoral research, we aim to enhance the foundational performance delineated in Wang and Ji's study by investigating the relatively underexplored domain of neural network optimization through diverse activation functions.

The selection of activation functions in neural networks exerts a profound influence on the characteristics of the features acquired as well as on the model's efficacy across various tasks. Activation functions incorporate critical non-linearities enabling networks to discern intricate patterns within data. Furthermore, divergences in their mathematical attributes—such as smoothness, sparsity, and continuity—impact the model's capability to develop robust and generalizable representations [180, 181]. Although fixed functions such as ReLU, GELU, and Swish are recognized for their established strengths, recent scholarly investigations have delved into learnable activation functions that are capable of adapting during the training process to more effectively align with specific tasks or datasets.Moreover it has been evidenced in several studies [117, 118] that incorporating nonlinear and tunable activation functions can significantly enhance model performance. These investigations further indicate a knowledge gap concerning the application of such specialized activation functions, as state-of-the-

art models predominantly employ traditional activation functions. Notably, nonlinear and tunable functions may be better suited to real-world data contexts.These adaptable functions offer a flexible mechanism for adjusting gradient flow and representational capacity, thereby enabling the tailored customization of activation patterns to meet specific task requirements during the training phase. Empirical studies demonstrate that such adaptability can lead to significant enhancements in domains such as image recognition, language modeling, and speech processing [182, 183]. Considering the nascent stage of research in text generation from EEG data, compounded by the inherently complex and noisy characteristics of EEG signals, selecting the activation function as a variable for experimentation could potentially reveal a more suitable alternative to traditional activation functions. Moreover, the use of tunable activation functions facilitates the determination of the most optimal parameters for such functions by minimizing the training error.

In this section of the thesis, we conducted a methodical examination of a broad spectrum of activation functions to assess their comparative efficacy. The results illuminate that the integration of tunable activation mechanisms can enhance model performance without requiring modifications to the foundational architecture. Additionally, the deployment of polynomial activation functions of higher degrees showed superior performance compared to their linear counterparts. The significant contribution of this research segment resides in demonstrating the potential for optimizing a Transformer Encoder by employing alternative activation functions, whether through the utilization of high-degree polynomial models or tunable activation parameters.

5.2 Methodology

To rigorously assess the potential for performance enhancement through the deployment of alternative activation functions, a comprehensive re-training of the EEG-to-Text decoder framework, originally formulated by Wang and Ji [145], was undertaken. The preservation of the original network architecture ensured a controlled experimental environment, facilitating the systematic incorporation and evaluation of diverse activation functions. The experimental protocol necessitated the application of these varied

activation functions within the context of an analysis employing raw EEG signals. This methodological approach was based on constructing an experimental baseline, achieved by training the model in its unaltered form. Subsequently, the empirical results derived from the employment of alternative activation functions were meticulously compared to this established baseline, allowing for a detailed examination of the performance implications associated with each activation function adaptation.

5.2.1 Data

During the training phase of our computational model, we engaged comprehensively with the unprocessed EEG data, meticulously recorded for each linguistic unit within each successive sentence. Although the architectural framework of our model closely resembles that proposed by Wang and Ji [145], a fundamental distinction is our utilization of raw EEG data as opposed to the processed, word-level EEG attributes. The ZuCo dataset encompasses an extensive range of experimental tasks, thereby providing diversity. The significant study by Wang and Ji [145] demonstrated that enhancing the training sample pool leads to improvements in the model's performance efficiency.

In alignment with their findings, our research employs Task 1 and Task 2 from ZuCo-1.0 along with Task 2 from ZuCo-2.0 as the primary training stimuli. To align divergent data formats, we applied advanced transformation algorithms to our raw EEG corpus prior to integration into our model's architecture for training procedures. In stark contrast, whereas word-level features align to a unidimensional sequence of attributes, raw EEG data innately manifest as a bidimensional array.

The primary dimension encompasses the number of recordings per linguistic unit, while the secondary dimension describes the levels of electroactivity across individual electrodes. To harmonize these dimensional differences and achieve the desired unidimensional format, we calculated the mean electroactivity levels for each electrode by averaging across all recordings related to a specific word. The resultant values integrated into a singular mean activation metric per electrode, producing a unidimensional array of feature vectors. This data transformation was crucial in optimizing the integration of raw EEG data within our model framework, ensuring seamless compatibility

and facilitating an effective training process.

5.2.2 Model

Our model architecture, while sharing similarities with the framework introduced by Wang and Ji [145], significantly diverges in terms of the activation functions utilized within our Transformer Encoder. This design is fundamentally consonant with the paradigms underlying methodologies for translation tasks, employing neural networks structured with encoder and decoder components. The encoder architecture integrates a series of Transformer Encoder layers, each initialized with random weights, intended to distill high-level feature representations from raw EEG signals. These abstracted features subsequently inform the operations performed by the decoder.

A notable deviation from the prototypical study is our incorporation of distinct activation functions within these Transformer layers, aimed at enhancing the efficacy of the feature extraction mechanism. Our configuration specifically consists of X layers of Transformer Encoders, each possessing a dimensionality of N and incorporating S attention heads to ensure robust processing. During the decoding phase, we employ a pre-trained LLM, specifically the Bidirectional and Auto-Regressive Transformers (BART) model, which Wang and Ji [145] also utilized. BART's capabilities in generative tasks are exemplary, rendering it eminently suitable for converting EEG-derived features into logically structured and cohesive sentences. Furthermore, the selection of BART is predicated on its pre-training via a denoising auto-encoder objective, which is particularly advantageous for addressing the inherently noisy characteristics of EEG data. Moreover, BART's autoregressive generation capability allows for the prediction of each word contingent upon its predecessors, thereby potentially enhancing accuracy by leveraging the context generated at each time step as an ancillary modality, rather than solely relying on EEG data [184].

Model Operation

Encoder: The randomly initialised stack of Transformer Encoder layers processes the raw EEG data, extracting meaningful features that encapsulate the information within

the EEG signal.

Decoder: The BART model takes these high-level features from the encoder and generates the corresponding sentences. This combination of a customised Transformer Encoder and a robust LLM decoder allows our model to effectively perform the EEG-to-text translation task.

5.2.3 Activation Functions

During the training and evaluation stages of our investigation, a thorough exploration of different activation functions was executed. Activation functions are crucial to the efficacy of neural networks as they determine the level of non-linearity incorporated, thus greatly affecting the model's capability to learn intricate patterns.

In our experimental framework, we meticulously evaluated a variety of activation functions under controlled conditions with the aim of maximizing model performance. This approach sought to transcend the traditional use of the ReLU, which is extensively implemented as the default activation function in Transformer Encoder architectures. Each activation function was rigorously assessed using a range of performance metrics to ascertain those that might outperform ReLU, especially concerning the processing of EEG data within our model structure.

1. The Swish activation function, as articulated by Ramachandran et al. in their foundational paper [185], constitutes a noteworthy progression in the field of activation functions, particularly within the realms of machine learning and deep learning applications. Distinct from conventional activation functions such as the ReLU, Swish embodies a smooth, non-monotonic attribute that delineates it as an advanced alternative for neural network models. Its mathematical formulation as $f(x) = x \cdot \sigma(\beta x)$, wherein σ represents the sigmoid function, not only augments the networks' capacity to model complex datasets but also streamlines the training process by alleviating challenges linked to gradient-based optimization. The body of empirical evidence substantiating Swish's efficacy is substantial, with significant enhancements in performance observed in demanding image classification tasks, notably those assessed on the ImageNet dataset.

This activation function's adeptness in identifying and utilizing intricate patterns and relationships within data structures establishes it as a potent option for improving both the predictive accuracy and generalization aptitudes of modern deep learning architectures. Swish's refined ability to surpass the confines of monotonic activation functions renders it an essential element in our experimental investigation, aimed at advancing beyond the traditional ReLU capabilities.

2. The Gaussian Error Linear Unit (GELU), as described in [186], constitutes a smooth, non-monotonic activation function that is increasingly recognized in the domains of machine learning and AI. Its distinctive operational mechanism within computational architectures involves the weighting of input signals based on their inherent values, thereby creating a gradient landscape enriched with informational content compared to that employed by the ReLU.

   This enhanced propagation of gradients endows GELU with the capability to excel across a wide array of deep learning paradigms. GELU exhibits notable efficacy within NLP frameworks, as exemplified by models such as the BERT. This efficacy stems from GELU's capability to capture detailed semantic and syntactic features of language which are crucial for advanced NLP tasks. Thus, the preference for GELU as the activation function of choice is substantiated by its superior performance metrics relative to ReLU within these rigorous contexts.

3. The Exponential Linear Unit (ELU) [187] constitutes a notable advancement within the domain of deep learning, yielding significant improvements in performance metrics. These enhancements are especially evident in addressing issues related to vanishing gradients and biases associated with activations. The ELU function is distinguished by its smooth and non-monotonic nature, setting it apart from traditional activation functions such as the ReLU. It possesses the advantageous properties of ReLU for positive inputs while integrating a uniquely formulated exponential curvature for managing negative inputs. This dual nature is regulated by a parametric variable α, which permits mechanistic control over the saturation characteristics of the function.

In the broader context of activation functions employed in deep neural networks, the design of ELU arises as a strategic amalgamation of linear and exponential components, facilitating a more adaptive response to the varied input distributions encountered during training. In contrast to its ReLU counterpart, which remains strictly linear for positive values and zero for negatives, the exponential decay characteristic of ELU in the negative domain inhibits the deactivation of neurons, thereby addressing the 'dying ReLU' issue while maintaining simplicity and computational efficiency. As a result, this innovative formulation of ELU not only accelerates convergence rates but also fosters the consistent acquisition of robust hierarchical feature representations. Thus, in sophisticated AI architectures, particularly those that encompass expansive parameter spaces and complex data patterns, ELU emerges as an optimally balanced activation function, offering both theoretical sophistication and empirical effectiveness in high-performance neural computation tasks.

4. In the domain of AI, particularly within the framework of deep learning and neural network architectures, the employment of advanced activation functions is crucial for optimizing performance and fostering the development of sophisticated models. In this context, the Leaky Rectified Linear Unit (Leaky ReLU) [188] and its derivative, the Parametric ReLU (PReLU) [189], have been strategically selected owing to their capacity to maintain non-zero gradients even for negative input values. This is a deliberate methodological choice aimed at addressing the pervasive issue known as the 'dying ReLU' problem, a condition where neurons become inactive due to the nature of conventional ReLU activation, which renders zero output for negative inputs. Unlike the traditional ReLU, which may lead to sparse gradients and suboptimal learning in such negative regions, both Leaky ReLU and PReLU afford enhanced gradient propagation, thereby mitigating the risk of neuron inactivation.

   The Leaky ReLU adopts a predetermined, fixed slope for negative inputs, enabling improved gradient flow during backpropagation—albeit at the expense of introducing an additional layer of structural complexity. This characteristic allows

Leaky ReLU to ensure continuity and gradual transitions in the learning process, preventing abrupt disruptions that might otherwise hinder convergence. Conversely, PReLU offers an adaptive advantage by dynamically learning the slope coefficient as part of the training process itself. By endogenously calibrating the slope's steepness throughout model refinement, PReLU provides a dynamic learning capability that optimizes the activation function's responsiveness to varied and complex data inputs. This adaptability potentially leads to superior model performances, although it incurs a computational cost associated with optimizing an additional parameter. Thus, the choice to incorporate Leaky ReLU and PReLU is a testament to the pursuit of a balance between structural simplicity and the efficacy of intricate learning paradigms within the neural network.

5. Within the domain of neural network architectures, the majority of activation functions typically exhibit monotonic behavior, thereby restricting their capacity to effectively model complex and periodic phenomena characterized by non-monotonic attributes. To address this limitation, our research has elected to employ the sine function as the activation mechanism. This decision is predicated on the sine function's intrinsic oscillatory qualities, which render it exceptionally competent in capturing rhythmic fluctuations and cyclical patterns inherent in data derived from periodic inputs. By exploiting these inherent periodicities, neural networks utilizing sine-based activation functions may potentially demonstrate superior performance in domains where other activation functions, such as ReLU, may be inadequate due to their linear characteristics and insufficient modulation ability.

   Despite its advantages in specific contexts, it is imperative to acknowledge that the sine function and its associated variations have not achieved the same level of acceptance or prevalence as more traditional, monotonic functions, such as the ReLU, or other canonical activation functions within standard machine learning practices. Consequently, although the sine function may not be as widely recognized or implemented within conventional neural network frameworks, its application in our research is grounded in the hypothesis that its non-linear, pe-

riodic properties could enhance model performance in complex, non-monotonic modelling scenarios that necessitate the capture of dynamic and fluctuating data patterns.

6. Chebyshev polynomials represent a class of orthogonal polynomials of critical importance, distinguished by their extensive analytical properties and applications in diverse fields, including scientific computing, numerical analysis, and approximation theory. Their unique structure is defined by a trigonometric formula involving cosine functions, which imparts them with distinctive attributes conducive to both theoretical exploration and practical application. Consistent with the insights presented in Chapter 2, empirical research, such as the study by [119], has demonstrated that Chebyshev polynomials significantly enhance the predictive accuracy of models applied to the dynamics of real-world atmospheric phenomena.

   This empirical evidence highlights their potential as a valuable methodological resource, making them a promising prospect for enhancing our predictive modelling framework by capitalizing on their inherent ability to approximate complex functions effectively and precisely.

7. Learnable polynomials represent a class of polynomial equations characterized by dynamically adjustable constants, which undergo optimization during the training phase of a computational model. This adaptability permits the precise calibration of the polynomial equation to address the complexities inherent in the modelled phenomenon. Recent academic inquiries, including the investigation by Bilonoh et al. (2022) [118], have systematically evaluated the efficacy of these constructs. Their research highlights the considerable potential of learnable polynomials in achieving highly accurate approximations of complex functions within predictive modelling frameworks.

   Consequently, these adaptive polynomial constructs present themselves as a promising candidate for integration into our experimental methodologies. The capability to alter their coefficients through iterative learning processes further enhances

their suitability as a methodological resource, thereby augmenting the robustness and effectiveness of our experimental configurations and heightening the overall precision and reliability of the modelling outcomes.

5.2.4 Training process and Evaluation

Our comprehensive training protocol is systematically partitioned into two symmetric phases of training repetition. In the initial phase, the learnable parameters of BART are meticulously constrained, effectively immobilizing its weights. Simultaneously, the loss generated by BART, aligning with its standard loss function, is utilized as a supervisory signal to explicitly optimize the parameters of the Transformer Encoder. This critical stage enables the extraction of salient, high-dimensional features from the neural architecture of the Transformer Encoder, which subsequently channel these features as inputs into the BART framework. Consequently, this mechanism equips the Transformer Encoder with the capability to proficiently partition and synchronize EEG signal characteristics with the linguistically structured information processed by BART. The foundational training phase is imperative for ensuring that the Transformer Encoder acquires comprehensive, semantically rich representations of the EEG signal data; this serves as an optimal prerequisite for subsequent model evolutions.

Progressing to the second training iteration, the methodology remains consistent; however, fine-tuning of BART is initiated concurrently with the Transformer Encoder. This crucial development allows the adaptable weights of BART to undergo modification, thereby optimizing them during this dual-phase training schema. The rationale for orchestrating these dualistic training phases is twofold: the preliminary phase leverages the inherent capabilities of BART as an informative scaffold, augmenting the Transformer Encoder's capacity to effectively integrate EEG-derived signals with text-based constructs. Upon successful alignment, the subsequent phase focuses on the bespoke fine-tuning of BART, thereby enhancing its proficiency in serving as a conduit for EEG-to-text translation in downstream applications.

Upon the completion of the comprehensive training regimen, the model is utilized to generate textual output from BART, employing the advanced beam search algo-

rithm. The syntactic and semantic integrity of the resulting textual sequences is then meticulously evaluated through comparative analysis with reference sentences. Evaluation metrics, such as BLEU scores and ROUGE scores, are calculated to quantitatively assess the fidelity and coherence of the generated text relative to the source material.

In our experimental study, we systematically trained the machine learning model using each specific activation function over the course of 10 epochs in individual and separate runs, resulting in an aggregate total of 20 epochs across the training sessions. Our approach employed a consistent batch size of 16 in conjunction with a finely-tuned learning rate set at 5×10^{-5} across both trial scenarios. To rigorously evaluate the performance and generalizability of the constructed model, we combined the entire dataset collected from each subject, subsequently executing a randomized division into training, validation, and test subsets according to an 80% training sample, 10% validation, and 10% testing cohort for balanced evaluation.

Throughout both phases of model refinement, we employed the stochastic gradient descent algorithm with an unvarying learning rate of 5×10^{-5} to ensure convergence and optimization of the model parameters.

5.2.5 Evaluation Metrics

Within our extensive evaluation framework, we adopt a dual-metric methodology to thoroughly appraise the efficacy of our models. The primary metric employed is the BLEU score, an acronym for BLEU, as delineated in the foundational research by Siddhad, Gupta, Dogra, and Roy (2024) [147].

The BLEU score is a well-established quantitative metric frequently utilized in the evaluation of outputs from machine translation systems, which constitute essential components of NLP tasks. This metric precisely assesses the fidelity of generated outputs in comparison to one or more predefined reference translations. The BLEU score is depicted as a continuum ranging from 0 to 1, with scores approaching 1 indicating a closer alignment with the references, thus signifying high-quality translations. Such a metric facilitates the objective assessment of translation accuracy and is instrumental in identifying opportunities for enhancement in the quality of machine-generated text.

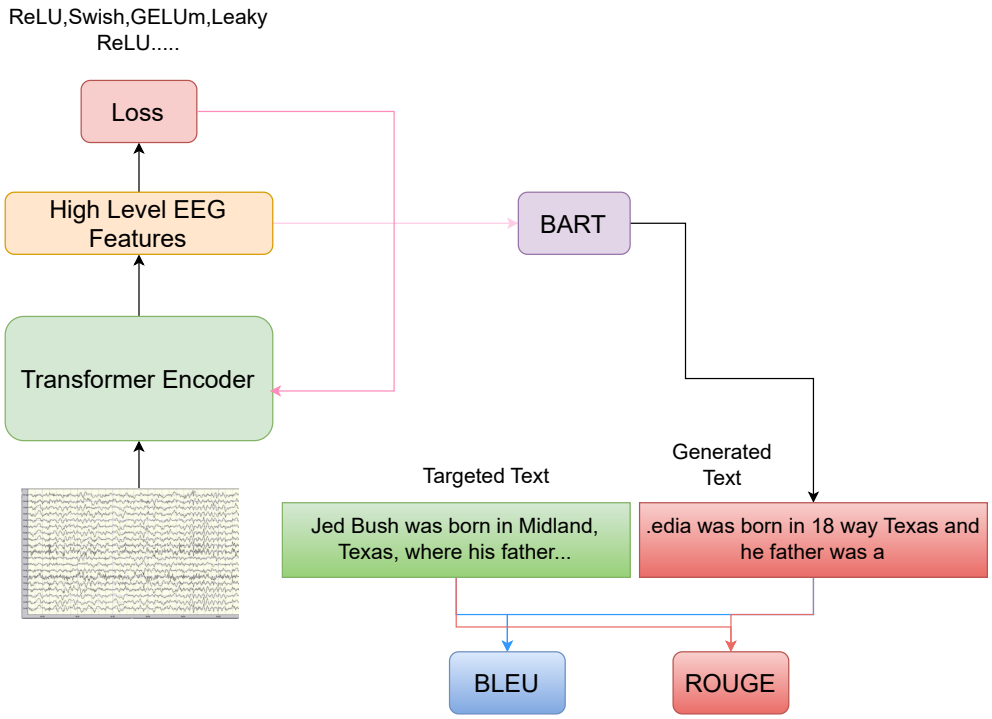


Figure 5.1: This figure shows the flow used to fine-tune BART alongside a transformer encoder and generate text from brain data.

In the domain of computational linguistics and NLP, the BLEU metric is widely regarded as an essential benchmark for evaluating the performance accuracy of machine-generated translations and text generation tasks. BLEU measures the precision of n-grams, which are contiguous sequences of 'n' items (typically words or tokens) extracted from a text sample, thereby quantifying the extent to which these n-grams in the generated text align with those in the reference text. In practical applications, n-grams of sizes ranging from 1 (unigram) to 4 (quartgram) are predominantly utilized, as they offer a comprehensive representation of text coherence across various linguistic levels.

In conjunction with BLEU, the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) metric is employed. ROUGE comprises a set of evaluation metrics specifically crafted to navigate the complexities involved in assessing the quality of summaries and translations produced by NLP models. Differing from BLEU, which emphasizes precision, ROUGE primarily focuses on recall and the degree of n-gram overlap between

the generated and reference textual contents. ROUGE metrics are particularly valued in tasks related to summarization and other forms of text-generation evaluations due to their nuanced capacity to capture semantic similarities and the presence of content. The ROUGE score itself includes various iterations, each tailored to address different text evaluation requirements. In our research, we utilize ROUGE-N, which is adept at quantifying the overlap of n-grams between the generated content and reference materials. This involves calculating unigrams (ROUGE-1) and bigrams (ROUGE-2), thereby revealing both individual word matches and paired word sequences.

Furthermore, we incorporate ROUGE-L, a metric that assesses the Longest Common Subsequence (LCS) present between the generated and reference texts. By doing so, ROUGE-L encapsulates not only the presence of n-grams but also the sequential order and relational dynamics of the words in the text, thus presenting a more comprehensive view of content alignment. [190]

5.3 Results

In our thorough evaluation of model performance, internationally respected evaluation metrics, specifically the BLEU score and the ROUGE score, were employed. These metrics have achieved extensive recognition in the domains of machine translation and question-answering tasks due to their effectiveness in measuring alignment between machine-generated text and reference materials.

Our investigative strategy was further distinguished by a comparative analysis with an established pipeline as described in [145], wherein our resulting data visualizations consistently refer to this reference configuration under the label "plain". In addition we compared our results to random baseline denotated as "random baseline" in our results table.

Moreover, not withstanding the considerable challenges presented by inter-subject variability and noise within EEG data, a decision was made to amalgamate data across all subjects. This choice was informed by insights from previous investigations [119], which indicated that an enriched training dataset correlates with improved model performance outcomes. We scrupulously aligned our experimental parameters with those

from preceding studies to ensure the highest possible fidelity in our comparative analysis.

Upon scrutiny of Table 5.2, it becomes apparent that configurations such as "torch poly 3" and "head same as layers" consistently demonstrated superior BLEU scores, thus positioning them as leading configurations in translation quality, with scores converging in the 0.09 to 0.10 range. In contrast, configurations incorporating custom activation functions, such as "use custom activation Chebysev degree 3" and "use swish activation function," yielded comparatively lower BLEU scores, predominantly falling below the 0.08 threshold. This notable variability highlights the differing efficacy levels across various configuration paradigms.

A comprehensive analysis of the preceding results has led to the identification of several significant observations that merit further scientific investigation. Firstly, a notable discrepancy was observed in the performance metrics between the Chebyshev and Torch polynomials, although both utilize equivalent 3rd-degree polynomial frameworks. This discrepancy is hypothesized to be due to the presence of learnable parameters within the Torch polynomial. By enabling the parameters of the activation functions to be learnable, it is apparent that the model's performance can be considerably enhanced.

Additionally, configurations described as 'head identical to layers' gain prominence when combined synergistically with a custom activation function. The 'head is the same as layers' configuration pertains to a Transformer Encoder where the number of attention heads equals the number of its layers. Implementing such a transformative approach reveals that the learnable activation function surpasses the performance of traditional configurations, highlighting the essential role of learnable activation functions in improving model accuracy. However, these findings are predominantly evident when the BLEU score is assessed using a 1-gram evaluation method.

This indicates that, while our model demonstrates competence in recognizing individual words within sentences, it tends to predict their positions inaccurately. Furthermore, a detailed analysis of BLEU scores for 2-gram and 3-gram setups suggests the superior performance of the 'head identical to layers' configuration, with the leaky ReLU activation function being the leading evaluative metric. Additionally, an im-

Activation Function	Rouge Score 1		Rouge Score 2	
	P±(STD)	R±(STD)	P±(STD)	R±(STD)
Random Baseline	0.05±0.01	0.05±0.01	0.05±0.01	0.05±0.01
plain	0.04±0.007	0.009±0.007	0.000±0.0	0.000±0.0
swish	0.051±0.007	0.013±0.007	0.000±0.0	0.000±0.0
gelu	0.087±0.006	0.060±0.006	0.0026±0.0001	0.0023±0.0001
elu	0.156±0.007	0.046±0.007	0.0037±0.0005	0.0016±0.0004
leaky relu	0.155±0.007	**0.115±0.007**	**0.0142±0.007**	**0.0124±0.007**
swish norm first	0.018±0.001	0.013±0.001	0.0002±0.00007	0.0001±0.0001
parametric relu	0.158±0.007	0.114±0.001	0.0102±0.0004	0.0065±0.0003
sine	0.009±0.001	0.008±0.001	0.0002±0.00001	0.0001±0.00001
chebysev degree 3	0.170±0.007	0.053±0.007	0.0070±0.0001	0.0028±0.0001
chebysev degree 3 norm first	0.018±0.007	0.013±0.007	0.0002±0.00007	0.0001±0.0
chebysev degree 2	0.136±0.007	0.048±0.007	0.0007±0.0001	0.0002±0.0001
torch poly 2	0.106±0.007	0.041±0.007	0.0005±0.0001	0.0002±0.0003
torch poly 3	**0.173±0.007**	0.06±0.007	0.0055±0.0002	0.0024±0.0002
negative positive poly	0.142±0.007	0.044±0.007	0.0027±0.0002	0.0010±0.0002
negative positive poly norm first	0.018±0.007	0.013±0.007	0.0002±0.00001	0.0001±0.00002

Table 5.1: Rouge Scores results table

portant observation was made: modifying the position of the normalization layer to precede the encoder leads to a reduction in performance, diminishing it to a tenth of its previous efficiency, irrespective of the activation function used. As depicted in Table 5.2, applying the Chebyshev polynomial with a 3rd-degree along with a prior normalization layer results in a decrease in performance from 0.087 to just 0.008. A similar reduction pattern is observed with the use of a negative-positive polynomial configuration.

Activation Function	BLEU 1-Gram ±(STD)	BLEU 2-Gram ±(STD)
Random Baseline	0.05±0.0	0.005±0.0
Plain	0.07±0.005	0.009±0.005
Neg-Positive Poly Norm	0.009±0.001	0.001±0.0001
Neg-Positive Poly	0.08±0.005	0.015±0.0001
Torch Poly Degree 3	**0.1±0.005**	0.02±0.0001
Torch Poly Degree 2	0.080±0.005	0.017±0.0005
Chebysev Poly Degree 3 Norm	0.009±0.007	0.001±0.0001
Chebysev Poly Degree 2	0.09±0.001	0.003±0.0001
Chebysev Poly Degree 3	0.09±0.005	0.002±0.01
Sine	0.004±0.005	0.001±0.0001
Parametric ReLu	0.07±0.007	0.02±0.007
Swish Norm First	0.008±0.007	0.001±0.0007
Leaky ReLu	0.09±0.005	**0.03±0.007**
ELU	0.08±0.005	0.017±0.001
GELU	0.04±0.005	0.009±0.001
Swish	0.07±0.005	0.006±0.0001
Head Same as Layers	0.09±0.005	0.03±0.005
Norm First	0.008±0.001	0.001±0.0005
Half Layers	0.03±0.001	0.001±0.0005

Table 5.2: In this table we present our BLEU score results for 1-Gram and 2-Gram BLEU configuration.

To rigorously evaluate the performance and efficacy of our computational models, we have integrated the ROUGE-score into our assessment metrics. The results of these evaluations are comprehensively documented and presented in Table 5.1. In Table 5.1 we have only included our precision and recall for ROUGE-1 and ROUGE-2 for

simplicity. A more complete table with all our results can be found in the Appendix Chapter A containing all our results. Similarly, we observe parallel trends concerning the BLEU-score. Although ROUGE is widely regarded as a metric with a strong emphasis on recall, our analysis indicates that the application of ROUGE-1 metrics demonstrates that optimal performance is achieved when utilizing either a learnable activation function or a homogeneous configuration for the head as compared to the layers. However, when the evaluative focus shifts to 2-grams, the findings suggest that the leaky ReLU activation function proves to be the most effective.

To ensure statistical significance of our results and the validity of our results we conducted a pairwise t-test on all our approaches against the random baseline and the plain baseline and calculated the standard deviation(STD) accross 20 runs for each method. Moreover the results presented is the calculated average accross our different runs.

In both our Tables (5.2 & 5.1) our STD values range from 0.0001 to 0.001 with the most dominant value to be 0.007. Since the value S calculated for the STD is $S \leq 0.01$ which typically this indicates a low variance accross different runs we showed that there is a consistency accross the performance for the different activation functions.

Furthermore, we performed a pair-wise t-test to assess the statistical significance of our top-performing methods in comparison to both random and plain baselines. The methods exhibiting the highest performance were: 'Torch Polynomial Degree 3', 'Torch Polynomial Degree 2', 'Chebyshev Polynomial Degree 2', 'Chebyshev Polynomial Degree 3', and 'Leaky ReLU'. Each of these methodologies demonstrated a p-value of $p \leq 0.05$ when evaluated against the random baseline across all assessment metrics, signalling statistical significance relative to the random baseline. Additionally, when compared to the plain baseline, each of our best approaches showed a p-value of $p \leq 0.05$ for all 1-Gram metrics and the longest metric for Rouge-L. Furthermore, for the 2-Gram metrics (BLEU-2, Rouge-2), the second-degree polynomial approaches ('Torch Polynomial Degree 2', 'Chebyshev Polynomial Degree 2') and Leaky ReLU exhibited a p-value of $p \leq 0.07$, indicating a potential trend towards statistical significance since this value approaches 0.05. This observation is reinforced as the third-degree polynomial meth-

ods show statistical significance with respect to the 2-Gram metrics, suggesting that higher polynomial degrees correlate with improved activation performance, statistically significant against the plain baseline.

5.4 Chapter Summary

In this study, a comprehensive investigation was undertaken into a range of activation functions, alongside two distinct transformation methodologies, with the aim of enhancing the current effectiveness of an EEG-to-Text decoder model that relies on the analysis of EEG data. Our empirical findings reveal the improved performance of these models in real-world contexts when subjected to alternative activation functions, thus demonstrating a substantive enhancement in the decoder's capabilities. The study accentuates the potential benefits inherent in the use of learnable activation functions. Notably, the Leaky ReLU activation function exhibited superior performance metrics in the context of 2-gram sequences and higher, as assessed by ROUGE and BLEU scores, indicating a robust model output in these areas.

In contrast, the learnable third-degree polynomial function showed a tendency for enhanced performance in 1-gram evaluations, suggesting its potential applicability in specific modelling scenarios. It is hypothesized that employing a diverse array of activation functions in the training of such models constitutes a promising, yet inadequately explored, research area. Our findings underscore the need for increased scholarly attention in this direction. Although the outcomes derived from the EEG-to-Text decoding process offer promising insights, it remains clear that achieving an applied, accurate, real-time model capable of generating text from raw neurological data requires further exploration. Initial investigations indicate the viability of brain-derived text generation, establishing a preliminary basis for future research undertakings.

A salient constraint identified is the limited frequency of sentence repetition within the current dataset, whereby each subject encounters each sentence only once or twice. This deficiency may adversely affect the learning dynamics of deep learning models, which benefit from recurrent exposure to samples. It is posited that increasing the frequency of repetition within the EEG dataset could substantially enhance performance

outcomes. Given the inherently noisy characteristics of EEG data, it is hypothesized that the model's capacity for representation learning is obstructed by an inability to adequately model the noise, potentially due to its simplistic architectural design. Hence, the deployment of a more sophisticated encoder, designed to extract high-order EEG features before their integration into the extensive language model, represents a potential enhancement strategy. Finally, due to temporal constraints, experimentation with the activation function was limited to a single random seed. Therefore, future research will involve multi-seed testing to evaluate the impact of randomness on the reported performance metrics.

Chapter 6

Development of a Generic Brain-to-Text Decoding Module

6.1 Introduction

The intricate workings of the human brain have engaged the interest of the scientific community for numerous decades, precipitating extensive research and exploration within the field of neuroscience [191–195]. Historically, the endeavor to elucidate the mysteries of brain function has been propelled by the fundamental aspiration to comprehend the complex cognitive processes that characterize the human species. Recent technological advancements have profoundly enhanced our capacity to examine the living brain in a non-invasive manner. Techniques such as EEG, Magnetic Resonance Imaging (MRI), and MEG have emerged as pioneering methodologies facilitating the visualization and quantification of brain activity in real time. These methodologies have unveiled novel dimensions in the sphere of cognitive neuroscience, permitting the analysis of neural substrates associated with cognitive and emotional functions.

In the early phases of brain research, investigators employed relatively rudimentary experimental paradigms to elicit discernible patterns of neural activation. For instance, studies often utilized straightforward binary conditions, such as exposing subjects to emotionally charged scenes (e.g., happy versus sad imagery), thereby facilitating the examination of the neural correlates of emotional processing. These early experiments

established a foundational understanding of emotional representation within the cerebral architecture. As scientific inquiry matured, researchers embarked on more sophisticated experiments intended to elucidate the nuanced functions of specific cerebral regions, such as the Superior Temporal Sulcus and the Human Insula. Studies conducted by Hein et al. (2008) and Chang et al. (2013) exemplify this trajectory [196, 197]. Their findings underscored the multifaceted nature of brain regions, highlighting both the diversity of functions they serve and the intricate web of interconnectivity among neural networks. This paradigm shift paved the way for a deeper appreciation of the brain's capacity to integrate multifarious stimuli, thereby facilitating complex processes such as audiovisual integration and linguistic comprehension.

Among the notable endeavours in contemporary neuroscience is the pursuit to elucidate the cerebral mechanisms by which the human brain interprets natural language and semantic constructs. Attaining a thorough comprehension of these processes promises not only to enhance theoretical models of language processing but also to provide profound insights into the evolutionary progression of the human brain across millennia. The emerging field of neurolinguistics seeks to bridge this knowledge gap by analysing the neural foundations of language comprehension, thereby contributing to a more profound understanding of both language and cognition within the neural context. The implications of effectively mapping these cognitive pathways are extensive, carrying the potential to revolutionize educational, clinical, and computational paradigms. Within this dynamic research landscape, the pursuit to unravel the complexities of human cognition and language understanding persists, charting a course toward a more profound understanding of the human condition.

Prior research within the realm of neurolinguistics has extensively depended on the acquisition of empirical data via techniques such as fMRI. As noted in scholarly studies [1,37,60,198], fMRI is acknowledged for its limitations, including its slow speed and high cost, and it does not support real-time data processing when compared with other neuroimaging methodologies. Recent scholarly efforts aim to elucidate the cerebral representation of information needs [73, 80]. Information needs are intrinsically linked to the discipline of neurolinguistics, primarily due to their crucial role as driving forces be-

hind the execution of information retrieval operations. A dominant perspective suggests that artificial intelligence frameworks are ineffective in capturing the authentic semantic essence, consequently often failing to provide users with the desired informational content. Previous approaches have been restricted to merely classificatory functions, concentrating on specific areas such as classifying information needs [134], assessing mental workload [89,90], engaging in reading tasks [143], or imagining categories [199], among others.

To the best of our knowledge, this study constitutes a foundational exploration into the application of raw EEG signals for the precise identification of linguistic components and the generation of sentences as direct cerebral outputs. The significance of our methodology is underscored by the advantageous attributes of EEG, which encompass real-time data acquisition, cost efficiency, and operational simplicity, especially in comparison to fMRI. By decoding brain activity using unprocessed EEG signals, our research endeavours to accelerate the understanding of neural functions, offering a method that is both more rapid and methodologically rigorous. Furthermore, within the current academic discourse, our study is established as a pioneering initiative to utilize raw EEG signals as the fundamental input for our computational model intended for decoding purposes.

The following sections of this work are structured as follows, Section 6.2 discusses the methodology of the research, Section 6.3 highlights the results of our investigation, and lastly, Section 6.4 a summary of the current Chapter is presented.

6.2 Methodology

6.2.1 Introduction

This section delineates the systematic methodology employed in this study to facilitate the successful training of the brain-to-text decoder. It elaborates on the research design and the analytical procedures implemented. Additionally, it presents a comprehensive overview of the machine learning methodologies and tools employed, along with the justification for their selection based on the results obtained from each integration.

Furthermore, it provides a succinct overview of the two datasets utilized, as well as the reasoning behind their selection.

Each experimental concept and instrument employed underwent multiple iterations to ensure uniformity across results. A detailed exposition of our findings is provided in Section 6.3, accompanied by the standard deviation(STD) and statistical significance tests for each method. In the methodology section, we provide only a succinct overview of the results to justify the inclusion and transition from one method to another.

6.2.2 Data

In the process of the development and refinement of our Brain Decoder system, we adopted a strategy involving the training on two distinct datasets of cerebral data. This dual-data strategy was pivotal in achieving a comprehensive understanding of neural activities across diverse experimental protocols. Initially, we employed publicly accessible EEG datasets, ZuCo 1.0 [140] and ZuCo 2.0 [141] provided by the University of Zurich. These EEG datasets constituted the foundational basis for our preliminary analyses and facilitated the development of an initial model of neural activity patterns, which served as the cornerstone for our subsequent research endeavours. Building upon the foundational insights derived from the EEG data, we elected to expand the scope of our research by integrating a second dataset type, which consisted of intra-cortical microelectrode arrays data. This supplementary dataset was provisioned and made available through the research conducted by Willett et al. [59].

The incorporation of the microelectrode array datasets enables the acquisition of neural signals with markedly superior spatial and temporal resolution in comparison to EEG data, thus facilitating more profound insights into the cortical processing mechanics on a microscopic level. The rationale for the utilization of these two distinct datasets is comprehensively explored in Section 6.2.2. This section presents a thorough narrative that substantiates our choice of these particular datasets. It provides a succinct description of each dataset, clarifies their unique attributes, and specifies the particular segments of these datasets that were employed as training data to refine and augment the capabilities of the Brain Decoder.

EEG

This comprehensive investigation leverages the highly regarded and publicly accessible Zurich Cognitive Language Processing Corpus, specifically the ZuCo 1.0 [140] and ZuCo 2.0 datasets [141]. These datasets are distinguished by their inclusion of EEG data in conjunction with eye-tracking records. These records are meticulously gathered from a robust sample of participants.ZuCo 1.0 encompasses data from 12 native English-speaking healthy participants, while ZuCo 2.0 includes data from 18 participants, culminating in a collective sample of 30 individuals. Furthermore, ZuCo 1.0 comprises 1,107 sentences, and ZuCo 2.0 consists of 739 sentences, resulting in an aggregate of 1,846 sentences across both datasets. By leveraging these datasets, the model was trained on approximately 30,000 words, each of which is accompanied by corresponding EEG recordings. In both datasets the participants systematically engaged in both Normal Reading (NR) and Task-Specific Reading (TSR) activities for about 4-6 hours. The content for these activities has been carefully curated, focusing on the in-depth analysis of movie reviews and informative Wikipedia articles. In Normal Reading (NR) activities, participants engaged in reading the text devoid of any associated tasks. Conversely, Task-Specific Reading (TSR) required participants to perform a semantic relation annotation task concurrently with reading the text. This methodological approach facilitates the capture of neural and ocular signals indicative of goal-directed language processing, thereby enhancing the dataset's utility for investigating profound semantic comprehension and cognitive load during reading. Additionally, it allows for the development of more robust models capable of detecting task engagement through physiological data.

A crucial characteristic of the ZuCo datasets is the exact temporal alignment of EEG data with text-based stimuli. This alignment is accomplished through meticulous monitoring of fixation points, which are precisely recorded by advanced eye-tracking technology. The datasets comprise an extensive range of EEG features that are closely linked with specific eye-tracking metrics. These metrics include, among others, First Fixation Duration (FFD), which denotes the duration of initial fixations on a particular text element.

Furthermore, the datasets offer insights into total reading time (TRT), which encompasses the cumulative duration of all fixations on a specific text, as well as gaze duration (GD), which quantifies the total time spent in first-pass reading before any regressions occur. The single first fixation (SFD) is another parameter of interest, representing the isolated duration of an initial fixation event on an object, while go-past time (GPT) is essential for examining backward eye movements and re-reading patterns in reading tasks.

The ZuCo corpus, characterized by its extensive dataset composition, represents an essential asset for the investigation of cognitive language processing. By integrating EEG and eye-tracking data, it enables researchers to examine the intricate dynamics of reading behaviour and neural processing. This amalgamation of data enhances comprehension of the physiological and psychological dimensions of reading, thereby advancing developments in brain decoding technologies and cognitive neuroscience. Eye tracking played a crucial role in the experimental setup, as the sentences were presented to the participants in their entirety. Utilizing eye tracking data enabled the determination of the precise interval during which a word was read, allowing for the corresponding EEG recording to be aligned with this word.

Within the scope of this scholarly investigation, the comprehensive exploration of raw EEG data at the sentence level has been undertaken. This methodological approach is underpinned by several compelling justifications. Chiefly, sentence-level data demonstrates an alignment with the structure of conventional speech-to-text datasets, a consideration of significant importance as thoroughly expounded in the Methodology section of this dissertation. This section rigorously elucidates the manner in which the architecture of the Brain Decoder system is inherently influenced by the paradigms of established speech-to-text and ASR frameworks, thus rendering sentence-level data especially advantageous and relevant.

Moreover, a comprehensive analysis of the dataset uncovered significant inconsistencies at the word level. It was apparent that while raw EEG data were recorded for some words, such recordings were intermittently missing for others. This inconsistency exhibited variability among different subjects. The authors of the original study offer

an explanation, stating that if the time required by a subject to read a particular word was shorter than the EEG sampling interval, it resulted in the absence of recorded data for that specific word during the pertinent fixation period.

Intra-cortical Microelectrode Arrays (IMA)

In light of the detection of inconsistencies within the ZuCo dataset, which raised significant concerns about its reliability and suitability for our research framework, an investigation was undertaken to identify an alternative dataset that satisfied the requisite parameters. The central criterion guiding our search was the need for a dataset replicating the structure of ZuCo.

The congruence of the data configuration would allow for its seamless integration with our pre-existing analytic codebase, thus enabling the advancement of our research efforts without requiring substantial alterations or the restructuring of our code. Our objective was to maintain the integrity and efficiency of our computational processes while minimizing disruption.

The dataset introduced by Willet et Al [59]. was utilised to expand our research. This dataset as mentioned in Section 2.2.3 compromises of a single subject reading natural text whilst intra cortical arrays are recording his or her brain activity. The data were gathered as the participant endeavored to express sentences prompted on a computer monitor. On each assessment day, between 260 and 480 sentences (amounting to 41 $\pm$ 3.7 minutes of data) were recorded for the purpose of training. The subject executed these tasks over multiple days, culminating in a cumulative training dataset of 10,850 sentences by the concluding day of data collection. On average, data acquisition and RNN training persisted for 140 minutes each day.

Their research discusses advancements in speech neurolinguistics, specifically focusing on assistive technologies designed for individuals with severe speech impairments due to conditions like ALS. A significant study highlights the creation of a high-performance speech BCI capable of decoding neural signals linked to speech production. This BCI achieves a word error rate of 9.1% for a vocabulary of 50 words and 23.8% for a lexicon of 125,000 words, with a communication speed of 62 words per

minute, close to natural conversation rates.

This research underscores the relevance of factors like vocabulary size, electrode density, and training data in optimizing BCI performance, advocating for microelectrode expansion and language model refinement. While challenges regarding system resilience and long-term use remain, the study signifies a substantial leap in neurolinguistics, offering new possibilities for enhancing communication for individuals with speech impairments, thus improving their quality of life.

Although the technique is invasive, it provided an opportunity to conduct a more in-depth evaluation of our model and determine whether the identified technical challenges were the cause of inadequate model training.

Data Summary

The primary motivation for the integration of Intra-cortical Microelectrode Arrays (IMA) [59] data into the analytical paradigm is largely due to the limitations inherent in the experimental framework of the ZuCo dataset. Although the ZuCo dataset is robust across various dimensions, it notably lacks repetitive data collection, an element deemed crucial for the training and progression of AI models. Repetition in learning datasets is essential for success, as it allows AI systems to repeatedly assess the same instances with slight variations. This systematic repetition enables the development of diverse representations of the same entity, thereby enhancing the AI's ability to identify and distinguish the entity amidst subtle differences.

For elucidation, consider the analogy of various feline breeds. Although individual breeds may exhibit significant variations, the overarching classification as 'cat' remains constant. In contrast, these breeds are distinctly different from canines. In ZuCo's experimental framework, subjects were not subjected to repetitive stimuli, and only a restricted number of sentences were common across different experimental tasks. To address the challenge of repetition, data were aggregated across multiple subjects to construct a comprehensive dataset. Nonetheless, conducting cross-subject data training in the field of Neuroscience is acknowledged as being highly unstable and unreliable due to the inherent variability in brain structure and function among individual subjects.

In contrast, the IMA dataset originated from a single participant, effectively eliminating complexities associated with cross-subject variability. Importantly, the experimental design required the subject to read the same sentences on five distinct occasions, thereby ensuring the necessary repetition within the dataset. Such repetitive exposure was crucial in enhancing the model's ability to discern significant patterns and insights. As a result, transitioning to this dataset was advantageous, as the data collection method conformed to previously established configurations, facilitating a seamless transition with minimal need for code alterations. This transformation enabled an efficient adaptation process and significantly reduced disruptions during implementation.

6.2.3 LLM substitution

In the preliminary experimental framework, we adopted the architectural paradigm outlined in Section 5.1. The basis of our methodology was founded upon employing a stable transformer encoder, which was enhanced by modifications in the activation function to support text generation. During the initial phases, our aim was to improve the quality of the model's output by systematically replacing the LLM employed at the conclusion of the processing pipeline. Initially, we selected BART as our core LLM, recognizing the limitations imposed by the technological advancements available at that time.

In the subsequent evolution of research on LLMs, a diverse array of advanced models have been integrated into the machine learning ecosystem. The introduction of the Large Language Model Meta AI (LLaMa) [200] by Meta AI, alongside OpenAI's subsequent releases, GPT-3.5 and GPT-4.0 [201], signified a significant transformation in the domain of NLP. These pioneering models demonstrated exceptional capabilities by surpassing existing benchmarks, thereby establishing novel paradigms of excellence.

In light of our primary focus on text generation through LLMs, the inherent capabilities of these advanced models offered a significant opportunity to improve the quality of textual outputs beyond the potential of BART. We hypothesized that the implementation of these state-of-the-art models in place of BART would result in more refined textual generation since these models were state of the art and achieve the best

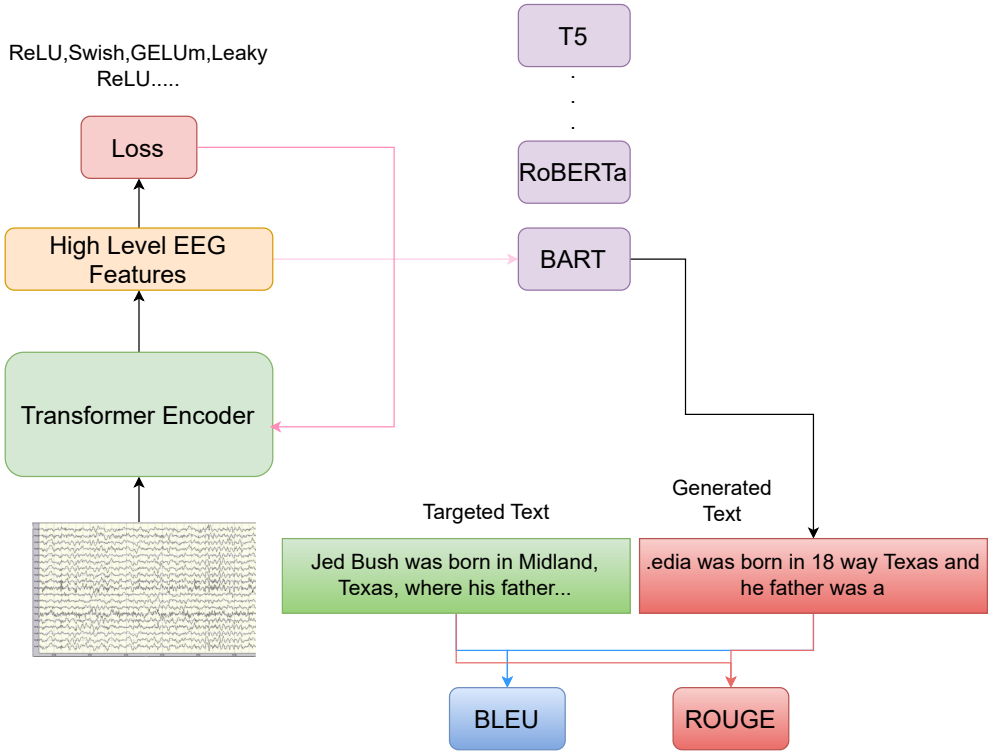


Figure 6.1: This figure shows the enhancement from the previous implementation illustrated in Figure 5.1 with the introduction of several state-of-the art LLMs

results in text generation at the time. Therefore, a systematic transition was executed to replace BART with these improved models, facilitating comprehensive documentation and analysis of our empirical observations throughout this developmental phase. A comprehensive list of all the models evaluated within our pipeline is presented in Table 6.1. As previously described, we adhered to the original pipeline outlined in Section 5.1. However, to facilitate the incorporation of the new LLMs, it was necessary to adjust the output configuration of our Transformer Encoder to align with the input configuration of the intended LLM. This adjustment was an unavoidable constraint, as the features produced by our encoder needed to conform to the dimensional requirements for effectively utilizing these LLMs as decoders.

Model Name
EleutherAI-GPT-Neo-1.3B [202]
Facebook-BlenderBot 400M [203]
Google BigBird Pegasus Large [204]
Microsoft Prophetnet Large Uncased [205]
RoBERTa [168]
T5 [206]

Table 6.1: List of Model Names Tested as an enhancement to our pipeline.

6.2.4 CTC integration

Initially, our efforts to utilize various LLMs did not demonstrate significant enhancements in the performance of the pipeline. We postulated that the absence of observed improvements may be attributed to inadequately constructed neural features derived from our Transformer Encoder. This prompted a thorough investigation aimed at uncovering potential areas for improvement, which ultimately highlighted the effectiveness of incorporating CTC loss [207] when constructing our initial Brain Encoder. CTC loss, a method predominantly employed during the training stages of speech-to-text systems, could effectively addresses the inherent characteristics and format similarities that exist between speech data, EEG, and IMA data. Each of these modalities is characterized by a wave-based structure and is segmented into time-steps as determined by sampling rates, presenting specific challenges related to their variable input and target lengths. These variations result from the heterogeneity in text sizes and discrepancies in reading or speaking durations, which are influenced by individual speeds and comprehension capacities.

The CTC loss function provides a robust methodology for addressing the intricate task of classifying unsegmented data. It is adept at mapping sequences of variable lengths into coherent outputs. Unsegmented data, which lack the explicit segmentation often exemplified in auditory character depiction where the precise timing of the spoken element is known, pose a significant challenge. The resolution of this issue involves the adjustment of the sampling rate; however, variability persists, with discrepancies occurring when characters are articulated over extended time frames, influenced by complexity and speech velocity. The task of segmenting brain data, which shares similar

complexities, is particularly challenging. Hence, the integration of CTC loss into our pipeline was considered optimal, effectively alleviating segmentation challenges. In the realm of speech-to-text tasks, as documented, CTC loss has shown exceptional performance, frequently achieving notably superior outcomes [208, 209].

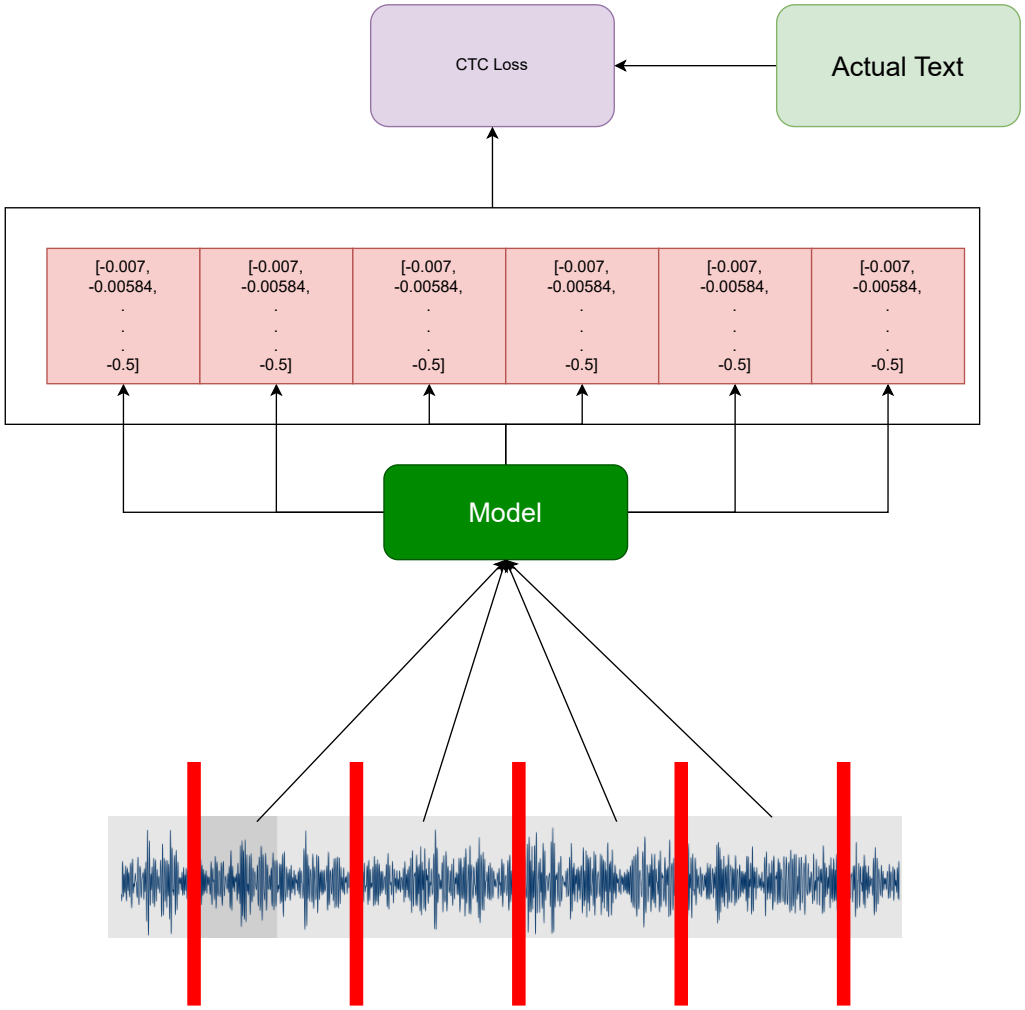


Figure 6.2: This figure illustrates how CTC loss was integrated to the pipeline with the hope of learning positional alignment of characters and brain data. For each time step the log probabilities of each character in the vocabulary were calculated and then the CTC loss was calculated between the predicted and actual sentence.

Further elucidation on this approach highlights the rationale for associating each time-step with a character. This approach simplifies the classification problem, reducing

the classification space significantly compared to word-level classification driven by extensive dictionaries. The limited scope of the English alphabet, comprising only 26 letters, constrains the classification space, facilitating the generation of any conceivable word and thereby extending the vocabulary's reach. Furthermore, in spoken discourse, a character's sound varies, and clusters of characters may coalesce into singular phonemic sounds, such as the 'ch' sound in speech. To address these nuances, phonemes were incorporated, significantly boosting the effectiveness of speech-to-text models. These observations steered the decision to tokenize our text data using either character or phonemic granularity, striving for precise classification within the temporal framework of our data. To enable this functionality, we implemented two custom tokenizers: one employing the English alphabet as its vocabulary, and the other utilizing a phonetic vocabulary [210, 211].

Considering the aforementioned characteristics, we developed a novel training pipeline for our Transformer Encoder by incorporating the CTC Loss function. Notably, the sentence representation of EEG recordings was utilized as both training and testing data. Each sentence comprised numerous EEG recording frames, serving as time steps. As illustrated in Figure 6.2, for each time step, the EEG data is fed into our Transformer Encoder, where the log probabilities of each character in our vocabulary are calculated. Our vocabulary was determined based on the chosen tokenizer (phoneme versus character), including the addition of a blank character. The CTC is pivotal, particularly when the input sequence is significantly longer than the target output, and explicit frame-wise labels are unavailable, underscoring the importance of the blank character for the input frames. Instead of requiring a fixed alignment between input frames and output labels, CTC allows the model to predict a special blank symbol alongside the target labels and establishes a many-to-one mapping that collapses consecutive repeated labels and removes blanks. Subsequently, the CTC loss computes the negative log probability of all valid alignments capable of producing the desired label sequence after this collapsing process. Formally, given an input sequence $\mathbf{x} = (x_1, x_2, \ldots, x_T)$ and a target label sequence $\mathbf{y} = (y_1, y_2, \ldots, y_U)$, where $U \leq T$, the CTC loss is mathematically defined

as: ,

$$\mathcal{L}_{\text{CTC}} = -\log\left(\sum_{\pi\in\mathcal{B}^{-1}(\mathbf{y})}\prod_{t=1}^{T} p(\pi_t|x_t)\right)$$

, where π signifies a path (a length-T sequence including blanks), $\mathcal{B}(\pi)$ represents the collapsing function that transforms paths to the final output sequence by eliminating repeated symbols and blanks, and $p(\pi_t|x_t)$ denotes the model's predicted probability of symbol π_t at time step t. This formulation facilitates end-to-end model training without necessitating manual alignment between inputs and outputs, significantly simplifying the learning process for sequential data. Upon completion of the CTC loss training, we employed a greedy decoding approach, selecting the maximum probability of each character at each time step, and subsequently passed this information to the LLM to conduct the decoding and evaluate the generated text against the ground truth using BLEU and Rouge scores. We trained our models using a batch size of 16, 100 epochs using the Adam optimizer with a learning rate l where: $0.01 \leq l \geq 0.001$.

Upon the integration of the CTC loss within the structure of our current pipeline, an enhancement in performance metrics was anticipated. Nevertheless, the expected advancement in performance did not manifest as anticipated. We believed that the initial Transformer Encoder was not able to capture good temporal and spatial features for the EEG so we replaced it with a more suitable model architecture that has exhibit notable state-of-the art results when used in time series data.The model utilized was the Conformer model, extensively documented in the literature, such as the work by Gulati et al. (2020) [2], for exhibiting exceptional results in accuracy and efficiency within speech-to-text tasks. The Conformer model is renowned for its ability to effectively manage and process sequential data inputs, dynamically adapting to the temporal variations inherent in speech signals.

The Conformer architecture synergizes CNNs and self-attention mechanisms, providing an enhanced ability to capture both local and global dependencies in acoustic signals. The design of the model integrates essential characteristics required for high-performance speech recognition systems, including scalability and robustness against perturbations and noise. Given these attributes, it was logical to foresee that the

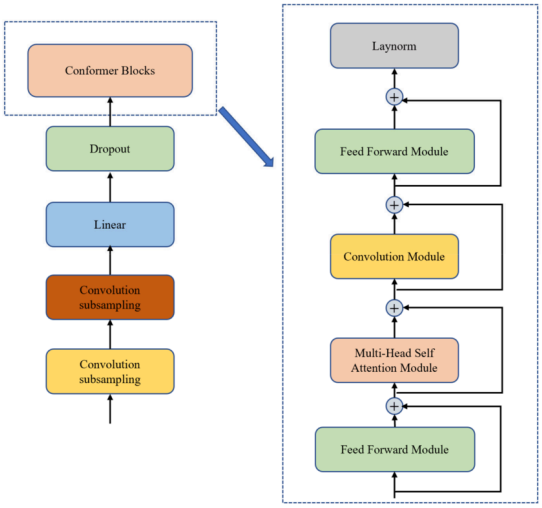


Figure 6.3: This figure shows the architecture of a Conformer Neural Network as proposed by Gulati et Al. [2] and how Convolution Layer can be integrrated with a Multi Headed Attention Layer.

augmentation of our pipeline with CTC loss would lead to significant performance improvements. However, the empirical findings of our experiments did not align with these expectations, necessitating a deeper investigation into the underlying factors hindering the anticipated performance enhancements.

6.2.5 Wav2Vec2 and Data2vec Implementation

A detailed examination of the training process across various models reveals that, in numerous instances, data is subjected to augmentation and feature extraction prior to being input into the conformer model. In the domain of speech models, features are typically derived by generating MEL spectrograms from audio files or by employing a pre-trained feature model on speech data to extract features from individual audio samples. Contrastingly, there exists a lack of standardized processing pipelines for deriving features from EEG and IMA data. The existing pipelines are not only cumbersome but also resource-intensive, necessitating significant processing power.

Wav2Vec2 [3] and Data2Vec [4], the latter being proposed as a more general implementation, offer a training framework aimed at the construction of 'encoder models' that are cable of creating task agnostic feature of a modality. These models leverage

self-supervised learning through the organization of unlabelled data in a supervised framework using contrastive loss, thereby acquiring generic features pertinent to each modality. Data2Vec has been adeptly employed across various modalities, such as text, video, and images, successfully developing potent encoders for each. Additionally, given the complex nature and advanced development of novel machine learning models, these modalities can be cohesively integrated to forge robust multimodal models capable of effectively processing a wide array of data types.

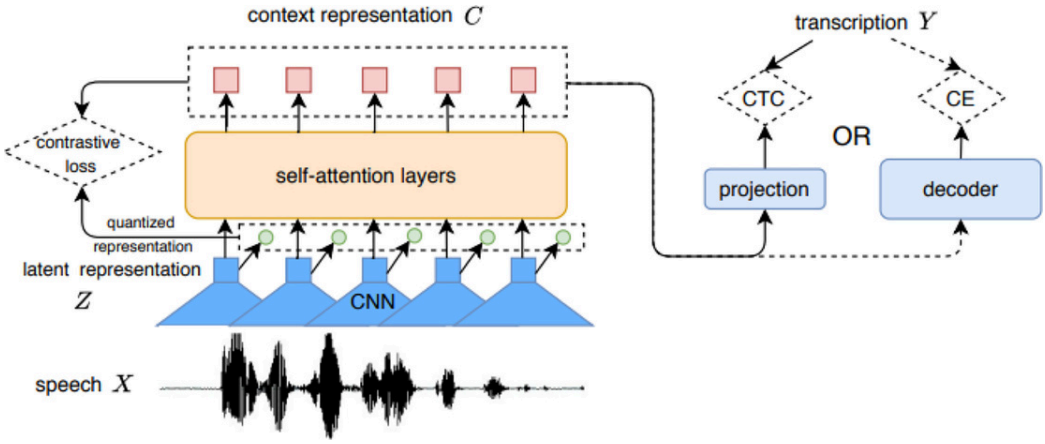


Figure 6.4: This Figure shows the proposed architecture for a Wave2Vec2 model training regime as proposed by Baevski et Al. [3]

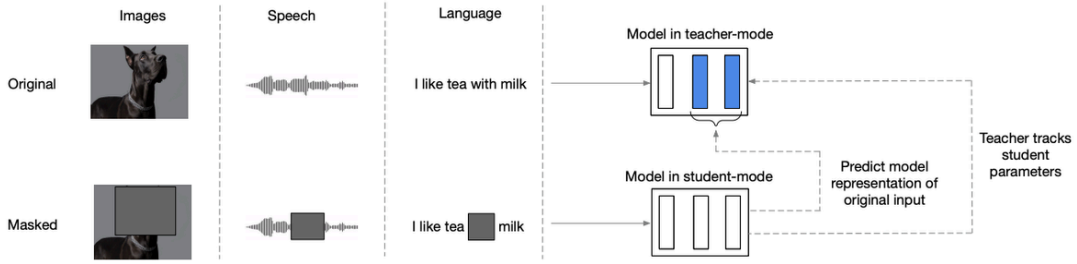


Figure 6.5: This figure illustrates the proposed architecture for training effectively a Data2Vec model as proposed by Baevski et Al. [4]

In our pursuit of developing a model that closely emulates brain-like data, we identified a substantial gap, as no existing models, to the best of our knowledge, have successfully accomplished this objective. This gap in the current research prompted us to explore the potential application of Data2Vec in the development of a truly representa-

tive brain modality. Our experimental model is primarily based on the well-established Conformer architecture, which is extensively detailed in Figure 6.3. The unique advantage of Conformer blocks resides in their hybrid structure. They astutely combine the feature extraction capabilities of CNNs with the comprehensive attention mechanisms employed in transformer architectures. This is accomplished through a strategic arrangement of self-multiheaded attention layers within each conformer block. The structural design of our model integrates multiple Conformer blocks in a tiered manner, with each tier enhancing model complexity. The flexibility in the number of blocks enabled us to methodically increase model intricacy and capture more nuanced patterns in our data. Subsequent to the arrangement of conformer blocks, the architecture transitions into a fully connected layer, culminating in a projection layer designed to output character probabilities within our designated classification framework. Here, the implementation of a Log Softmax activation function is particularly noteworthy.

Unlike the conventional Softmax, Log Softmax offers distinct advantages, especially in machine learning contexts such as neural networks. The superiority of Log Softmax is demonstrated through its contribution to numerical stability, a challenge frequently encountered due to overflow and underflow issues associated with large score values (logits) in neural computations. By employing the log-sum-exp trick, Log Softmax effectively addresses these issues, simplifying the Softmax and logarithm computations into a streamlined operation that enhances computational efficiency. This efficiency leads to faster convergence during training by improving error penalization for inaccurate predictions, thereby ensuring a more precise trajectory for gradient descent. Furthermore, the integration of Log Softmax with loss functions, particularly cross-entropy, highlights its seamless compatibility. Since these loss functions necessitate log probabilities, Log Softmax optimizes the training cycle not only by facilitating stable gradient computation, crucial for robust back-propagation, but also by efficiently adapting to extreme data values. Such attributes contribute to a balanced probability output, ultimately affirming Log Softmax as a prudent choice in numerous deep learning contexts.

We have modified our existing pipeline to integrate a Conformer-based encoder,

which will generate feature representations from the EEG and IMA data. Subsequently, these extracted features will be employed as input to our pre-existing deep learning model, which will utilize Connectionist Temporal Classification (CTC) loss to interpret the provided features.

Our methodology has evolved into a two-step training pipeline. Initially, a generic encoder is trained utilizing contrastive learning to discern generic features from brain data. The contrastive learning pipeline adheres to the principle of generating self-supervised examples through the random masking of a certain percentage of sentence time steps. Subsequently, the model endeavors to reconstruct the masked time steps. In our experimental design, we masked 15% of the time steps. Contrastive loss is employed to learn embeddings by encouraging similar input pairs to possess closer representations in the feature space, while dissimilar pairs are driven apart by at least a specified margin. It is formally defined as:

$$\mathcal{L}_{\text{contrastive}} = \frac{1}{2} y\, D^2 + \frac{1}{2}(1-y)\, \max(0, m - D)^2$$

where $y \in \{0, 1\}$ designates whether the pair is similar ($y = 1$) or dissimilar ($y = 0$), D represents the distance (commonly Euclidean) between the two embeddings, and m is a margin hyperparameter dictating the minimum desired separation between dissimilar pairs. Taking these considerations into account, we designated a positive pair as the ground truth and a negative pair as a randomly sampled EEG or IMA recording from our dataset. Our initial encoder model was trained employing a batch size of 16 for 100 epochs with a learning rate l, where: $0.01 \leq l \geq 0.001$, using the Adam optimizer.

The second stage of our training pipeline mirrors the approach detailed in Section 6.2.4. The sole distinction is that the deep learning model trained with CTC employs featurized inputs in contrast to raw EEG data. Each time step is initially featurized by the encoder and subsequently inputted into the decoding model. The underlying rationale is that the model can execute the classification task more effectively, thereby improving the CTC loss and leading to more optimal adjustments of the model's weights.

6.2.6 Bendr And EEG-Conformer Integration

In efforts to optimize and improve the pre-existing processing pipeline we hypothized that a generic brain encoder was needed for better performance by providing better brain features. Alongside our training phase in the previous section we were performing a research in parallel for any new advancments regarding the construction of a generic brain encoder. In our research came to light 2 distinct cases :Bendr [58] and EEG-Conformer [212]. As a result a substantial modification was done to the pipeline for the incorporation of two adaptable EEG encoding models: Bendr [58] and EEG-Conformer [212]. These encoders, initially crafted for the purpose of generating generalized EEG features, fulfil distinct functions: Bendr is primarily employed in the domain of EEG feature classification tasks, while EEG-Conformer strives to encapsulate both local and global features within a unified EEG classification framework.

It is crucial to recognize that neither encoding model was originally devised or intentionally crafted for the purpose of generating text directly from EEG-derived data. Nevertheless, it was postulated that with judicious fine-tuning, these models could be tailored to accommodate the distinct requirements of our particular task objectives. To enable this functional adaptation, a structural modification to the existing architecture was instituted.We substitute our training using contrastive learning to just use one of the 2 pre-trained encoders. We hypothesized that an increase in performance could indicate a faulty training regime for our encoder model.

This integration involved the strategic integration of an additional pair of layers at the final stage of the pipeline. Specifically, a fully connected layer was introduced, succeeded by a projection layer. The primary objective of this structural alteration was to proficiently compute and produce a probability distribution over prospective character outputs, thereby augmenting the model's proficiency in generating text-based representations from EEG data.

Nevertheless, even with the incorporation of these two models as replacements for our brain encoder, a significant enhancement in performance was not observed. A comprehensive discourse on our comparative analysis of the three approaches is provided in Section 6.3, wherein we delineate our conclusions and formulate our hypotheses.

6.3 Results

In assessing the performance of our models, we have adjudged it appropriate to utilize the BLEU score and ROUGE score as our principal evaluation metrics. This decision is consistent with the evaluation framework articulated in Chapter 5. Given the homogeneous nature of the task, the continued application of these metrics was deemed judicious for two compelling reasons. Primarily, as previously expounded upon in this thesis, the BLEU score and ROUGE score are esteemed among the most widely acknowledged standards within the text generation evaluation domain. Their established standing in the field substantiates our selection.

We compared each of our results against a random baseline and the current baseline set by Wang et Al. [145] using the appropriate statistical tests to ensure the validity of our results. Moreover in Table 6.2 we present our results alongside, as the average value accross runs for our different approaches, and their STD value. To obtain the STD we run our experimentation 18 times. As you can see from our results the STD S is $S \leq 0.01$ *and that signifies a consistency for our different model approaches.*

Moreover, these metrics played an integral role in the evaluation of the seminal transformer Encoder. Consequently, their continued use in assessing our models facilitates a coherent and seamless comparison of results across various model iterations. Maintaining such consistency is essential for accurately interpreting the performance improvements and the efficacy of our novel approaches within the framework of pre-established benchmarks.

Our empirical findings have unequivocally demonstrated that the mere substitution of the LLM within the extant pipeline did not yield a significant improvement in the results. This observation is consistent with the existing literature, which suggests that LLMs, when employed for the translation of various modalities into text [213–215], primarily operate as correctional mechanisms rather than transformative agents. For instance, within the realm of converting spoken language into written form, certain lexemes such as "red" and its homophonic counterpart "read" (past tense) remain phonetically indistinguishable.

Scenario	**Technique**	**EEG (BLEU±STD, ROUGE±STD)**	**IMA (BLEU±STD, ROUGE±STD)**
Baselines	Random Baseline	(0.05, 0.05)	(0.05, 0.05)
	Plain Baseline	(0.07, 0.05)	(0.07, 0.05)
CTC	CTC+Phoneme	(0.02±0.007, 0.0±0.001)	(0.01±0.001, 0.0±0.001)
	CTC+Character	(0.09±0.01, 0.0±0.0)	(0.07±0.02, 0.001±0.01)
Generic Algorithms	Data2Vec+Phoneme	(0.02±0.001, 0.0005±0.0003)	(0.008±0.001, 0.0005±0.003)
	Data2Vec+Character	(0.1±0.01, 0.02±0.007)	(0.05±0.01, 0.02±0.01)
	Wav2Vec2+Phoneme	(0.02±0.01, 0.0002±0.00001)	(0.008±0.0001, 0.0005±0.001)
	Wav2Vec2+Character	(0.1±0.0, 0.02±0.007)	(0.05±0.005, 0.02±0.01)
Brain Encoders	Bendr+Phoneme	(0.03±0.007, 0.0002±0.00007)	(0.02±0.006, 0.0±0.0)
	Bendr+Character	(**0.13±0.02**, **0.02±0.01**)	(0.04±0.005, 0.0±0.0)
	EEG-Conformer+Phoneme	(0.02±0.0001, 0.0004±0.00007)	(0.009±0.002, 0.0006±0.0004)
	EEG-Conformer+Character	(0.1±0.001, **0.02±0.001**)	(0.05±0.005, 0.03±0.002)

Table 6.2: Performance comparison of different techniques and scenarios using EEG and IMA metrics (BLEU, ROUGE).

In such circumstances, ascertaining the intended meaning demands the utilization of surrounding sentence structure and contextual cues, which remain beyond the reach of the decoder sans a suitably pre-trained LLM. Consequently, while the LLM occupies an essential role as an integral component within the system, it is imperative that the text generation mechanism facilitated by the decoder achieves a level of sophistication adequate for the precise extraction and formation of coherent and contextually pertinent sentences. The incorporation of an LLM is, therefore, fundamental to fulfilling this objective, ensuring that translations are not only syntactically accurate but also semantically significant. Furthermore, upon scrutinizing our pipeline for errors or defects, it became evident that our pipeline was affected by teacher forcing, consequently rendering our results invalid. In light of the aforementioned considerations, we

resolved to exclude our results, as they were deemed invalid and beyond the scope of our undertaking.

Given the dependence of LLMs on the decoder component, we commenced the integration of CTC loss into our workflow. This initiative was motivated by the aspiration to develop a more efficient decoder model prior to the deployment of the LLM. As detailed in our methodology Section 6.2.4, CTC loss is a well-recognized technique within the sphere of speech-to-text applications. These applications share the data format challenges encountered in our task, thus justifying our selection of this particular loss function.

Moreover, the speech-to-text field frequently contends with the challenge posed by variable-length input recordings, a dilemma analogous to the issues we face. The efficacy of CTC is demonstrated in its ability to manage this variability effectively. Nonetheless, our experiments employing only the CTC loss did not result in the expected improvement in the performance of our decoder, as depicted in Table 6.2. We hypothesize that the lack of progress is due to the absence of an Encoder mechanism. In particular a more focused Brain Encoder

It is imperative to situate this research within the extant body of literature, with particular emphasis on the encoder-decoder framework, which could yield further understanding of our hypothesis. We particularly wanted to utilised encoder-decoder architecture since many times previously in the literature has shown immense performance on different sequence-to-sequence tasks [216–218]. Since our task is generating a text sequence from a brain sequence an encoder-decoder architecture fits perfectly. This foundational gap underscores the possible need for the incorporation of an encoder to attain enhanced results. Therefore, although CTC loss provides certain benefits, its solitary application proves insufficient, necessitating the exploration of an augmented architecture to achieve superior outcomes.

To test the hypothesis stated above we incorporate different techniques to train different encoders and test them in our pipeline. Firstly we tried training our encoder model using 2 generic techniques : Data2Vec and Wave2Vec2. Both techniques outperform the current state-of-the art models in modalities such as videos , text and audio

and both offer a unique perspective of constructing agnostic modality encoders, hence why we wanted to test how they fair against brain data. We couldn't obtain any results for this 2 modalities for a very specific reason. The CTC loss obtain at the last training step was negative. A negative CTC means that some of the probabilities obtain in each time step using log Softmax are positive. Log Softmax's probabilities should be always negatives since it measures the negative log likelihood. We hypothesized that this is because our encoder was not learning correct representations could not provide good features to our decoder.

To address the identified challenges associated with encoder obstacles, we employed two existing pre-trained models: BENDR [58] and EEG-Conformer [212]. These models were originally developed as general-purpose EEG encoders across various datasets and were adapted through minimal modifications for use in classification tasks. Similarly, we have applied minimal adjustments to fine-tune these models on our datasets and subsequently assessed their efficacy as encoders within our processing pipeline. As illustrated in Table 6.2, the utilization of both encoders yielded results; however, these results were insufficient and inconclusive to achieve effective brain-to-text decoding.

Method	**Baseline**							
	Plain Baseline				**Random Baseline**			
	BLEU-EEG	ROUGE-EEG	ROUGE-IMA	BLEU-IMA	BLEU-EEG	ROUGE-EEG	ROUGE-IMA	BLEU-IMA
CTC+Character	0.5	**0.05**	**0.01**	0.41	0.2	**0.001**	**0.02**	0.66
Data2Vec+Character	**0.001**	**0.05**	**0.03**	**0.001**	**0.001**	**0.05**	0.12	**0.001**
Wave2Vec+Character	**0.001**	**0.05**	**0.02**	**0.01**	**0.001**	**0.05**	**0.05**	0.42
Bendr+Character	0.29	0.1	**0.001**	**0.009**	0.2	0.15	**0.001**	**0.05**

Table 6.3: P-values for statistical significance comparison between different methods against the Random and Plain baseline.

To further substantiate the validity of our results, we conducted a pair-wise t-test for all the methods outlined in Table 6.2 against both the random baseline and the baseline provided by Wang et al. [145]. Based on the findings presented in Table 6.2, our four

best-performing methods are as follows: "CTC+Character," "Data2Vec+Character," "Wav2Vec2+Character," and "Bendr+Character." The first method surpasses both the random and plain baselines in analyses of both EEG and IMA data. Conversely, the latter three methods exhibit superiority over the two baselines when applied to the EEG data.

Our statistical analysis, detailed in Table 6.3, demonstrates that in the majority of experimental conditions, within both datasets and for various evaluation metrics, there is a statistically significant difference when comparing the results to the plain and random baselines. This is supported by our p-value calculation, which yields a score of p and $p \leq 0.05$, indicating statistical significance. However, in some instances, a significant difference is not observed between the two baselines, suggesting that while most results indicate statistical significance, it is not possible to definitively conclude the superiority of one method over all others. Nevertheless, the Data2Vec+Character method exhibited the highest statistical significance against both baselines across both datasets and metrics. Moreover, the Wav2Vec2+Character method achieved a comparable level of statistical significance. These findings, combined with the application of these methods for developing a generic brain-encoder to enhance brain feature extraction for the decoder, suggest that employing a brain encoder can indeed improve the model's performance.

During the course of our analysis, we identified a previously overlooked inconsistency within our dataset: certain scenarios depicted in Table 6.2 exhibited a lack of data in the results section. A thorough scrutiny of the execution logs disclosed that throughout the model's training phase, the loss metric remained consistently high. Typically, loss metrics are expected to fluctuate within a range of approximately 0.1 to 1.0, indicative of satisfactory model performance. In contrast, our observations revealed that the loss values exceeded the threshold of 2.0, with certain training instances exhibiting values as high as 30.0 and above. Such elevated loss values precipitate the occurrence of 'gradient explosion,' whereby the model's gradients diverge, culminating in the computation of infinite loss. Consequently, this unfavourable phenomenon disrupted the training process, rendering the model incapable of effectively learning from the supplied data.

This finding is crucial as it highlights potential deficiencies in the training phase that could impede model convergence, thereby providing valuable insights for enhancing our methodology in subsequent experiments.

6.4 Chapter Summary

In summary, the chapter entitled "Brain-to-Text Decoder" examines the progression of brain research with an emphasis on advancements in neuroimaging technologies such as EEG, MRI, and MEG. These advancements have significantly enhanced our comprehension of brain activity and cognitive processes. The study underscores the complexity of brain regions and their interconnections, highlighting a particular focus on the brain's mechanisms for language comprehension, thus contributing to the field of neurolinguistics.

The study utilizes raw EEG data for the identification of linguistic components, thereby underscoring the greater efficiency of EEG compared to the slower and more expensive fMRI. By employing the ZuCo datasets, which integrate EEG and eye-tracking methodologies, the research aims to decode brain activity to produce coherent interpretations. These datasets are crucial for advancing the comprehension of reading behaviour and neural processes.

Difficulties in maintaining dataset consistency necessitated the incorporation of intra-cortical microelectrode arrays to achieve enhanced data resolution. The research recognizes the limitations inherent in the ZuCo dataset, particularly its deficiency in repetitive data, which is essential for effective AI training. Consequently, single-subject IMA data was utilized to enhance the reliability of the training process.

Moreover, the research investigates the implementation of sophisticated language models, in conjunction with CTC loss for the management of unsegmented data. Contrary to expectations, these advancements did not significantly enhance performance, thereby necessitating the incorporation of data augmentation and feature extraction methodologies.

This chapter advocates for the utilization of the Conformer model architecture, which combines CNNs with self-attention mechanisms, to augment the processing of

sequential data. It incorporates Log Softmax to achieve greater computational efficiency and stability. The research comprehensively highlights the inherent challenges and recent advancements in the development of brain-to-text models, bearing significant implications for technologies aimed at aiding speech-impaired communication.

Our findings indicate that merely substituting the LLM within an extant pipeline does not substantially enhance outcomes. This concurs with literature suggesting that LLMs primarily function as corrective mechanisms rather than transformative agents in tasks such as speech-to-text. The accurate interpretation of homophones, for instance, is heavily contingent upon contextual cues, necessitating a well-trained LLM integrated with an advanced decoder.

In order to enhance the efficiency of the decoder prior to the deployment of a LLM, the researchers conducted experiments using CTC loss, which is prevalent in speech-to-text applications managing variable input lengths. Nevertheless, the application of CTC loss in isolation did not result in improved performance of the decoder, possibly attributable to the lack of an encoding mechanism.

Subsequent experimentation with encoders utilizing Data2Vec and Wave2Vec2 methodologies encountered challenges, as evidenced by negative CTC loss values and high loss resulting in exploding gradients. While the application of pre-trained models such as BENDR and EEG-Conformer as encoders yielded some promising outcomes, they were insufficient for proficient brain-to-text translation. Furthermore, the statistical significance analysis did not conclusively determine whether some of our approaches outperformed the random and plain baseline; however, it demonstrated statistical significance in most cases across the two datasets. The most promising approach, with statistical significance, was identified as Data2Vec+Character. These findings emphasize the necessity for a more cohesive integration of an encoder-decoder framework ,with a dedicated brain encoder to enhance overall results, as a future direction for this research. We believe that the creation of such an encoder should be investigated further to enhance the results.

Chapter 7

Design and Development of a Customizable Brain-Powered Chatbot System

7.1 Introduction

As explored in Chapter 6, a primary factor contributing to the unsuccessful training of the Brain-To-Text decoder is the inadequate quality of the underlying data. This issue is critical since, akin to all machine learning experiments, the performance and precision of models are intrinsically linked to the quality and robustness of the datasets they are trained on. Therefore, it is imperative to meticulously capture, curate, and organize these datasets to ensure their reliability and applicability.

In pursuit of this objective, significant efforts have been made to gather EEG data, resulting in the creation of various EEG datasets. Notable contributions to this field include datasets referenced in [140,141,219]. A common attribute among these datasets is the introduction of stimuli in diverse forms such as text, images, or videos, intended to elicit responses from the subjects' brains. Thus, employing an appropriate stimuli presentation medium is essential in facilitating such interactions.

Moreover a prevailing trend in the field of Neuroscience is the utilization of EEG data to develop machine learning models for a diverse range of applications, including

elementary classification tasks [76, 220–223] and more intricate pursuits such as text generation from EEG signals [145]. The real-time acquisition capabilities of EEG data, coupled with its cost-effectiveness relative to other neuroimaging techniques like fMRI, render it an ideal approach for deployment in real-world scenarios, thereby contributing to its increasing popularity in recent years.

However even though a lot of attempts have been made still some of the datasets lack on quality for machine learning integration. This can happen for many reasons.One of these reasons identified is the lack of a standardized medium amongst the community [224], with often different datasets using different medium for stimuli presentation. Moreover another reason identified is the lack of customization, flexibility and exact timing of the stimuli presented [225–228].

The objective of this chapter is to introduce NeuraSearch Chat (NSChat) along with its capabilities. This system is specifically designed to address the identified challenges and facilitates the presentation of stimuli within a more authentic environment. It also serves to bridge the gap between the interaction with chatbot agents and the collection of neural data during these interactions, including those with large language models (LLMs). Furthermore, it provides precise timing and customization, allowing for seamless use across various settings. Additionally, a notable feature of this system is its capacity to concurrently record EEG signals as the system operates. This functionality offers a foundational design interface for the conceptual system intended for development in this thesis..

The rest of the Chapter is formatted as followed: in Section 7.2 we showcase the features of NSChat and how they can be used to overcome the current difficulties and finally in Section 7.3 we state future improvements coming to NSChat.

7.2 System architecture

In this section, we present an exhaustive examination of the functionalities inherent within the NSChat System. We explicate the complexities of the current user interface, delivering a detailed analysis of its design and operational attributes. Moreover,

we explore the various features that have been carefully implemented to address and mitigate the challenges and difficulties outlined in Section 7.1.

7.2.1 User Interface

The NSChat System is accessible via both web and mobile platforms, utilizing the React framework for its front-end development. React provides the capability to develop a singular application compatible with various platforms, thereby ensuring a responsive design adaptable to a wide range of screen sizes.

Upon initial entry into the platform, users are prompted to input a username and an experiment code. The username serves as a means to verify and correlate the data recorded by the system for each participant. Our objective is to employ NSChat as an experimental instrument rather than a conventional chatbot; hence, the experiment code corresponds to the ongoing experiment, enabling researchers to subsequently verify the specific experiment in which the participant was engaged within the provided dataset.

Upon the successful input of user credentials, the primary chat interface is displayed. The user interface is designed to be straightforward, featuring a solitary chat window through which users can submit queries and receive responses from NSChat. Situated on the left side is a settings icon, which presently provides five sets of parameter customization options.

1. Modify the language model responsible for generating the response.
2. Alter the manner in which the results are displayed.
3. Adjust the interval at which each word is rendered on the screen.
4. Change the font size.
5. Change the line spacing.

At the bottom of the interface, there is a text input field designated for user entries, which employs the currently selected language model to generate a response. Addition-

ally, the user is afforded the capacity to evaluate the utility of a response by utilizing the thumbs up or thumbs down buttons situated adjacent to each response.

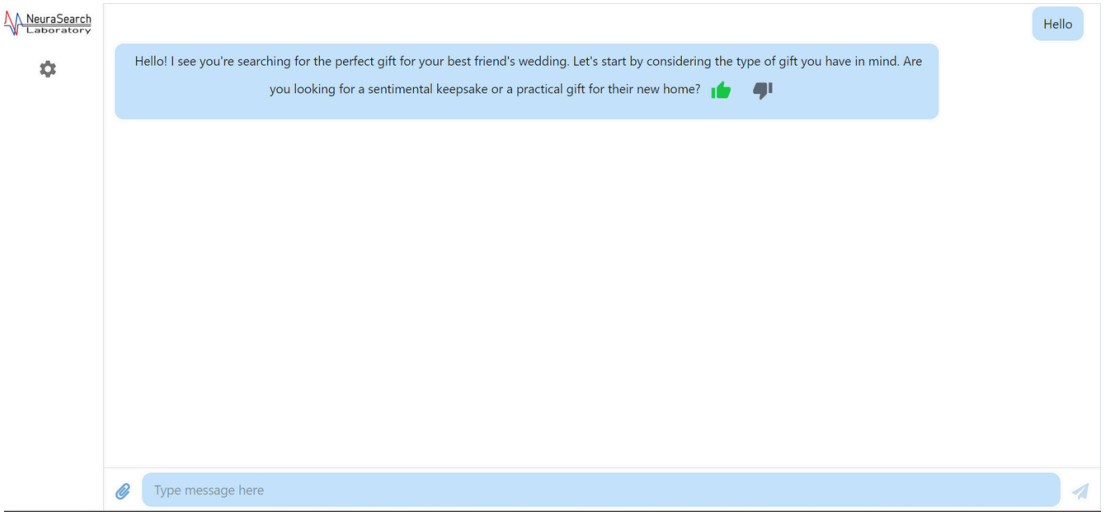


Figure 7.1: Basic Usage of the NSChat System: The user transmits a message and subsequently receives a response, which they then evaluate as pertinent by selecting the thumbs-up icon..

7.2.2 Flexibility and Customization

The NSChat offers considerable flexibility and customization for researchers, facilitating adaptation to diverse experimental scenarios with minimal modifications to the codebase. The system framework permits the augmentation of functionalities, enabling seamless implementation of new features. As detailed in Section 7.2.1, the user interface permits customization across four distinct parameters. The first parameter involves the selection of the LLM employed to generate responses. The second parameter pertains to the modification of response presentation to participants, with current options including "Show whole response," "Show one word at a time," "Show one sentence at a time," and "Typing."

The initial option delivers the complete response to the user instantaneously, whereas the subsequent three options present the response incrementally—either one word, one sentence, or one character at a time. This characteristic is especially beneficial in neuroscience research, as it aids in the examination of reading intervals when users interact

with paragraphs or discrete words.

The two additional adjustable parameters encompass the duration for which each word is displayed on the screen and the inter-word fixation interval. Users are afforded the ability to regulate the display duration of words and the intervals between them when utilizing formatting response options two and three. Notably, these temporal modifications can be effected without any code alterations; researchers are enabled to make modifications directly via the user interface.

Furthermore, the NSChat system incorporates two additional features for customization. Researchers have the capability to modify the font size and line spacing of the textual responses. This adaptability serves two primary functions. Firstly, it establishes a platform usable by individuals with visual impairments without necessitating code alterations, thus enhancing accessibility for a broader audience. Secondly, it facilitates the precise alignment of eye tracking with the actual reading behaviour of users when an eye tracker is employed, thereby preventing misinterpretation of word positioning. In neuroscience experiments, eye trackers are frequently utilized as a supplementary input to ascertain which specific word a user is reading. By allowing for the precise identification of words without the risk of the eye tracking circle moving over two words, due to adjustments in line spacing, the quality of the data collected is enhanced.

Finally, NSChat offers a template configuration file that researchers can configure in advance of conducting experiments to enhance the customization of system behaviour. Users have the capability to incorporate prompts into NSChat, enabling them to adjust the model's responses without necessitating modifications to the code. Additionally, a secondary configuration file is provided, allowing researchers to establish default values prior to experimentation, thereby facilitating extensive customization of system parameters and behaviours.

7.2.3 Utilizing Different LLMs

The NSChat system exhibits a notably advantageous capability that facilitates the integration of responses from a variety of large language models (LLMs) while preserving

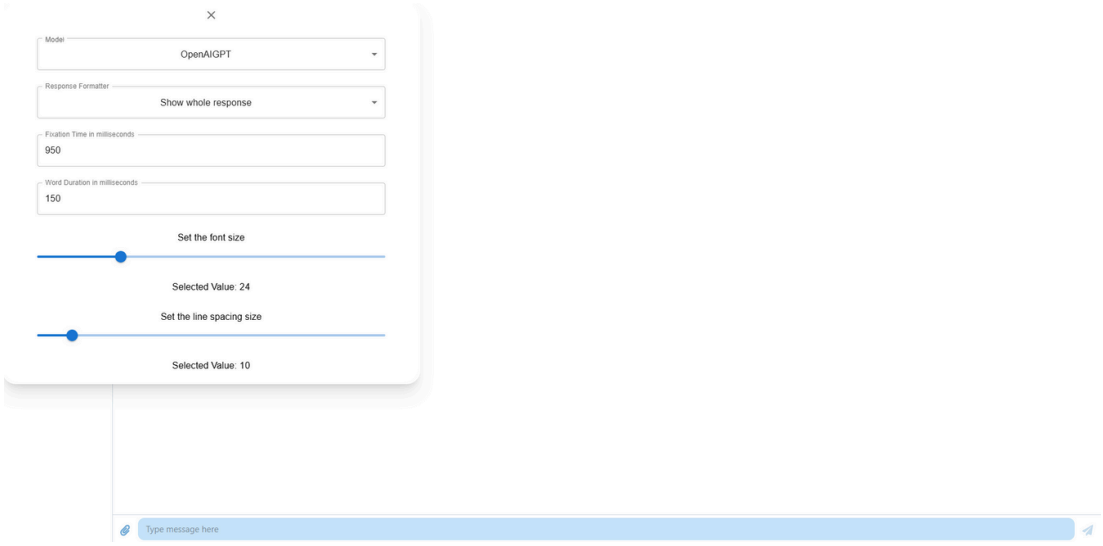


Figure 7.2: This Figure showcases the customization available for the NSChat in a set of parameters that can be set by the user.

the continuity of an active dialogue. By maintaining the current conversation at a level superior to that managed by individual LLMs, NSChat permits users to dynamically transition between different models, as depicted in Figure 7.2. This functionality is especially advantageous in experimental scenarios, as it permits researchers to concurrently assess and compare responses from multiple LLMs. The entire backend of NSChat is implemented in Python, which streamlines the process of incorporating and integrating new models into the system. Furthermore, NSChat affords researchers the latitude to introduce novel or custom models, contingent upon their ability to be loaded and executed within a Python environment. This architectural design not only augments the adaptability of NSChat but also promotes a more thorough examination of LLM capabilities across diverse research settings.

7.2.4 Logging Mechanism

The NSChat system incorporates a customizable logging mechanism that is specifically engineered to document various user events and interactions with the interface. This mechanism was developed to fulfil two primary objectives: firstly, to ensure that the system can be readily expanded to accommodate new foundational events as required;

and secondly, to enable researchers to integrate their own logging events with minimal programming effort, thereby obviating the need for implementing a distinct logging mechanism. NSChat systematically records user interactions across both the front-end and back-end, thereby safeguarding against data loss in scenarios of server or user interface malfunctions by archiving all interactions in a backup location for subsequent retrieval.

By default, the system chronicles an array of events, encompassing user interactions such as clicking the approval or disapproval buttons, hovering over response indicators, dispatching responses, receiving model-generated replies, and noting the initiation and termination times for response display. The logging framework is architected with a general approach, empowering users to selectively implement the specific events they endeavor to log while seamlessly integrating these bespoke events within the extant system. Such adaptability augments the utility of NSChat for researchers engaged in EEG experiments and other interactive investigations, thereby ensuring expansive data acquisition without superfluous complexity.

7.2.5 PyLSL integration

Within the domain of EEG and neurophysiological data collection, a considerable difficulty lies in achieving precise synchronization of stimuli and accurately documenting the timing of stimulus presentations to participants. Researchers in psychology frequently employ basic software packages that lack the advanced customization options present in NSChat. To augment the neuroscience functionality of NSChat, the PyLSL library has been incorporated into its backend infrastructure. The Lab Streaming Layer (LSL) offers a robust synchronization mechanism that generates precise timestamps for stimulus presentations. Moreover, it facilitates seamless integration with any EEG recording software proficient in capturing LSL streams, as LSL operates over a local network, obviating the need for direct connections between devices.

Researchers are able to enable or disable the PyLSL integration through a simplified setting in the configuration file. By default, the system logs all events delineated in Section 7.2.4 as triggers within PyLSL. This architecture allows users to integrate new

triggers for EEG experiments without the necessity to manually incorporate PyLSL or its related capture mechanisms. Consequently, NSChat not only streamlines the data collection process but also substantially enhances the flexibility and adaptability of experimental arrangements in neuroscience research. To assess the synchronization between the LSL protocol and our system, identical events are fired both with and without LSL to examine the temporal duration between events. Figure 7.3 demonstrates that nearly 85% of the events exhibit an offset of less than 1 ms compared to the LSL event, thereby affirming that our system is well synchronized.

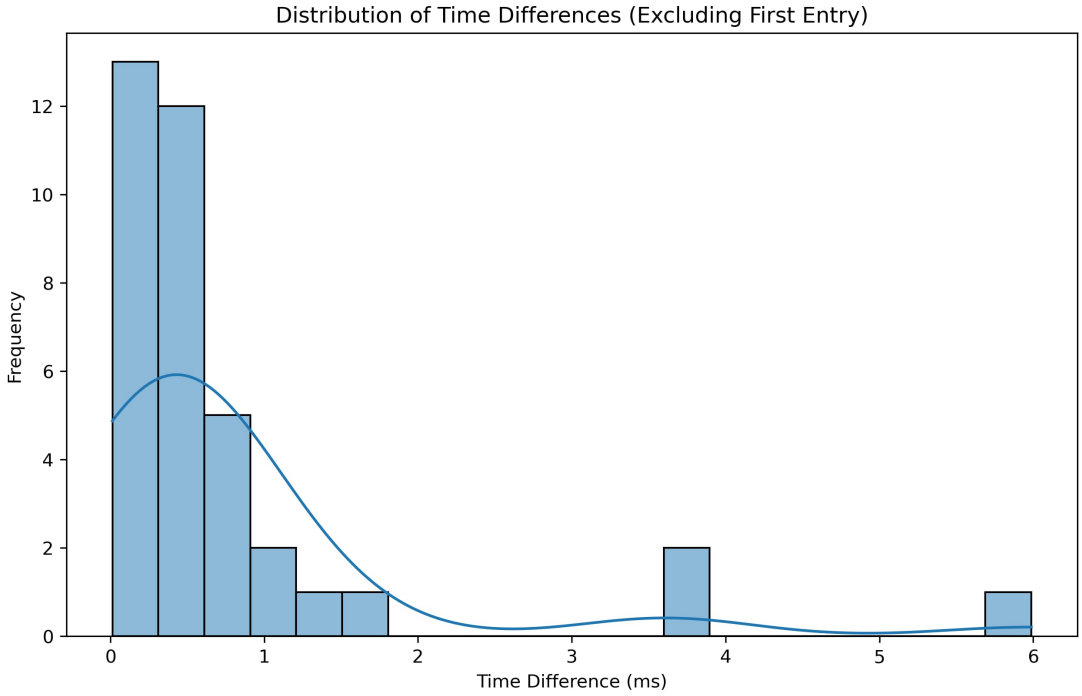


Figure 7.3: This figure illustrates that the majority of events in NSChat exhibit an offset of less than 1 millisecond.

7.3 Conclusion and Future Work

In this study, we introduce NSChat, an online chatbot system specifically crafted to aid neuroscience research. The system is distinguished by its flexibility and adaptability across diverse research contexts, while retaining a user-friendly interface. The configuration process is simplified through the employment of configuration files, thereby

obviating the necessity for programming expertise. Moreover, NSChat incorporates a comprehensive logging mechanism and a trigger system that can be immediately deployed or effortlessly expanded as required. Prospectively, we intend to incorporate a mechanism that addresses server delays, thus enabling NSChat to operate proficiently in distributed online environments beyond the constraints of local access. It is pertinent to highlight that although our system is principally constructed to support neuroscience research, it is not confined to this domain; it can be readily adapted to support information retrieval research or interactions with chatbot agents in general. Ultimately, the features of the system have been meticulously developed to address the requirements of standard experimental designs prevalent in various neuroscience studies [1, 140, 141]. The system has yet to undergo a comprehensive evaluation by users, except for a preliminary pilot test conducted by a single test user. As part of future work, we intend to disseminate the system to a broader user base in order to facilitate a thorough evaluation process and obtain feedback from diverse users.

7.4 Chapter Summary

In summary, the chapter examines the obstacles encountered in training the Brain-To-Text decoder, identifying inadequate data quality as a principal cause of failure. It underscores the indispensable role of robust datasets in machine learning and details initiatives to compile electroencephalogram (EEG) datasets with diverse stimuli. Although EEG data holds promise for applications such as text generation and basic classification, challenges regarding inconsistencies in the medium of stimuli and a paucity of customization options remain. To mitigate these issues, the chapter presents NSChat, a system devised to enhance data collection by incorporating neural data acquisition during interactions, including those with large language models (LLMs). NSChat is characterized by its precise stimuli timing, high degree of customizability, and adaptability across web and mobile platforms.

Leveraging its Python-based architecture, the system facilitates dynamic model integration and offers extensive configurability through template files. It employs PyLSL to achieve synchronization with EEG recording software, thereby ensuring precision

in data logging and integration, aspects that are particularly beneficial for both neuroscience and broader research applications. The chapter concludes by highlighting NSChat's adaptability for applications extending beyond the realm of neuroscience, with proposed future enhancements targeting the reduction of server delays and the augmentation of distributed system functionalities.

Chapter 8

Conclusion And Future Work

8.1 Introduction

This section furnishes a comprehensive overview of the contributions and principal conclusions derived from this thesis. It encapsulates an exhaustive summary of the findings and meticulously explores how the research questions, initially posited at the commencement of this study, have been addressed. By elucidating the core essence of the work executed, it ensures that the pivotal research outcomes are cogently articulated, underscoring the methodological thoroughness and analysis employed.

Furthermore, this section re-examines the fundamental questions that directed the research process, elucidating the systematic approach undertaken to explore and resolve each inquiry throughout the study. Accordingly, it underscores the thesis's significance and pertinence within its respective field, while establishing a basis for potential future research and academic contributions.

8.2 Chapter Outline

In the ensuing sections of this thesis, we furnish a thorough elucidation of our capacity to address the research questions posited at the outset of this study. Initially, in Section 8.3.1, we engage in an in-depth reflection on our originally proposed high-level architectural paradigm, evaluating its feasibility and consistency with our research objectives. Subsequently, in Section 8.3.2, we explore our findings and explicate the

methodologies employed to establish a consistent alignment between NLP models and the cognitive representation of language within the brain. This section emphasizes our strategy in bridging the divide between computational models and neurological evidence. Progressing further, Section 8.3.3 expounds upon the insights derived from training these models on a meticulously curated dataset, aimed at optimizing text generation tasks. We concentrate on key discoveries and identify crucial influences that such data exerts on the models' efficacy. Furthermore, Section 8.3.5 provides a comprehensive account of our initiatives towards developing a fully functional brain-to-text decoder. In this context, we elucidate both the results obtained and the challenges encountered throughout the development process. Concluding this discourse, the final Section 8.4 offers a reflective overview of the thesis as a cohesive body of research. Here, we conduct a critical evaluation of the path undertaken and propose potential avenues for future investigation, establishing foundational pathways that may amplify and broaden the scope of this vital research frontier.

8.3 Contributions

8.3.1 High Level Architecture

The primary objective and central research question formulated at the inception of this thesis concerned the development of a comprehensive framework for a system facilitating user interaction through the application of neural data. The principal aim was to conceptualize a high-level architectural blueprint that explicates the mechanisms enabling such interactions, with a particular focus on incorporating neural signals as input data. Consequently, this thesis seeks to meticulously analyse and examine one of the critical components identified as essential to the proposed system—a Brain-To-Text decoder. This component functions as a vital interface for converting neural impulses into coherent textual outputs.

In Chapter 3 of this thesis, a thorough and comprehensive investigation is presented, meticulously dissecting all the necessary components essential for the construction of a brain-controlled communication system. Each element of the system has been ex-

amined independently and in detail, offering a comprehensive understanding of their functions and interconnections. The analysis further elucidates the numerous potential challenges and technical obstacles that must be overcome during the development of these components, thereby offering a roadmap for future progress in this emerging field of research.

In our extensive research, we have systematically identified three essential components that are fundamental for the development of this complex system. The first component is an advanced module designed to meticulously manage the recording of data, while simultaneously maintaining a comprehensive history of these recordings. This historical record is crucial as it allows for the provision of these recordings to the Word Perceive Model in a strictly chronological order, ensuring seamless integration and analysis. Secondly, the implementation of a robust Word Perceive Model is indispensable and critically vital to the success of the system. This component fulfils the important function of accurately detecting the precise instances when the user conceptualizes a word. Considering the continuous activity within the human brain, it is essential to accurately determine when and which of the recordings should be transmitted to the Brain-To-Text decoder.

This process ensures an exact and thorough classification of the words considered by the user. The third and pivotal component of the system is the Brain-To-Text decoder. This fundamental element undertakes the intricate task of translating brain data into textual output. It subsequently transmits this textual data to various applications, which may include a chatbot, virtual assistant, or even a search engine system. Within this framework, our initial research attempts have effectively demonstrated the feasibility of developing such a system. Additionally, we have elucidated the critical components essential for the construction and operation of this innovative system.

8.3.2 Are the current state-of-the art NLP models brain aligned

This thesis primarily aims to investigate the potential existence of a substantive link between state-of-the-art NLP models and the field of cognitive neuroscience. It entails a comprehensive examination of whether there are significant parallels in the mechanisms

of representation learning that are shared by both artificial intelligence systems and the human brain.

Chapter 4 this thesis significantly extends the current body of scholarly literature through an in-depth analysis of four state-of-the-art NLP models: RoBERTa, DistilBERT, ALBERT, and ELECTRA. Each model, though based upon the foundational BERT architecture which introduced the groundbreaking transformer and attention mechanisms, demonstrates unique features designed to address specific computational challenges within the NLP field. Our detailed examination emphasizes the distinctive attributes inherent in each model, attributes which have been astutely optimized to address the constraints of the original BERT model, including but not limited to challenges of memory usage and computational efficiency.

Throughout our comprehensive investigation, we uncovered a noteworthy finding: despite the inherent disparities among these models, primarily resulting from varied technical priorities and limitations, all four models exhibited an unexpected capacity to generate representations that more closely resemble neural representations observed in the human brain. This phenomenon indicates that the methodological advancements intended to address computational requirements have inadvertently fostered an enhanced congruence with human cognitive processes. Specifically, the models' acquired representations, initially honed to fulfil technical objectives, also encompassed a degree of alignment with brain function that was not explicitly pursued during the initial development of these technologies. We also uncover that the first layers of these models are more brain aligned than the subsequent one, a result which is aligned with the initial research done by Toneva at Al. [1]

In addition, this thesis delineates substantive progress in methodological frameworks, particularly through the refinement of models used to assess the alignment between brain functions and NLP models. Driven by the objective of enhanced efficiency, we have realized a remarkable reduction in evaluation duration, enhancing the speed of this process by an unprecedented 300-fold relative to preceding methodologies. This advancement holds profound implications, significantly increasing our capability to evaluate novel NLP models. It facilitates the expedited analysis of whether their

internal representations pertaining to natural language processing tasks are consonant with those employed by the human brain, thus providing an efficacious trajectory for future academic inquiry and practical implementation in the converging domains of Artificial Intelligence and Cognitive Neuroscience.

Moreover in this investigation, our research was advanced by executing experiments across four distinct scenarios, each employing varying context lengths to assess the role of punctuation in natural text interpretation. Our findings consistently reveal that as context length increases, the significance of punctuation in language comprehension markedly decreases. This suggests that as the volume of context within a sentence expands, the sentence's meaning can be comprehended effectively without reliance on punctuation marks. By examining these varying context lengths, we highlight the flexible nature of language processing and its capacity to derive meaning even in the absence of conventional punctuation symbols.

In summary, our primary research question seeks to determine whether there exists a relationship between state-of-the-art NLP models and cognitive neuroscience, specifically focusing on similarities in representation learning between artificial systems and the human brain. This study examines four NLP models—RoBERTa, DistilBERT, ALBERT, and ELECTRA—that are rooted in the BERT architecture, although they have been optimized to address particular computational challenges. Despite their varied optimizations, all models unexpectedly yield neural-like representations akin to those found in the human brain, not as a result of deliberate design but as an unintended outcome of addressing computational necessities. The research further introduces significant advancements in evaluation methodologies, achieving a 300-fold reduction in evaluation time, thereby enhancing the analysis of NLP models' congruence with brain functions. This advancement holds considerable implications for future research and practical applications within the fields of artificial intelligence and cognitive neuroscience. We posit that by establishing these neural connections and developing a sufficiently rapid framework for evaluating emerging models, we are positioned to affirmatively address our initial research question.

8.3.3 Effectively train NLP models on brain data

The second research question this thesis aims to answer is if we can effectively train deep learning models to generate text from brain data. Chapter 5 encompasses a detailed analysis followed by the optimization of neural network models, accomplished through the integration and application of a diverse spectrum of activation functions. It has been conclusively established that the employment of non-linear and tunable activation functions can substantially enhance the performance of these models while preserving the integrity of their foundational architectures. A meticulous and thorough retraining process was executed on the EEG-to-Text decoder. This process entailed the deployment of various activation functions, which provided substantiated evidence that tunable functions, such as polynomial activation functions of higher degrees, yielded significantly superior results by suppressing the traditional use of ReLU activation function.

A Transformer Encoder was employed in this framework, where it underwent meticulous optimization through experiments with various activation functions. The architecture was engineered to incorporate raw EEG data, processed through sophisticated transformation algorithms, highlighting the complexity introduced by managing bidimensional raw EEG data. The encoder comprised multiple layers initialized with random weights, focused on extracting features from EEG signals. This was further complemented by the integration of a pre-trained BART model for decoding. The evaluation utilized BLEU and ROUGE scores as metrics, demonstrating that certain configurations significantly improved performance, particularly within the scope of brain-to-text translation tasks.

It was observed that configurations, particularly the one referred to as "head identical to layers,",which employs the same number of attention heads and hidden layers, significantly enhanced performance, especially when employed in conjunction with learnable activation functions such as leaky ReLU. Moreover, the deliberate placement of normalization layers was found to substantially affect performance, with notable declines occurring when alterations were misaligned with the activation function's dynamics. Consequently, the study substantiates the assertion that the integration of

advanced activation functions within neural architectures not only augments EEG-to-text translation performance but also pinpoints specific configurations that yield the most favourable outcomes.

To summarise Chapter 5 investigates whether deep learning models can be effectively trained to generate text from brain data. It finds that non-linear and tunable activation functions significantly enhance model performance while maintaining architectural integrity. Through optimizing a Transformer Encoder with various activation functions and integrating a pre-trained BART model, the study achieves significant improvements in brain-to-text translation tasks, as measured by BLEU and ROUGE scores. A "head identical to layers" configuration with learnable activation functions like leaky ReLU, combined with strategic normalization layers, is identified as particularly effective. The research confirms that advanced activation functions improve EEG-to-text model performance and identifies specific configurations for optimal outcomes. The answer to the questions here however is not clear. Yes we are able to suppress the current pipeline but still the results were not very promising on generating how quality text from our pipeline. That is why moving to Chapter 6 we explored a more generic technique that excel on generating text from different modalities than text.

8.3.4 Brain-To-Text Decoder

The third research inquiry centred on evaluating the feasibility of generating textual outputs directly from raw brain data. To investigate this possibility in detail, we conducted an extensive series of experiments as outlined in Chapter 6. This exploration involved the utilization of two distinct datasets, each presenting specific challenges and insights. We systematically applied three diverse experimental designs to rigorously test and validate our hypotheses. This methodological diversity facilitated a comprehensive analysis of our primary research question, allowing us to derive meaningful conclusions regarding the potential transformation of brain data into coherent text.

Throughout the course of our experimental investigation, we conducted a comprehensive analysis of the implementation of Connectionist Temporal Classification (CTC)

loss in conjunction with two generic techniques for embedding generation, namely Data2Vec and Wav2Vec2. These techniques supplied a foundational framework essential for comprehending the potential evolution of embeddings within our system. To further enhance our experimental pipeline, we incorporated two specialized, pretrained EEG encoders, namely BENDR and EEG-Conformer. These encoders were meticulously selected to augment our capacity to accurately interpret intricate EEG data, thereby bolstering the analytical efficiency of our trials. The integration of these components significantly contributed to the refinement of EEG signal processing.

Our experimental approach was augmented by the implementation of two distinct tokenizers: one employing a standard English vocabulary and the other utilizing a phonetic vocabulary. This bifurcated methodology permitted a diversification in our analysis of language-based data, thereby improving our capacity to effectively handle and process varied data inputs. Notably, our entire experimental framework was predicated upon the Conformer architecture, a decision that was pivotal in enabling the extraction of both spatial and temporal features from the EEG data. This facilitated a more comprehensive and refined understanding of the information encapsulated within these signals.

Unfortunately, the development of an effective brain-to-text model that performs satisfactorily in both BLEU and ROUGE metrics was not achievable. A major challenge encountered during the implementation was the occurrence of gradient explosion, attributable to excessively elevated loss values. Such elevated loss values may result from several factors, including excessively high learning rates, the presence of data noise, or a lack of alignment between the dataset's quality and the task requirements. Additionally, it is noteworthy that the ZuCo dataset lacked adequate repetition at the subject level. Consequently, it necessitated the aggregation of data from all subjects for both training and validation phases, thereby escalating the risk of high loss values and subsequent gradient explosions.

In conclusion, Chapter 6 presents an investigation of various experimental scenarios aimed at developing a machine learning model capable of generating coherent text solely from brain signal data. However, despite our efforts, the experimental design did

not facilitate the effective or comprehensive training of such a model, thus leaving our central research question only partially addressed and unresolved. Notwithstanding these limitations, the research into text generation from brain signals continues to be a significant and open-ended question within the scientific community. Although our research did not yield a definitive conclusion regarding the feasibility of this generation process, it does not entirely negate the potential for future success in this domain. In Section 8.4, we present an overview of preliminary research initiated throughout the course of this thesis, proposing potential pathways forward. This includes an analysis of recent research initiatives and observations that were not explored in detail due to time constraints. Accordingly, our findings represent preliminary steps towards advancing research in this field, highlighting pathways that could be further explored to achieve significant breakthroughs in the future.

8.3.5 NSChat

The final advancement elucidated in this thesis is the development of a novel system termed NSChat. This system adroitly integrates the functionalities of conventional chatbot agents with the complex discipline of Neuroscience. NSChat is meticulously designed to support researchers within scholarly domains, furnishing them with an advanced instrument to collect data with heightened efficacy in environments that more accurately replicate real-world conditions than previously attainable. The system features exceptional customizability and flexibility, alongside capabilities such as event timing precision at the millisecond level. These attributes ensure that NSChat is adaptable for utilization across a diverse array of settings for data collection, thereby extending its applicability even into specialized fields like Neuroscience. In achieving this, NSChat not only improves the efficiency and precision of data collection but also expands its utility, rendering it an indispensable resource for academic research pursuits.

NSChat serves a critical role as a foundational interface within our proposed Brain-To-Text system. By leveraging its capability to seamlessly incorporate EEG data in parallel with its operations with minimal code modifications, we have successfully im-

plemented a configuration where the EEG data operates as a direct input channel. However, the current absence of an adequately robust Brain-To-Text decoder constrains the system's functionality to merely recording the incoming data, thus precluding its deployment as a comprehensive end-to-end model. Notwithstanding these present constraints, the interface is inherently structured to support and facilitate the integration and application of various emerging models that may be developed for extensive research purposes in the foreseeable future. It also shows that is feasible to construct a computer interface integrated with non-invasive brain data.

8.4 Future Work

In this study, a thorough examination was undertaken of the innovative concept concerning the creation of an advanced computer system amenable to direct manipulation via neural input methods. This ambitious venture required an extensive investigation, culminating in the formulation of a comprehensive blueprint that delineates the core architecture and systematically addresses all foreseeable challenges associated with the development of such a sophisticated technological system. The project highlighted the imperative for a multifaceted approach to achieve successful implementation.

The central focus of this thesis lies in the comprehensive clarification of the Brain-To-Text decoder, an essential part of the system. This component underwent an extensive, in-depth examination owing to its crucial role and inherent complexity. The experimental investigations predominantly concentrated on the realization of this critical component, in line with the hypothesis that it constitutes the most technically demanding and challenging facet of the proposed system's architecture.

Nevertheless, the realization of such a system within the scope of a single PhD thesis proves impractical. This is attributed to the intricate and extensive nature of the work involved, requiring a substantial degree of further research and exploration. Consequently, engaging in this undertaking in the future offers numerous burgeoning opportunities for subsequent scholarly inquiry and investigation. Numerous challenges and unresolved issues persist, necessitating meticulous attention and critical examination. Therefore, addressing these multifaceted aspects will advance future research and

significantly contribute to the field.

The first challenge that offers a noteworthy research opportunity is the development and implementation of the Word Perceived Model. While its utility and the challenges it aims to address have been acknowledged, conducting thorough research and proposing a comprehensive implementation extend beyond the scope of this thesis. Nevertheless, we maintain that embarking on a rigorous examination of this component will not only uncover further complexities beyond those presently identified but also enhance the field by introducing additional avenues for research. Therefore, we assert that this constitutes a significant and promising direction for future academic investigations.

The second future opportunity identified resides in the capacity to markedly enhance the performance of the Brain-To-Text decoder. At the time of authoring this thesis, this module has been developed utilizing the most advanced and state-of-the-art techniques currently accessible in the field. Nonetheless, the domain of machine learning is advancing at a remarkable pace, especially in recent years. It is projected that more sophisticated methodologies will be introduced, enabling models to be trained more rapidly and with increased efficiency. Additionally, these advancements could significantly bolster the performance of existing models. Given that our objective is to render this system accessible to the general populace, attaining the utmost level of accuracy is imperative. Therefore, there exists substantial potential for refinement and enhancement to realize this ambitious goal. A prominent area of interest involves the development of a sophisticated brain encoder to deliver high-quality EEG features to the decoder. In any generative task, the quality of the embeddings or features produced is crucial for the decoder to effectively learn the relationship between these features and the text intended for generation. Specifically, it is posited that defining an appropriate loss function or employing a pre-existing one that more accurately aligns with brain data is imperative for enhancing the learning process.

Additionally, we have identified the need for enhanced curated datasets, which may serve as foundational components in the development of the Brain-To-Text decoder. To achieve this objective, the creation of a more naturalistic interface capable of attaining millisecond accuracy is required, as this will be crucial for the collection of precise

data. We assert that future research efforts should utilize NSChat as a methodological instrument for the acquisition of EEG data specifically tailored for the EEG-to-text task.

The next opportunity identified for future research involves the empirical evaluation of the system using human participants. At present, the conceptual blueprints and high-level architectural designs remain theoretical, as no practical testing has been conducted with live participants. Additionally, the Brain-To-Text Decoder has been assessed solely through offline test data sets, and not with actual human subjects, largely due to time constraints. Thus, it is posited that once the system reaches maturity, extensive research should be conducted to thoroughly assess its functionality. This will enable the identification and analysis of further challenges and potential issues that may emerge in real-world applications.

Ultimately, we have identified two new dimensions that we hypothesize require further exploration to effectively train a Brain-To-Text model capable of generating text from cerebral data. The initial dimension involves the integration of an external guiding modality into the encoder. Researchers at Meta AI have pioneered an innovative, non-invasive technique for speech decoding from cerebral activity through the employment MEG [229]. This methodology enables communication via a computer interface for individuals with speech impairments. The system's architecture encompasses a 'brain module' tasked with extracting information from MEG recordings and a 'speech module' dedicated to decoding speech representations. By providing the pre-trained speech module with the auditory stimuli experienced by participants and simultaneously processing the MEG recordings through the brain module, the researchers achieved the alignment of the two modules using contrastive learning. This process enabled the brain encoder to assimilate auditory-analogous features from the neural data.

Building on this observation, there is evidence suggesting that individuals engage in more complex cognitive processes when employing inner speech, as opposed to merely relying on contextual cues in natural language. Research indicates that cognition may occur through visual or auditory modalities [230–233]. In our study, we strived to synchronize models containing contextual information pertinent to the target text with

neural data. We posit that this approach constitutes a segment of the intended methodological pipeline. The exploration of other modalities and their integration with neural data represents an area that remains largely unexamined. We contend that this is an open domain warranting further investigation, especially in light of recent breakthroughs in machine learning and neuroscience.

Appendix A

Optimizing Brain Decoding using different Activation Functions

Activation Function	Rouge Score 1			Rouge Score 2			Rouge Score L		
	P±(STD)	R±(STD)	F±(STD)	P±(STD)	R±(STD)	F±(STD)	P±(STD)	R±(STD)	F±(STD)
Random Baseline	0.05±0.01	0.05±0.01	0.05±0.01	0.05±0.01	0.05±0.01	0.05±0.01	0.05±0.01	0.05±0.01	0.05±0.01
plain	0.04±0.007	0.009±0.007	0.001±0.007	0.000±0.0	0.000±0.0	0.000±0.0	0.041±0.007	0.009±0.0007	0.015±0.0007
swish	0.051±0.007	0.013±0.007	0.021±0.007	0.000±0.0	0.000±0.0	0.000±0.0	0.050±0.007	0.013±0.007	0.020±0.007
gelu	0.087±0.006	0.060±0.006	0.067±0.006	0.0026±0.0001	0.0023±0.0001	0.0022±0.0006	0.079±0.006	0.055±0.006	0.061±0.006
elu	0.156±0.007	0.046±0.007	0.069±0.007	0.0037±0.0005	0.0016±0.0004	0.0021±0.00007	0.150±0.005	0.044±0.005	0.066±0.007
leaky relu	0.155±0.007	**0.115±0.007**	**0.126±0.007**	**0.0142±0.007**	**0.0124±0.007**	**0.0122±0.007**	0.130±0.007	**0.098±0.007**	**0.106±0.007**
swish norm first	0.018±0.001	0.013±0.001	0.015±0.001	0.0002±0.00007	0.0001±0.0001	0.0002±0.0001	0.018±0.0001	0.013±0.0001	0.014±0.0001
parametric relu	0.158±0.007	0.114±0.001	0.114±0.001	0.0102±0.0004	0.0065±0.0003	0.0074±0.007	0.131±0.007	0.083±0.007	0.095±0.007
sine	0.009±0.001	0.008±0.001	0.008±0.001	0.0002±0.00001	0.0001±0.00001	0.0001±0.00001	0.009±0.0001	0.008±0.0001	0.008±0.0001
chebysev degree 3	0.170±0.007	0.053±0.007	0.078±0.007	0.0070±0.0001	0.0028±0.0001	0.0037±0.0002	**0.161±0.007**	0.050±0.007	0.074±0.007
chebysev degree 3 norm first	0.018±0.007	0.013±0.007	0.015±0.007	0.0002±0.00007	0.0001±0.0	0.0001±0.0	0.018±0.007	0.013±0.007	0.014±0.007
chebysev degree 2	0.136±0.007	0.048±0.007	0.069±0.007	0.0007±0.0001	0.0002±0.0001	0.0003±0.0001	0.126±0.007	0.045±0.007	0.063±0.007
torch poly 2	0.106±0.007	0.041±0.007	0.057±0.007	0.0005±0.0001	0.0002±0.0003	0.0003±0.0001	0.097±0.007	0.037±0.007	0.052±0.007
torch poly 3	**0.173±0.007**	0.06±0.007	0.087±0.007±0.0003	0.0055±0.0002	0.0024±0.0002	0.0031±0.0002	0.159±0.007	0.056±0.007	0.079±0.007
negative positive poly	0.142±0.007	0.044±0.007	0.065±0.007±0.0001	0.0027±0.0002	0.0010±0.0002	0.0014±0.0002	0.140±0.007	0.043±0.007	0.063±0.007
negative positive poly norm first	0.018±0.007	0.013±0.007	0.015±0.007	0.0002±0.00001	0.0001±0.00002	0.0002±0.00002	0.018±0.007	0.013±0.007	0.014±0.007

Table A.1: Rouge Scores results table

Bibliography

[1] M. Toneva and L. Wehbe, "Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain)," in *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc., 2019.

[2] A. Gulati, Y. Wu, Y. Zhang, J. H. Lee, Q. V. Le, and M. S. Kahn, "Conformer: Convolution-augmented transformer for speech recognition," in *Proceedings of the 2020 International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2020, pp. 7454–7458.

[3] A. Baevski, H. Zhou, E. A. Conneau, and M. Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," *arXiv preprint arXiv:2006.11477*, June 2020, version 3, last revised on 22 Oct 2020.

[4] A. Baevski, W.-N. Hsu, Q. Xu, A. Babu, J. Gu, and M. Auli, "Data2vec: A general framework for self-supervised learning in speech, vision and language," in *International Conference on Machine Learning*. PMLR, 2022, pp. 1298–1312.

[5] K. Kingphai and Y. Moshfeghi, "Mental workload assessment using deep learning models from eeg signals: A systematic review," *NeuraSearch Laboratory, University of Strathclyde*, 2024. [Online]. Available: https://strathprints.strath.ac.uk/90686/1/Kingphai-and-Moshfeghi-Mental-workload-assessment-using-deep-learning-models-from-EEG-S pdf

[6] J. Gonzalez *et al.*, "Classification of mental workload using brain connectivity and machine learning models," *Scientific Reports*, vol. 14, no. 1, pp. 1–12, 2024. [Online]. Available: https://www.nature.com/articles/s41598-024-59652-w

[7] F. Putze *et al.*, "Reproducible machine learning research in mental workload assessment using eeg data," *Frontiers in Neuroergonomics*, vol. 1, 2024. [Online]. Available: https://www.frontiersin.org/journals/neuroergonomics/articles/10.3389/fnrgo.2024.1346794/full

[8] A. Kumar *et al.*, "Online speech synthesis using a chronically implanted brain-computer interface in a man with impaired articulation due to als," *Scientific Reports*, vol. 14, no. 1, p. 60277, 2024. [Online]. Available: https://www.nature.com/articles/s41598-024-60277-2

[9] L. He *et al.*, "Brain-computer interface: Applications to speech decoding and synthesis," *Frontiers in Neuroscience*, vol. 16, 2022. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9130409/

[10] H. Que *et al.*, "Hellobench: Evaluating long text generation capabilities of large language models," *arXiv preprint arXiv:2409.16191*, 2024. [Online]. Available: https://arxiv.org/abs/2409.16191

[11] Y. Zhao, H. Zhang, S. Si, L. Nan, X. Tang, and A. Cohan, "Investigating table-to-text generation capabilities of large language models in real-world information seeking scenarios," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track*. Singapore: Association for Computational Linguistics, 2023, pp. 160–175. [Online]. Available: https://aclanthology.org/2023.emnlp-industry.17

[12] C. Caucheteux *et al.*, "Studying language processing in the human brain with speech and language models," in *Proceedings of the 22nd China National Conference on Computational Linguistics*, 2023, pp. 17–23. [Online]. Available: https://aclanthology.org/2023.ccl-4.3.pdf

[13] A. Narayanan *et al.*, "Overview of recent advancements in deep learning and artificial intelligence," *Elsevier*, 2023. [Online]. Available: https://asu.elsevierpure.com/en/publications/overview-of-recent-advancements-in-deep-learning-and-artificial-i

[14] Y. Fu, D. Guo, Q. Li, L. Liu, S. Qu, and W. Xiang, "Advances in artificial intelligence, machine learning and deep learning techniques," *MDPI Electronics*, vol. 12, no. 18, p. 3780, 2023. [Online]. Available: https://www.mdpi.com/2079-9292/12/18/3780

[15] H. Zhang *et al.*, "Electroencephalograph-based emotion recognition using brain connectivity features and domain adaptive residual convolutional network," *Frontiers in Neuroscience*, vol. 16, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2022.878146/full

[16] Y. Cao *et al.*, "Emotion recognition with eeg-based brain-computer interfaces: A systematic review," *Journal of Ambient Intelligence and Humanized Computing*, vol. 15, pp. 1–20, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s11042-024-18259-z

[17] J. Huang *et al.*, "Detecting emotions through eeg signals based on modified convolutional fuzzy neural network," *Scientific Reports*, vol. 14, 2024. [Online]. Available: https://www.nature.com/articles/s41598-024-60977-9

[18] A. Zhang and A. Author2, "Transformers and meta-tokenization in sentiment analysis for software engineering," *Journal of Software: Evolution and Process*, vol. 34, no. 1, pp. 1–15, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s10664-024-10468-2

[19] U. Hasson *et al.*, "The neurobiology of semantic memory," *PubMed Central*, vol. 3350748, 2011. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC3350748/

[20] Z. Lamprou, F. Pollick, and Y. Moshfeghi, "Role of punctuation in semantic mapping between brain and transformer models," in *Machine Learning, Opti-*

mization, and Data Science, G. Nicosia, V. Ojha, E. La Malfa, G. La Malfa, P. Pardalos, G. Di Fatta, G. Giuffrida, and R. Umeton, Eds. Cham: Springer Nature Switzerland, 2023, pp. 458–472.

[21] Z. Lamprou, I. Tenedios, and Y. Moshfeghi, "On the role of activation functions in eeg-to-text decoder," in *Machine Learning, Optimization, and Data Science: 10th International Conference, LOD 2024, Castiglione Della Pescaia, Italy, September 22–25, 2024, Revised Selected Papers, Part III*. Berlin, Heidelberg: Springer-Verlag, 2025, p. 46–60. [Online]. Available: https://doi.org/10.1007/978-3-031-82487-6_4

[22] Z. Lamprou and Y. Moshfeghi, "Customizable llm-powered chatbot for behavioral science research," 2025. [Online]. Available: https://arxiv.org/abs/2501.05541

[23] ——, "On creating a brain-to-text decoder," 2025. [Online]. Available: https://arxiv.org/abs/2501.06326

[24] S. Ogawa, K. Kwong, and et al., "Brain activation visualized in vivo by functional magnetic resonance imaging," *Proceedings of the National Academy of Sciences*, vol. 89, no. 13, pp. 5951–5955, 1992. [Online]. Available: https://www.pnas.org/doi/10.1073/pnas.89.13.5951

[25] M. Wegrzyn, J. Aust, L. Barnstorf, M. Gippert, M. Harms, A. Hautum, S. Heidel, F. Herold, S. M. Hommel, A.-K. Knigge, D. Neu, D. Peters, M. Schaefer, J. Schneider, R. Vormbrock, S. M. Zimmer, F. G. Woermann, and K. Labudda, "Thought experiment: Decoding cognitive processes from the fMRI data of one individual," vol. 13, no. 9, p. e0204338.

[26] M. S. M. Chow, S. L. Wu, S. E. Webb, K. Gluskin, and D. T. Yew, "Functional magnetic resonance imaging and the brain: A brief review," vol. 9, no. 1, pp. 5–9. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5241538/

[27] M. Mather, J. T. Cacioppo, and N. Kanwisher, "How fMRI can inform cognitive theories," vol. 8, no. 1, pp. 108–113. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3610572/

[28] J. F. Baizabal-Carvallo *et al.*, "Neuroimaging in functional neurological disorder: State of the art and future directions," *Frontiers in Neurology*, vol. 12, pp. 1–12, 2021.

[29] V. Voon *et al.*, "Neuroimaging in functional movement disorders," *Current Neurology and Neuroscience Reports*, vol. 19, no. 2, pp. 1–11, 2019.

[30] S. E. Shaywitz and B. A. Shaywitz, "Functional magnetic resonance imaging (fmri)," in *Biological Psychiatry*. Elsevier, 2005, vol. 57, no. 11, pp. 1180–1190.

[31] A. D. Friederici, "The brain basis of language processing: from structure to function," *Physiological Reviews*, vol. 91, no. 4, pp. 1357–1392, Oct. 2011.

[32] J. Brennan, Y. Nir, U. Hasson, R. Malach, D. J. Heeger, and L. Pylkkänen, "Syntactic structure building in the anterior temporal lobe during natural story listening," *Brain and Language*, vol. 120, no. 2, pp. 163–173, Feb. 2012.

[33] Y. Lerner, C. J. Honey, L. J. Silbert, and U. Hasson, "Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story," *Journal of Neuroscience*, vol. 31, no. 8, pp. 2906–2915, Feb. 2011, publisher: Society for Neuroscience Section: Articles.

[34] L. Wehbe, B. Murphy, P. Talukdar, A. Fyshe, A. Ramdas, and T. Mitchell, "Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses," *PloS One*, vol. 9, no. 11, p. e112575, 2014.

[35] A. G. Huth, W. A. de Heer, T. L. Griffiths, F. E. Theunissen, and J. L. Gallant, "Natural speech reveals the semantic maps that tile human cerebral cortex," *Nature*, vol. 532, no. 7600, pp. 453–458, Apr. 2016.

[36] I. A. Blank and E. Fedorenko, "Domain-General Brain Regions Do Not Track Linguistic Input as Closely as Language-Selective Regions," *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, vol. 37, no. 41, pp. 9999–10 011, Oct. 2017.

[37] L. Wehbe, A. Vaswani, K. Knight, and T. Mitchell, "Aligning context-based statistical models of language with brain activity during reading," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 233–243.

[38] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, 2018, pp. 2227–2237.

[39] D. Cer, "Universal sentence encoder," *arXiv preprint arXiv:1803.11175*, 2018.

[40] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *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)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186.

[41] Z. Dai, "Transformer-xl: Attentive language models beyond a fixed-length context," *arXiv preprint arXiv:1901.02860*, 2019.

[42] S. Jain and A. G. Huth, "Incorporating Context into Language Encoding Models for fMRI," bioRxiv, Tech. Rep., Nov. 2018, section: New Results Type: article.

[43] J. O. Caro, A. H. de Oliveira Fonseca, S. A. Rizvi, M. Rosati, C. Averill, J. L. Cross, P. Mittal, E. Zappala, R. M. Dhodapkar, C. Abdallah, and D. van Dijk, "Brain language model: A foundation model for brain activity dynamics," in *Proceedings of the International Conference on Learning Representations (ICLR)*, January 2024, published: 16 Jan 2024, Last Modified: 21 Apr 2024.

[44] M. Soufineyestani, D. Dowling, and A. Khan, "Electroencephalography (EEG) technology applications and available devices," vol. 10, no. 21, p. 7453, number:

21 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2076-3417/10/21/7453

[45] J. H. Medicine, "Sleep study," *Johns Hopkins Medicine*, 2024. [Online]. Available: https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/sleep-study

[46] L. A. Rechtschaffen and A. Kales, "Eeg recording and analysis for sleep research," *PMC*, 2005. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC2824445/

[47] A. Santos *et al.*, "The electroencephalogram in metastatic brain tumors," *Arquivos de Neuro-Psiquiatria*, vol. 61, no. 1, pp. 19–25, 2003. [Online]. Available: https://www.scielo.br/j/anp/a/nj3Tw57784tZmVPtmKM7J4t/?lang=en

[48] V. S. Selvam and S. S. Devi, "Analysis of spectral features of eeg signal in brain tumor condition," *Measurement Science Review*, vol. 15, no. 4, pp. 219–224, 2015. [Online]. Available: https://www.measurement.sk/2015/Selvam.pdf

[49] K. Wagner *et al.*, "Early electroencephalography for outcome prediction of postanoxic coma after cardiac arrest: A multicenter prospective cohort study," *Annals of Neurology*, vol. 86, no. 2, pp. 203–214, 2019. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC6771891/

[50] S. Scarfo, Y. Moshfeghi, and W. J. McGeown, "Random forest classifiers to predict psychotic symptoms in alzheimer's disease," vol. 20, p. e092242, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/alz.092242. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/alz.092242

[51] H. Jeong *et al.*, "Real-time seizure detection using eeg: A comprehensive comparison of recent approaches under a realistic setting," *arXiv preprint arXiv:2201.08780*, 2022. [Online]. Available: https://arxiv.org/abs/2201.08780

[52] Y. Wang *et al.*, "Epileptic seizure detection based on eeg signals and cnn," *Frontiers in Neuroinformatics*, vol. 12, 2018. [Online]. Available: https://www.frontiersin.org/journals/neuroinformatics/articles/10.3389/fninf.2018.00095/full

[53] H. He *et al.*, "Decoding eeg brain activity for multi-modal natural language processing," *Scientific Reports*, vol. 11, 2021. [Online]. Available: https://www.nature.com/articles/s41598-021-93114-3

[54] K. Värbu, N. Muhammad, and Y. Muhammad, "Past, present, and future of EEG-based BCI applications," vol. 22, no. 9, p. 3331. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9101004/

[55] S. L. Frank, L. J. Otten, G. Galli, and G. Vigliocco, "The ERP response to the amount of information conveyed by words in sentences," *Brain and Language*, vol. 140, pp. 1–11, Jan. 2015.

[56] J. Hale, C. Dyer, A. Kuncoro, and J. R. Brennan, "Finding Syntax in Human Encephalography with Beam Search," *arXiv:1806.04127 [cs]*, Jun. 2018, arXiv: 1806.04127.

[57] W. Cui *et al.*, "Neuro-gpt: Towards a foundation model for eeg," *arXiv preprint arXiv:2311.03764*, November 2023. [Online]. Available: https://arxiv.org/abs/2311.03764

[58] D. Kostas, S. Aroca-Ouellette, and F. Rudzicz, "BENDR: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data," vol. 15, 2021.

[59] F. R. Willett, E. M. Kunz, C. Fan *et al.*, "A high-performance speech neuroprosthesis," *Nature*, vol. 620, pp. 1031–1036, August 2023.

[60] A. J. Reddy and L. Wehbe, "Can fMRI reveal the representation of syntactic structure in the brain?" in *Advances in Neural Information Processing Systems*, vol. 34. Curran Associates, Inc., 2021, pp. 9843–9856.

[61] C. Caucheteux and J.-R. King, "Language processing in brains and deep neural networks: computational convergence and its limits," bioRxiv, Tech. Rep., Jan. 2021, section: New Results Type: article.

[62] C. Shain, H. Kean, B. Lipkin, J. Affourtit, M. Siegelman, F. Mollica, and E. Fedorenko, "'Constituent length' effects in fMRI do not provide evidence for abstract syntactic processing," Neuroscience, preprint, Nov. 2021.

[63] D. J. Acunzo, D. M. Low, and S. L. Fairhall, "Deep neural networks reveal topic-level representations of sentences in medial prefrontal cortex, lateral anterior temporal lobe, precuneus, and angular gyrus," *NeuroImage*, vol. 251, p. 119005, May 2022.

[64] A. Søgaard, "Evaluating word embeddings with fMRI and eye-tracking," in *Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP*. Berlin, Germany: Association for Computational Linguistics, Aug. 2016, pp. 116–121.

[65] A. Fyshe, P. P. Talukdar, B. Murphy, and T. M. Mitchell, "Interpretable semantic vectors from a joint model of brain- and text- based meaning," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Baltimore, Maryland: Association for Computational Linguistics, Jun. 2014, pp. 489–499.

[66] Y. Moshfeghi, "Neurasearch: Neuroscience and information retrieval," *CEUR Workshop Proceedings*, vol. 2950, pp. 193–194, Sep. 2021, presented at: DESIRES 2021, Design of Experimental Search Information REtrieval Systems; Proceedings of the Second International Conference on Design of Experimental Search Information REtrieval Systems, Padova, Italy, September 15-18, 2021.; 2nd International Conference on Design of Experimental Search and Information REtrieval Systems, DESIRES 2021 ; Conference date: 15-09-2021 Through 18-09-2021.

[67] V. Mueller, Y. Brehmer, T. Von Oertzen, S.-C. Li, and U. Lindenberger, "Electrophysiological correlates of selective attention: a lifespan comparison," *BMC neuroscience*, vol. 9, pp. 1–21, 2008.

[68] G. R. Müller-Putz, R. Riedl, S. C Wriessnegger *et al.*, "Electroencephalography (eeg) as a research tool in the information systems discipline: Foundations, measurement, and applications," *Communications of the Association for Information Systems*, vol. 37, no. 1, p. 46, 2015.

[69] J. Gwizdka, Y. Moshfeghi, and M. L. Wilson, "Introduction to the special issue on neuro-information science," *Journal of the Association for Information Science and Technology*, vol. 70, pp. 911–916, 2019.

[70] R. Martínez-Castaño, A. Htait, L. Azzopardi, and Y. Moshfeghi, *Early Risk Detection of Self-Harm Using BERT-Based Transformers.* Cham: Springer International Publishing, 2022, pp. 183–206. [Online]. Available: https://doi.org/10.1007/978-3-031-04431-1_8

[71] Y. Moshfeghi and J. M. Jose, "An effective implicit relevance feedback technique using affective, physiological and behavioural features," in *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 133–142.

[72] Y. Moshfeghi, L. R. Pinto, F. E. Pollick, and J. M. Jose, "Understanding relevance: An fmri study," in *Advances in Information Retrieval: 35th European Conference on IR Research, ECIR 2013, Moscow, Russia, March 24-27, 2013. Proceedings 35.* Springer, 2013, pp. 14–25.

[73] Y. Moshfeghi, P. Triantafillou, and F. E. Pollick, "Understanding information need: An fMRI study," in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, ser. SIGIR '16. Association for Computing Machinery, 2016, pp. 335–344.

[74] Y. Moshfeghi, P. Triantafillou, and F. Pollick, "Towards predicting a realisation of an information need based on brain signals," in *The World Wide Web Conference*, ser. WWW '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1300–1309.

[75] Y. Moshfeghi and F. E. Pollick, "Neuropsychological model of the realization of information need," *Journal of the Association for Information Science and Technology*, vol. 70, no. 9, pp. 954–967, 2019.

[76] N. McGuire and Y. Moshfeghi, "Prediction of the realisation of an information need: An EEG study," in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '24. Association for Computing Machinery, pp. 2584–2588. [Online]. Available: https://dl.acm.org/doi/10.1145/3626772.3657981

[77] D. Michalkova, M. P. Rodriguez, and Y. Moshfeghi, "Understanding feeling-of-knowing in information search: An eeg study," *ACM Trans. Inf. Syst.*, vol. 42, no. 3, jan 2024.

[78] L. Kangassalo, M. Spapé, G. Jacucci, and T. Ruotsalo, "Why do users issue good queries? neural correlates of term specificity," in *Proceedings of the 42nd international acm sigir conference on research and development in information retrieval*, 2019, pp. 375–384.

[79] Y. Moshfeghi and F. E. Pollick, "Search process as transitions between neural states," in *Proceedings of the 2018 World Wide Web Conference*, ser. WWW '18. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2018, p. 1683–1692.

[80] M. Allegretti, Y. Moshfeghi, M. Hadjigeorgieva, F. E. Pollick, J. M. Jose, and G. Pasi, "When relevance judgement is happening? an EEG-based study," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '15. Association for Computing Machinery, 2015, pp. 719–722.

[81] M. J. Eugster, T. Ruotsalo, M. M. Spapé, O. Barral, N. Ravaja, G. Jacucci, and S. Kaski, "Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals," *Scientific reports*, vol. 6, no. 1, p. 38580, 2016.

[82] H. H. Kim and Y. H. Kim, "Erp/mmr algorithm for classifying topic-relevant and topic-irrelevant visual shots of documentary videos," *Journal of the Association for Information Science and Technology*, vol. 70, no. 9, pp. 931–941, 2019.

[83] Z. Pinkosova, W. J. McGeown, and Y. Moshfeghi, "The cortical activity of graded relevance," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 299–308.

[84] Z. "Pinkosova, W. J. McGeown, and Y. Moshfeghi, "Revisiting neurological aspects of relevance: An eeg study," in *International Conference on Machine Learning, Optimization, and Data Science*. Springer, 2022, pp. 549–563.

[85] P. Zuzana, M. J, and M. Yashar", "Moderating effects of self-perceived knowledge in a relevance assessment task: An eeg study," *Computers in Human Behavior Reports*, vol. 11, p. 100295, 2023.

[86] S. Paisalnan, Y. Moshfeghi, and F. Pollick, "Neural correlates of realisation of satisfaction in a successful search process," *Proceedings of the Association for Information Science and Technology*, vol. 58, no. 1, pp. 282–291, 2021.

[87] S. Paisalnan, F. Pollick, and Y. Moshfeghi, "Neural correlates of satisfaction of an information need," in *Machine Learning, Optimization, and Data Science*, G. Nicosia, V. Ojha, E. La Malfa, G. La Malfa, P. Pardalos, G. Di Fatta, G. Giuffrida, and R. Umeton, Eds. Cham: Springer Nature Switzerland, 2023, pp. 443–457.

[88] K. Kingphai and Y. Moshfeghi, "Mental workload prediction level from eeg signals using deep learning models," Sep. 2021, the 3rd Neuroergonomics Conference 2021, NEC21 ; Conference date: 11-09-2021 Through 16-09-2021.

[89] K. "Kingphai and Y. Moshfeghi, "On EEG preprocessing role in deep learning effectiveness for mental workload classification," in *Human Mental Workload: Models and Applications*, ser. Communications in Computer and Information Science, L. Longo and M. C. Leva, Eds. Springer International Publishing, 2021, pp. 81–98.

[90] K. Kingphai and Y. Moshfeghi, "On time series cross-validation for deep learning classification model of mental workload levels based on eeg signals," in *Machine Learning, Optimization, and Data Science*. Cham: Springer Nature Switzerland, 2023, pp. 402–416.

[91] N. McGuire and Y. Moshfeghi, "On ensemble learning for mental workload classification," in *Machine Learning, Optimization, and Data Science*, G. Nicosia, V. Ojha, E. La Malfa, G. La Malfa, P. M. Pardalos, and R. Umeton, Eds. Cham: Springer Nature Switzerland, 2024, pp. 358–372.

[92] J.-P. Kauppi, M. Kandemir, V.-M. Saarinen, L. Hirvenkari, L. Parkkonen, A. Klami, R. Hari, and S. Kaski, "Towards brain-activity-controlled information retrieval: Decoding image relevance from meg signals," *NeuroImage*, vol. 112, pp. 288–298, 2015.

[93] S. Paisalnan, F. Pollick, and Y. Moshfeghi, "Towards understanding neuroscience of realisation of information need in light of relevance and satisfaction judgement," in *Machine Learning, Optimization, and Data Science*, G. Nicosia, V. Ojha, E. La Malfa, G. La Malfa, G. Jansen, P. M. Pardalos, G. Giuffrida, and R. Umeton, Eds. Cham: Springer International Publishing, 2022, pp. 41–56.

[94] Z. Lamprou, F. Pollick, and Y. Moshfeghi, "Role of punctuation in semantic mapping between brain and transformer models," in *Machine Learning, Optimization, and Data Science*, G. Nicosia, V. Ojha, E. La Malfa, G. La Malfa, P. Pardalos, G. Di Fatta, G. Giuffrida, and R. Umeton, Eds. Cham: Springer Nature Switzerland, 2023, pp. 458–472.

[95] J. Gwizdka, R. Hosseini, M. Cole, and S. Wang, "Temporal dynamics of eye-tracking and eeg during reading and relevance decisions," *Journal of the Association for Information Science and Technology*, vol. 68, no. 10, pp. 2299–2312, 2017.

[96] G. Jacucci, O. Barral, P. Daee, M. Wenzel, B. Serim, T. Ruotsalo, P. Pluchino, J. Freeman, L. Gamberini, S. Kaski *et al.*, "Integrating neurophysiologic relevance feedback in intent modeling for information retrieval," *Journal of the Association for Information Science and Technology*, vol. 70, no. 9, pp. 917–930, 2019.

[97] A. Jimenez-Molina, C. Retamal, and H. Lira, "Using psychophysiological sensors to assess mental workload during web browsing," *Sensors*, vol. 18, no. 2, p. 458, 2018.

[98] N. McGuire and Y. Moshfeghi, "What song am i thinking of?" in *Machine Learning, Optimization, and Data Science*, G. Nicosia, V. Ojha, E. La Malfa, G. La Malfa, P. M. Pardalos, and R. Umeton, Eds. Cham: Springer Nature Switzerland, 2024, pp. 418–432.

[99] A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Łukasz Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017. [Online]. Available: https://arxiv.org/abs/1706.03762

[100] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021. [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy

[101] G. Chochlakis *et al.*, "Vault: Augmenting the vision-and-language transformer for sentiment classification on social media," *arXiv preprint arXiv:2208.09021*, 2023. [Online]. Available: https://arxiv.org/abs/2208.09021

[102] B. Sandwidi and S. P. Mukkolakal, "Transformers-based approach for a sustainability term-based sentiment analysis (stbsa)," in *Proceedings of the Second Workshop on NLP for Positive Impact (NLP4PI)*. Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics, 2022, pp. 157–170. [Online]. Available: https://aclanthology.org/2022.nlp4pi-1.19

[103] A. S. Team, "Taming transformers for text classification with millions of classes," in *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2023. [Online]. Available: https://www.amazon.science/blog/natural-language-processing-techniques-text-classification-with-transformers-at-scale

[104] M. Tuteja and D. G. Juclà, "Long text classification using transformers with paragraph selection strategies," in *Proceedings of the Natural Legal Language Processing Workshop 2023*. Singapore: Association for Computational Linguistics, 2023, pp. 17–24. [Online]. Available: https://aclanthology.org/2023.nllp-1.3

[105] A. Scherp, "Transformers are short text classifiers: A study of inductive short text classifiers on benchmarks and real-world datasets," *arXiv preprint arXiv:2211.16878*, 2022. [Online]. Available: https://arxiv.org/abs/2211.16878

[106] I. Silfverskiöld, "Fine-tune smaller transformer models: Text classification," *Towards Data Science*, 2023. [Online]. Available: https://towardsdatascience.com/fine-tune-smaller-transformer-models-text-classification-77cbbd3bf02b

[107] R. Koncel-Kedziorski, D. Bekal, Y. Luan, M. Lapata, and H. Hajishirzi, "Text generation from knowledge graphs with graph transformers," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2019, pp. 1957–1967. [Online]. Available: https://aclanthology.org/P19-1190.pdf

[108] L. Gong, J. Crego, and J. Senellart, "Enhanced transformer model for data-to-text generation," in *Proceedings of the 3rd Workshop on Neural Generation and*

Translation. Hong Kong: Association for Computational Linguistics, 2019, pp. 148–156. [Online]. Available: https://aclanthology.org/D19-5615

[109] A. Author and B. Author, "Knowledge based transformer model for information retrieval," *HAL*, 2021. [Online]. Available: https://hal.science/hal-03263784

[110] J. Qadrud-Din *et al.*, "Transformer based language models for similar text retrieval and ranking," *arXiv preprint arXiv:2005.04588*, 2020. [Online]. Available: https://arxiv.org/abs/2005.04588

[111] A. Zhang, "Pretrained transformers for efficient and robust information retrieval," Ph.D. dissertation, University of Waterloo, 2024. [Online]. Available: https://uwspace.uwaterloo.ca/items/05e9d9fb-dac1-4ff8-ae04-a9a6a83933c5

[112] P. Lewis, E. Perez, A. Piktus, F. Petroni, Y. Wu, T. Rocktäschel, S. Ruder, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive nlp tasks," in *Proceedings of the 38th International Conference on Machine Learning*. PMLR, 2021, pp. 10 078–10 090. [Online]. Available: http://proceedings.mlr.press/v139/lewis21a.html

[113] S. Antol, J. Huang, M. Mitchell, M. Paluri, D. Parikh, and P. Young, "Vqa: Visual question answering," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 2425–2433.

[114] N. R. Group, "Next-gpt: Any-to-any multimodal llm," *arXiv preprint arXiv:2304.12345*, 2023. [Online]. Available: https://arxiv.org/abs/2304.12345

[115] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. [Online]. Available: https://ieeexplore.ieee.org/document/726791

[116] V. Nair and G. Hinton, "Rectified linear units improve restricted boltzmann machines," *Proceedings of the 27th International Conference on Machine Learning (ICML)*, pp. 807–814, 2010. [Online]. Available: http://www.cs.toronto.edu/~hinton/absps/reluicml.pdf

[117] M. Goyal, R. Goyal, and B. Lall, "Learning activation functions: A new paradigm for understanding neural networks," 2020.

[118] B. Bilonoh, Y. Bodyanskiy, B. Kolchygin, and S. Mashtalir, "Tunable activation functions for deep neural networks," in *Lecture Notes in Computational Intelligence and Decision Making*, ser. Lecture Notes on Data Engineering and Communications Technologies, S. Babichev and V. Lytvynenko, Eds. Springer International Publishing, 2022, pp. 624–633.

[119] J. Wang, L. Chen, and C. W. W. Ng, "A new class of polynomial activation functions of deep learning for precipitation forecasting," in *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, ser. WSDM '22. Association for Computing Machinery, 2022, pp. 1025–1035.

[120] "Evaluation of chatgpt's responses to information needs and information seeking of dementia patients," *Scientific Reports*, 2024. [Online]. Available: https://www.nature.com/articles/s41598-024-61068-5

[121] "Assessing the quality of chatgpt responses to dementia caregivers," *PMC*, 2024. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11089887/

[122] "Evaluation of chatgpt's responses to information needs and information seeking of dementia patients," 2024. [Online]. Available: https://www.researchgate.net/publication/380347377_Evaluation_of_ChatGPT's_responses_to_information_needs_and_information_seeking_of_dementia_patients

[123] "Use of search engines for the retrieval of scholarly information," *ResearchGate*, 2014. [Online]. Available: https://www.researchgate.net/publication/279218885_Use_of_Search_Engines_for_the_Retrieval_of_Scholarly_Information_A_study_of_the_Kerala_University_Library

[124] "Google and the scholar: The role of google in scientists' information-seeking behaviour," *ResearchGate*, 2010. [Online]. Available: https://www.researchgate.net/publication/220207532_Google_and_the_scholar_The_role_of_Google_in_scientists'_information-seeking_behaviour

[125] "Use of search engines as predictors of research skills of postgraduate students," *Digital Commons @ University of Nebraska - Lincoln.* [Online]. Available: https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=15156&context=libphilprac

[126] Y. Moshfeghi, "Intelligent rss aggregator," publisher: Citeseer. [Online]. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c3e16c674c1034d03ce493da5cf51b532b2d7803

[127] Y. Moshfeghi, D. Agarwal, B. Piwowarski, and J. M. Jose, "Movie recommender: Semantically enriched unified relevance model for rating prediction in collaborative filtering," in *Advances in Information Retrieval*, M. Boughanem, C. Berrut, J. Mothe, and C. Soule-Dupuy, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 54–65.

[128] P. J. McParlane, Y. Moshfeghi, and J. M. Jose, "On contextual photo tag recommendation," in *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 965–968. [Online]. Available: https://doi.org/10.1145/2484028.2484160

[129] I. Paun, Y. Moshfeghi, and N. Ntarmos, "Are we there yet? estimating training time for recommendation systems," in *Proceedings of the 1st Workshop on Machine Learning and Systems*, ser. EuroMLSys '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 39–47. [Online]. Available: https://doi.org/10.1145/3437984.3458832

[130] I. Arapakis, Y. Moshfeghi, H. Joho, R. Ren, D. Hannah, and J. M. Jose, "Enriching user profiling with affective features for the improvement of a multimodal recommender system," in *Proceedings of the ACM International Conference on Image and Video Retrieval*, ser. CIVR '09. New York, NY, USA: Association for Computing Machinery, 2009. [Online]. Available: https://doi.org/10.1145/1646396.1646433

[131] I. "Arapakis, Y. Moshfeghi, H. Joho, R. Ren, D. Hannah, and J. M. Jose, "Integrating facial expressions into user profiling for the improvement of a multimodal recommender system," in *2009 IEEE International Conference on Multimedia and Expo*, 2009, pp. 1440–1443.

[132] Y. Moshfeghi, B. Piwowarski, and J. M. Jose, "Handling data sparsity in collaborative filtering using emotion and semantic based features," in *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 625–634. [Online]. Available: https://doi.org/10.1145/2009916.2010001

[133] Y. Moshfeghi and J. M. Jose, "An effective implicit relevance feedback technique using affective, physiological and behavioural features," in *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 133–142. [Online]. Available: https://doi.org/10.1145/2484028.2484074

[134] D. Michalkova, M. Parra-Rodriguez, and Y. Moshfeghi, "Information need awareness: An EEG study," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2022, pp. 610–621. [Online]. Available: https://dl.acm.org/doi/10.1145/3477495.3531999

[135] S. Whiting, Y. Moshfeghi, and J. M. Jose, "Exploring term temporality for pseudo-relevance feedback," in *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 1245–1246. [Online]. Available: https://doi.org/10.1145/2009916.2010141

[136] N. McGuire and Y. Moshfeghi, "Deeper: Dense electroencephalography passage retrieval," 2024. [Online]. Available: https://arxiv.org/abs/2412.06695

[137] L. R. Pinto, Y. Moshfeghi, F. E. Pollick, and J. M. Jose, "Neural features of video topical relevance," p. 51. [Online]. Available: http://www.neurois.org/papers/2014/2014%20Proceedings%20Gmunden%20Retreat%20on%20NeuroIS%20paper%2028.pdf

[138] Y. Moshfeghi and F. E. Pollick, "Search process as transitions between neural states," in *Proceedings of the 2018 World Wide Web Conference*, ser. WWW '18. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2018, p. 1683–1692. [Online]. Available: https://doi.org/10.1145/3178876.3186080

[139] M. A. Belabbes, I. Ruthven, Y. Moshfeghi, and D. R. Pennington, "Information overload: a concept analysis," vol. 79, no. 1, pp. 144–159, publisher: Emerald Publishing Limited. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/jd-06-2021-0118/full/html

[140] N. Hollenstein, J. Rotsztejn, M. Troendle, A. Pedroni, C. Zhang, and N. Langer, "ZuCo, a simultaneous EEG and eye-tracking resource for natural sentence reading," vol. 5, no. 1, p. 180291, 2018.

[141] N. Hollenstein, M. Troendle, C. Zhang, and N. Langer, "ZuCo 2.0: A dataset of physiological recordings during natural reading and annotation," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*. European Language Resources Association, 2020, pp. 138–146.

[142] U. Cop, N. Dirix, D. Drieghe, and W. Duyck, "Presenting geco: An eyetracking corpus of monolingual and bilingual sentence reading," *Behavior research methods*, vol. 49, pp. 602–615, 2017.

[143] N. Hollenstein, M. Tröndle, M. Plomecka, S. Kiegeland, Y. Özyurt, L. A. Jäger, and N. Langer, "Reading task classification using EEG and eye-tracking data," 2021.

[144] C. Herff, D. Heger, A. De Pesters, D. Telaar, P. Brunner, G. Schalk, and T. Schultz, "Brain-to-text: decoding spoken phrases from phone representations in the brain," *Frontiers in neuroscience*, vol. 9, p. 217, 2015.

[145] Z. Wang and H. Ji, "Open vocabulary electroencephalography-to-text decoding and zero-shot sentiment classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 5, 2022, pp. 5350–5358.

[146] D. Truong, M. Milham, S. Makeig, and A. Delorme, "Deep convolutional neural network applied to electroencephalography: Raw data vs spectral features," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, pp. 1039–1042.

[147] G. Siddhad, A. Gupta, D. P. Dogra, and P. P. Roy, "Efficacy of transformer networks for classification of eeg data," *Biomedical Signal Processing and Control*, vol. 87, p. 105488, Jan 2024.

[148] K. Kingphai and Y. Moshfeghi, "On channel selection for EEG-based mental workload classification," in *Machine Learning, Optimization, and Data Science*, G. Nicosia, V. Ojha, E. La Malfa, G. La Malfa, P. M. Pardalos, and R. Umeton, Eds. Springer Nature Switzerland, pp. 403–417.

[149] G. Buzsaki *et al.*, "Large-scale recording of neuronal ensembles," *Nature Reviews Neuroscience*, vol. 5, pp. 471–481, 2004. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10752959/

[150] J. Yoo *et al.*, "Offline ensemble co-reactivation links memories across days," *Nature*, 2024. [Online]. Available: https://www.nature.com/articles/s41586-024-08168-4

[151] R. Plucińska, K. Jedrzejewski, U. Malinowska, and J. Rogala, "Leveraging multiple distinct EEG training sessions for improvement of spectral-based biometric verification results," vol. 23, no. 4, p. 2057. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9963573/

[152] A. Sharma, J. Nigam, A. Rathore, and A. Bhavsar, "EEG classification for visual brain decoding with spatio-temporal and transformer based paradigms." [Online]. Available: http://arxiv.org/abs/2406.07153

[153] R. Mehta, P. Rajpura, and Y. K. Meena, "Can EEG resting state data benefit data-driven approaches for motor-imagery decoding?" [Online]. Available: http://arxiv.org/abs/2411.09789

[154] A. Gourav *et al.*, "Multi-modal retrieval for large language model based speech recognition," *Findings of the Association for Computational Linguistics ACL 2024*, pp. 4435–4446, 2024. [Online]. Available: https://aclanthology.org/2024.findings-acl.262

[155] Y. Liu *et al.*, "Evaluating speech recognition performance towards large language models," *Interspeech*, 2024. [Online]. Available: https://www.isca-archive.org/interspeech_2024/liu24c_interspeech.pdf

[156] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 2227–2237.

[157] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 328–339.

[158] X. Zhu, T. Li, and G. de Melo, "Exploring semantic properties of sentence embeddings," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 632–637.

[159] A. Conneau, G. Kruszewski, G. Lample, L. Barrault, and M. Baroni, "What you can cram into a single $&!#* vector: Probing sentence embeddings for linguistic properties," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 2126–2136.

[160] T. Linzen, E. Dupoux, and Y. Goldberg, "Assessing the Ability of LSTMs to Learn Syntax-Sensitive Dependencies," *arXiv:1611.01368 [cs]*, Nov. 2016, arXiv: 1611.01368.

[161] H. Peng, R. Schwartz, S. Thomson, and N. A. Smith, "Rational Recurrences," *arXiv:1808.09357 [cs]*, Aug. 2018, arXiv: 1808.09357.

[162] Y. Chen, S. Gilroy, A. Maletti, J. May, and K. Knight, "Recurrent neural networks as weighted language recognizers," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 2261–2271.

[163] G. Weiss, Y. Goldberg, and E. Yahav, "On the Practical Computational Power of Finite Precision RNNs for Language Recognition," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 740–745.

[164] I. Beltagy, K. Lo, and A. Cohan, "SciBERT: A Pretrained Language Model for Scientific Text," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 3615–3620.

[165] Q. Jia, J. Li, Q. Zhang, X. He, and J. Zhu, "RMBERT: News Recommendation via Recurrent Reasoning Memory Network over BERT," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in*

Information Retrieval, ser. SIGIR '21. New York, NY, USA: Association for Computing Machinery, Jul. 2021, pp. 1773–1777.

[166] W. Hong, K. Ji, J. Liu, J. Wang, J. Chen, and W. Chu, "GilBERT: Generative Vision-Language Pre-Training for Image-Text Retrieval," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '21. New York, NY, USA: Association for Computing Machinery, Jul. 2021, pp. 1379–1388.

[167] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter."

[168] L. Zhuang, L. Wayne, S. Ya, and Z. Jun, "A robustly optimized BERT pre-training approach with post-training," in *Proceedings of the 20th Chinese National Conference on Computational Linguistics*. Huhhot, China: Chinese Information Processing Society of China, Aug. 2021, pp. 1218–1227.

[169] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," Apr. 2020.

[170] K. Clark, M.-T. Luong, Q. V. Le, and C. D. Manning, "ELECTRA: Pre-training text encoders as discriminators rather than generators."

[171] D. Cer, Y. Yang, S.-y. Kong, N. Hua, N. Limtiaco, R. St. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, B. Strope, and R. Kurzweil, "Universal sentence encoder for English," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Brussels, Belgium: Association for Computational Linguistics, Nov. 2018, pp. 169–174.

[172] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. V. Le, and R. Salakhutdinov, "Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context," *arXiv:1901.02860 [cs, stat]*, Jun. 2019, arXiv: 1901.02860.

[173] N. Moore, "What's the point? The role of punctuation in realising information structure in written English," *Functional Linguistics*, vol. 3, no. 1, p. 6, May 2016.

[174] J. K. Rowling, *Harry Potter and the Philosopher's Stone*, 1st ed. London: Bloomsbury Publishing, June 1997, vol. 1.

[175] G. Sudre, D. Pomerleau, M. Palatucci, L. Wehbe, A. Fyshe, R. Salmelin, and T. Mitchell, "Tracking neural coding of perceptual and semantic features of concrete nouns," *NeuroImage*, vol. 62, no. 1, pp. 451–463, Aug. 2012.

[176] S. Nishimoto, A. T. Vu, T. Naselaris, Y. Benjamini, B. Yu, and J. L. Gallant, "Reconstructing visual experiences from brain activity evoked by natural movies," *Current Biology*, vol. 21, no. 19, pp. 1641–1646, Oct. 2011.

[177] T. Naselaris, K. N. Kay, S. Nishimoto, and J. L. Gallant, "Encoding and decoding in fmri," *Neuroimage*, vol. 56, no. 2, pp. 400–410, 2011.

[178] N. Kriegeskorte and P. K. Douglas, "Cognitive computational neuroscience," *Nature neuroscience*, vol. 21, no. 9, pp. 1148–1160, 2018.

[179] Y. Duan, J. Zhou, Z. Wang, Y.-K. Wang, and C.-T. Lin, "DeWave: Discrete EEG waves encoding for brain dynamics to text translation," 2023.

[180] T. Szandała, "Review and comparison of commonly used activation functions for deep neural networks," in *Bio-inspired neurocomputing*. Springer, 2020, pp. 203–224.

[181] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," *Neurocomputing*, vol. 503, pp. 92–108, 2022.

[182] M. Murray, V. Abrol, and J. Tanner, "Activation function design for deep networks: linearity and effective initialisation," *Applied and Computational Harmonic Analysis*, vol. 59, pp. 117–154, 2022.

[183] S. Hayou, A. Doucet, and J. Rousseau, "On the impact of the activation function on deep neural networks training," in *International conference on machine learning*. PMLR, 2019, pp. 2672–2680.

[184] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, Eds. Online: Association for Computational Linguistics, Jul. 2020, pp. 7871–7880. [Online]. Available: https://aclanthology.org/2020.acl-main.703/

[185] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions."

[186] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)."

[187] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)."

[188] A. L. Maas, "Rectifier nonlinearities improve neural network acoustic models."

[189] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification."

[190] C.-Y. Lin, "ROUGE: A package for automatic evaluation of summaries," in *Text Summarization Branches Out*. Barcelona, Spain: Association for Computational Linguistics, Jul. 2004, pp. 74–81. [Online]. Available: https://aclanthology.org/W04-1013

[191] J. Dockès, R. Poldrack, R. Primet, H. Gözükan, T. Yarkoni, F. Suchanek, B. Thirion, and G. Varoquaux, "NeuroQuery: comprehensive meta-analysis of human brain mapping," 2020. [Online]. Available: http://arxiv.org/abs/2002.09261

[192] J. Dockès, D. Wassermann, R. Poldrack, F. Suchanek, B. Thirion, and G. Varoquaux, "Text to brain: predicting the spatial distribution of

neuroimaging observations from text reports," 2018. [Online]. Available: http://arxiv.org/abs/1806.01139

[193] J. D. Power, D. A. Fair, B. L. Schlaggar, and S. E. Petersen, "The development of human functional brain networks," vol. 67, no. 5, pp. 735–748, 2010. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2941973/

[194] J. Power, A. Cohen, S. Nelson, G. Wig, K. Barnes, J. Church, A. Vogel, T. Laumann, F. Miezin, B. Schlaggar, and S. Petersen, "Functional network organization of the human brain," vol. 72, no. 4, pp. 665–678, 2011. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0896627311007926

[195] T. D. Wager, M. Lindquist, and L. Kaplan, "Meta-analysis of functional neuroimaging data: current and future directions," vol. 2, no. 2, pp. 150–158, 2007. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2555451/

[196] G. Hein and R. T. Knight, "Superior temporal sulcus—it's my area: Or is it?" vol. 20, no. 12, pp. 2125–2136, 2008. [Online]. Available: https://www.mitpressjournals.org/doi/abs/10.1162/jocn.2008.20148

[197] L. J. Chang, T. Yarkoni, M. W. Khaw, and A. G. Sanfey, "Decoding the role of the insula in human cognition: Functional parcellation and large-scale reverse inference," vol. 23, no. 3, pp. 739–749, 2013. [Online]. Available: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhs065

[198] D. Schwartz, M. Toneva, and L. Wehbe, "Inducing brain-relevant bias in natural language processing models," 2019. [Online]. Available: http://arxiv.org/abs/1911.03268

[199] N. Nieto, V. Peterson, H. L. Rufiner, J. E. Kamienkowski, and R. Spies, "Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition," vol. 9, no. 1, p. 52, 2022, number: 1 Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41597-022-01147-2

[200] A. Meta, "Introducing meta llama 3: The most capable openly available llm to date," *Meta AI*, 2024.

[201] OpenAI, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023. [Online]. Available: https://arxiv.org/abs/2303.08774

[202] S. Black, G. Leo, P. Wang, C. Leahy, and S. Biderman, "GPT-Neo: Large Scale Autoregressive Language Modeling with Mesh-Tensorflow," Mar. 2021, If you use this software, please cite it using these metadata. [Online]. Available: https://doi.org/10.5281/zenodo.5297715

[203] S. Roller, "Recipes for building an open-domain chatbot," *arXiv preprint arXiv:2004.13637*, 2020.

[204] M. Zaheer, G. Guruganesh, A. Dubey, J. Ainslie, C. Alberti, S. Ontanon, P. Pham, A. Ravula, Q. Wang, L. Yang, and A. Ahmed, "Big bird: Transformers for longer sequences," 2021.

[205] Y. Yan, W. Qi, Y. Gong, D. Liu, N. Duan, J. Chen, R. Zhang, and M. Zhou, "Prophetnet: Predicting future n-gram for sequence-to-sequence pre-training," *arXiv preprint arXiv:2001.04063*, 2020.

[206] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, vol. 21, no. 140, pp. 1–67, 2020. [Online]. Available: http://jmlr.org/papers/v21/20-074.html

[207] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks," in *Proceedings of the 23rd International Conference on Machine Learning (ICML)*. ACM, 2006, pp. 369–376. [Online]. Available: https://www.cs.toronto.edu/~graves/icml_2006.pdf

[208] Y. Wang and et al., "Residual convolutional ctc networks for automatic speech recognition," *arXiv preprint arXiv:1702.07793*, 2017. [Online]. Available: https://arxiv.org/abs/1702.07793

[209] A. Graves, A. R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 6645–6649. [Online]. Available: https://ieeexplore.ieee.org/document/6638947

[210] S. Singh, A. Gupta, A. Maghan, D. Gowda, S. Singh, and C. Kim, "Comparative study of different tokenization strategies for streaming end-to-end asr," in *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2021, pp. 388–394.

[211] T. Bañeras-Roux, M. Rouvier, J. Wottawa, and R. Dufour, "A comprehensive analysis of tokenization and self-supervised learning in end-to-end automatic speech recognition applied on french language," in *32th European Signal Processing Conference (EUSIPCO)*, 2024.

[212] Y. Song, Q. Zheng, B. Liu, and X. Gao, "Eeg conformer: Convolutional transformer for eeg decoding and visualization," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 710–719, 2023.

[213] T. B. Brown, B. H. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, G. Sastry, A. Askell, S. Agarwal, and et al., "Language models are few-shot learners," *arXiv preprint arXiv:2005.14165*, 2020. [Online]. Available: https://arxiv.org/abs/2005.14165

[214] Z. Lei, X. Na, M. Xu, E. Pusateri, C. V. Gysel, Y. Zhang, S. Han, and Z. Huang, "Contextualization of asr with llm using phonetic retrieval-based augmentation," in *Proceedings of the 2024 Conference on Machine Learning*, 2024. [Online]. Available: https://machinelearning.apple.com/research/asr-contextualization

[215] Y. e. a. Zhao, "Generative large language models for detection of speech recognition errors in radiology reports," *Radiology*, January 2024. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10982816/

[216] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *International Conference on Learning Representations (ICLR)*, 2015. [Online]. Available: https://arxiv.org/abs/1409.0473

[217] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," in *Proceedings of the 8th Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8)*, 2014, pp. 103–111. [Online]. Available: https://aclanthology.org/W14-4014

[218] I. Sutskever, O. Vinyals, and Q. Le, "Sequence to sequence learning with neural networks," in *Advances in Neural Information Processing Systems (NIPS)*, 2014, pp. 3104–3112. [Online]. Available: https://arxiv.org/abs/1409.3215

[219] W. L. Lim, "STEW: Simultaneous task EEG workload dataset." [Online]. Available: https://ieee-dataport.org/open-access/stew-simultaneous-task-eeg-workload-dataset

[220] J. J. Bird, L. J. Manso, E. P. Ribeiro, A. Ekárt, and D. R. Faria, "A study on mental state classification using EEG-based brain-machine interface," in *2018 International Conference on Intelligent Systems (IS)*, pp. 795–800, ISSN: 1541-1672. [Online]. Available: https://ieeexplore.ieee.org/document/8710576

[221] D. R. Edla, K. Mangalorekar, G. Dhavalikar, and S. Dodia, "Classification of EEG data for human mental state analysis using random forest classifier," vol. 132, pp. 1523–1532. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050918308482

[222] N. McGuire and Y. Moshfeghi, "On ensemble learning for mental workload classification," in *Machine Learning, Optimization, and Data Science*, G. Nicosia,

V. Ojha, E. La Malfa, G. La Malfa, P. M. Pardalos, and R. Umeton, Eds. Springer Nature Switzerland, pp. 358–372.

[223] K. Kingphai and Y. Moshfeghi, "Mental workload assessment using deep learning models from EEG signals: A systematic review," pp. 1–27, conference Name: IEEE Transactions on Cognitive and Developmental Systems. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10680430

[224] P. Sarma and S. Barma, "Review on stimuli presentation for affect analysis based on EEG," vol. 8, pp. 51 991–52 009, conference Name: IEEE Access. [Online]. Available: https://ieeexplore.ieee.org/document/9036897/?arnumber=9036897

[225] A. Özbeyaz and M. Korkmaz, "Stipresoft: An alternative stimuli presentation software synchronizing with current acquisition systems in eeg experiments," *SN Applied Sciences*, vol. 1, no. 5, pp. 1–10, 2019. [Online]. Available: https://link.springer.com/article/10.1007/s42452-019-1683-x

[226] X. Zhu, W. Rong, L. Zhao, Z. He, Q. Yang, J. Sun, and G. Liu, "Eeg emotion classification network based on attention fusion of multi-channel band features," *Sensors*, vol. 22, no. 14, p. 5252, 2022.

[227] S. ten Oever, T. A. de Graaf, C. Bonnemayer, J. Ronner, A. T. Sack, and L. Riecke, "Stimulus presentation at specific neuronal oscillatory phases experimentally controlled with tACS: Implementation and applications," vol. 10, publisher: Frontiers. [Online]. Available: https://www.frontiersin.org/journals/cellular-neuroscience/articles/10.3389/fncel.2016.00240/full

[228] M. Bola, P. Orłowski, K. Baranowska, M. Schartner, and A. Marchewka, "Informativeness of auditory stimuli does not affect EEG signal diversity," vol. 9, publisher: Frontiers. [Online]. Available: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2018.01820/full

[229] Using AI to decode speech from brain activity. [Online]. Available: https://ai.meta.com/blog/ai-speech-brain-activity/

[230] V. V. C. F. University, "Development of students' visual thinking based on the use of sketchnoting techniques," in *Proceedings of the IV International Scientific and Practical Conference "Modern Management Trends and Digital Economy: from Regional Development to Global Economic Growth" (MTDE 2022)*, vol. 141. EDP Sciences, 2022. [Online]. Available: https://doi.org/10.1051/shsconf/202214103005

[231] W. A. Wicaksono, "Visual thinking and reading comprehension: Foreign language setting as an example," *International Journal of Educational Management*, 2023. [Online]. Available: https://www.ijem.com/visual-thinking-and-reading-comprehension-foreign-language-setting-as-an-example

[232] P. Yenawine, *Visual Thinking Strategies: Using Art to Deepen Learning across School Disciplines*. Harvard Education Press, 2013. [Online]. Available: https://eric.ed.gov/?id=ED568850

[233] K. Huh, "Visual thinking strategies and creativity in english education," *Indian Journal of Science and Technology*, vol. 9, no. Special Issue 1, pp. 1–6, 2016. [Online]. Available: https://indjst.org/articles/visual-thinking-strategies-and-creativity-in-english-education