ON THE SHAPE OF BRAINSCORES FOR LARGE LANGUAGE MODELS (LLMS)

Jingkai Li
OpenSci.World
Montréal, Québec H4R 2R9, Canada
jingkai.li@opensci.world OR
jkli898@126.com

ABSTRACT

With the rise of Large Language Models (LLMs), the novel metric "Brainscore" emerged as a means to evaluate the functional similarity between LLMs and human brain/neural systems. Our efforts were dedicated to mining the meaning of the novel score by constructing topological features derived from both human fMRI data involving 190 subjects, and 39 LLMs plus their untrained counterparts. Subsequently, we trained 36 Linear Regression Models and conducted thorough statistical analyses to discern reliable and valid features from our constructed ones. Our findings reveal distinctive feature combinations conducive to interpreting existing brainscores across various brain regions of interest (ROIs) and hemispheres, thereby significantly contributing to advancing interpretable machine learning (iML) studies. The study is enriched by our further discussions and analyses concerning existing brainscores. To our knowledge, this study represents the first attempt to comprehend the novel metric brainscore within this interdisciplinary domain.

1 INTRODUCTION

There has been a notable surge in the proliferation of Large Language Models (LLMs) (Beeching et al., 2023) in recent times. Within this burgeoning landscape, certain models could be reasonably regarded as early, albeit incomplete, versions of artificial general intelligence (AGI) systems (Bubeck et al., 2023). Preceding the realization of AGI, a crucial imperative arises: we want to, and *need* to know the extent of human-likeness inherent in the LLMs under development.

Inspired by biological neural networks, Artificial Neural Networks (ANNs) were created in the 1940s by Warren McCullough and Walter Pitts (Russell & Norvig, 2010), marking the inception of one of the three principal schools of the machine learning community: *Connectionism* (Bognar, 2022). Subsequently, deep and very deep neural networks emerged, and LLMs nowadays. Nevertheless, *the extent to which ANNs have diverged from biological neural networks* remains largely unknown and vastly underexplored, thereby motivating our study.

In addressing the aforementioned inquiries, scientists have commenced employing state-of-the-art LMs to investigate neural activity in the human brain during language processing. Simultaneously, NLP researchers have been motivated to incorporate neuroimaging data to improve their model, as delineated in Karamolegkou et al. (2023). Within this context, the novel metrics "brainscores" (Schrimpf et al., 2018) and "brain hierarchy score" (Nonaka et al., 2021), have been created to assess the functional similarity between AI models and human brain/neural systems.

The Pearson Correlation/Brainscore metric, as surveyed in Karamolegkou et al. (2023), stands as the predominant method in mapping brains with LMs. In addition to those efforts to calculate "brainscores," our study seeks to address the fundamental questions: *What is the meaning of the score? Can we derive features to interpret it?* Building on prior efforts to quantify similarities between LMs and human neural responses, Li et al. (2023a)'s groundbreaking study employed Procrustes analysis, a form of statistical shape analysis, to align brain fMRI representations with those of LMs. This approach is substantiated by their discussions on the philosophical foundations regarding the non-

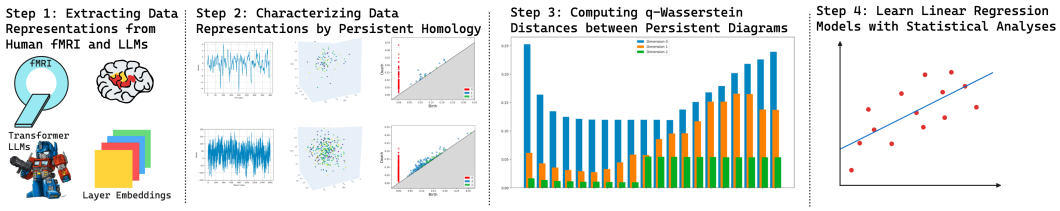

Figure 1: This figure delineates the four steps of our proposed approach: (1) Extracting data representations from human fMRI and diverse LLMs, (2) Characterising those data representations by Persistent Homology, (3) Constructing features by computing q-Wasserstein Distances between pairs of Persistence Diagrams, and (4) Systematically filter in reliable and valid features by learning Linear Regression Models with statistical analyses.

trivial nature of structural similarities. Our study extends the concept of "Shape" by systematically comparing the distinctions in the "shapes" between collected human brain fMRI data and the internal embeddings of various LLMs [1]. We employ the term "shape" deliberately, as our observation and quantification of the data are from a topological perspective, focusing solely on their "shapes" rather than their geometric properties such as lengths, areas, volumes, angles, etc. Leveraging the Topological Data Analysis (TDA) tool Persistent Homology (PH), we characterize data representations in both realms and construct features by computing q-Wasserstein Distances between pairs of their persistence diagrams. Subsequently, we learn Linear Regression Models to fit the existing "brainscores," followed by rigorous statistical analyses to identify reliable and valid features among our constructed ones. Our primary contributions are outlined below:

- We present detailed results concerning reliable and valid features derived from our constructed ones. It exhibits distinct feature combinations that facilitate the interpretation of existing brainscores for each ROI and each hemisphere. The identified features hold significant value for advancing studies in interpretable machine learning (iML). Our original research findings are augmented with further discussions and analyses on existing brainscores.
- We demonstrate the efficacy of the TDA tool Persistent Homology (PH), and the q-Wasserstein Distance metric in constructing features that are rarely seen in current methods, analytical frameworks, and metrics within the domain of brainscores research. This showcases the potential for learning from data apart from existing conventions.
- To the best of our knowledge, this study represents the inaugural endeavor in interpreting the novel metric brainscore within this interdisciplinary domain. Overall, we take small steps in trying to address the fundamental inquiries: *"In what sense are the LLMs we create human-like?"*

2 Methods and Data Processing Pipelines

2.1 Overview of Methods and Data Processing Pipelines

The brainscores that we try to interpret are based on Caucheteux et al. (2023). Figure 1 illustrates the four steps of our proposed approach.

We firstly extracted data representations from existing human fMRI data and embeddings from specific layers of various Transformer (Vaswani et al., 2017) based LLMs (Appendix A.1). Subsequently, we characterized these data representations using PH (Appendix A.2). Simultaneously, we reproduced the results from Caucheteux et al. (2023) and computed corresponding brainscores for our selected LLMs against our specified ROIs as the learning target. Following this, we constructed a set of features by computing q-Wasserstein Distances between pairs of persistence diagrams (Appendix A.3). Finally, we systematically identified reliable and valid features from our constructed ones by learning arrays of Linear Regression Models, accompanied by subsequent statistical analyses (Appendix A.4).

[1] A total of 39 models plus their untrained configurations, spanning from albert-base to Llama2-70B (quantized).

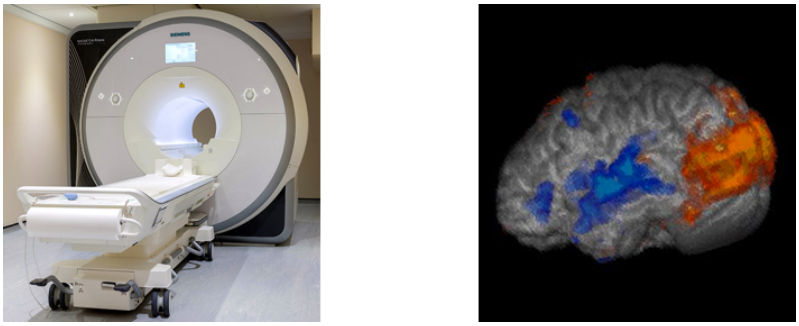

Figure 2: The original images are from Introduction to fMRI. The image on the left is a typical research scanner that has a field strength of 3 teslas (T), about 50,000 times greater than the Earth's field. The right one is the result of an fMRI experiment, and the processed image is the brain's "activation map" during the experiment.

2.2 Intuitions behind the Methods

This section proffers an intuitive explication of the foundational techniques underpinning our methodological approaches and data processing pipelines, namely, the neuroimaging technique of functional Magnetic Resonance Imaging (fMRI) (Section 2.2.1), the TDA tool of Persistent Homology (PH) (Section 2.2.2), and the q-Wasserstein Distance metric (Section 2.2.3). These expository overviews aim to facilitate a comprehensive understanding of the integral components constituting our investigative framework.

2.2.1 An Intuitive Introduction to fMRI

Functional Magnetic Resonance Imaging (fMRI) is a powerful technique used in neuroscience and medicine to study the brain's activity and functionality. It is a non-invasive imaging method that measures changes in blood oxygenation levels within the brain, which are related to neural activity (Martini et al., 2019).

Brain activity: When a particular region of the brain is active, it requires more oxygen and nutrients. This increased demand for oxygen is met by an increase in blood flow to that region. **Magnetic properties**: Oxygenated blood (blood carrying oxygen) and deoxygenated blood (blood without oxygen) have different magnetic properties. This difference is known as the "Blood Oxygenation Level Dependent" (BOLD) signal. **fMRI scanning**: During an fMRI scan, powerful magnetic fields and radio waves are used to detect and measure the BOLD signal in different parts of the brain. This allows researchers and doctors to infer which brain regions are more active or less active during a particular task or condition. **Spatial and temporal resolution**: fMRI provides good spatial resolution, meaning it can pinpoint the location of brain activity with reasonable accuracy. It also has a relatively good temporal resolution, allowing researchers to track changes in brain activity over time, although not as precise as some other techniques like electroencephalography (EEG). **Applications**: fMRI is widely used in cognitive neuroscience research to study brain function during various mental processes, such as perception, decision-making, memory, and language processing. It is also used in clinical settings to diagnose and study brain disorders, such as strokes, tumors, and neurodegenerative diseases (Ulmer & Jansen, 2010; 2013).

An MRI scanner and a corresponding brain "activation map", generated from an fMRI experiment (as illustrated in Introduction to fMRI), are depicted in Figure 2. The processed fMRI data for our current investigation are presented in Appendix C.

By measuring changes in blood oxygenation levels, fMRI provides a window into the brain's functional activity, enabling researchers and clinicians to better understand how the brain works and to investigate potential abnormalities or dysfunctions in different brain regions (Ulmer & Jansen, 2010; 2013).

In accordance with prior studies such as Caucheteux et al. (2023) and Oota et al. (2023), the fMRI data utilized in our investigation originate from the Narratives dataset (Nastase et al., 2021), which

is publicly accessible on OpenNeuro. The dataset is a collection which aggregates a variety of fMRI datasets collected while human subjects listened to real spoken stories.

Utilizing the same Narratives dataset (Nastase et al., 2021), Ye et al. (2024) investigated 8 participants, Hahamy et al. (2023) studied one task (21st year) with 25 participants, and Liu et al. (2023) examined two tasks: "Pie Man" with 75 participants and "Shapes" with 59 subjects, whereas, Oota et al. (2023) analyzed data from 82 subjects listening to the single story titled "Pie Man".

Considering the scale of our study, we extracted fMRI data from 190 out of 345 participants (subjects) listening to 3 out of 27 stories (tasks) from the dataset's pool. The selected tasks, namely "Pie man," "Shapes," and "It's Not the Fall That Gets You," were chosen due to their inclusion of the top 3 most subjects (refer to Appendix B and Table 1 for a concise overview).

Schrimpf et al. (2021) conducted a large-scale investigation into the relationships between brain activity and 43 contemporary neural network models (encompassing embedding, recurrent, and transformer architectures). Their analysis leveraged three datasets: two fMRI and one electrocardiography (ECG) signals. To assess model-brain alignment, they employed a novel metric termed the Brainscore, which incorporates noise ceiling normalization of the Pearson correlation coefficient.

As a large-scale study, we examined and analyzed 39 LLMs, spanning from albert-base to Llama-70B (quantized), plus their untrained counterparts [2]. To be in line with Schrimpf et al. (2021), we didn't make a distinction between masked language models and causal language models, but limit the architecture of our interest as Transformer (Vaswani et al., 2017) based one.

2.2.2 An Intuitive Introduction to Persistent Homology (PH)

Persistent homology (PH) is a powerful tool in topological data analysis (TDA) that allows us to study the shape and structure of data at different scales or resolutions. It provides a way to quantify and understand the topological features of data, such as connected components, holes, and higher-dimensional voids, and how these features persist or change as the scale or resolution changes (Fugacci et al., 2016; Carter, 2020; Koplik, 2022).

Imagine the data as a point cloud in a high-dimensional space. For example, if the data consists of images, each image can be represented as a point in a high-dimensional space, where each dimension corresponds to a pixel value. Then, imagine placing small balls or spheres around each data point. The size of these balls can be thought of as a scale parameter or resolution at which we are looking at the data. As the size of the balls increases, they start to overlap and merge, forming connected components, loops, and higher-dimensional voids or cavities. Persistent homology tracks the birth and death of these topological features (connected components, loops, voids) as the scale or resolution changes (i.e., as the ball size increases or decreases). The features that persist over a wide range of scales are considered more significant or robust, while features that appear and disappear quickly are considered noise or less important. Persistent homology represents this information in a concise way, typically using persistence diagrams or barcodes, which show the birth and death times (or scales) of the topological features.

In the example shown in Figure 3 (from the original notebook: TDA with Python using the Gudhi Library - Persistent homology and persistence diagrams) we witness the filtration given by a union of growing balls centered on the finite set of points C. We can see how the topological features emerge, change, and disappear at different scales or resolutions.

The power of persistent homology lies in its ability to capture and quantify the topological structure of data at different scales, without relying on specific assumptions about the data distribution or shape. It can reveal hidden patterns, clusters, and higher-dimensional structures that may not be apparent from traditional geometric or statistical methods (Carlsson, 2009; Edelsbrunner & Harer, 2010; Carlsson & Vejdemo-Johansson, 2021; Virk, 2022).

Koplik (2022) provides a non-technical introduction to PH, supplemented by illustrative examples. Another accessible one catering to newcomers is available in Fugacci et al. (2016). Readers seeking more formal and in-depth treatments may refer to Carlsson (2009); Edelsbrunner & Harer

[2]Untrained versions were unavailable for LLMs: Llama-70B (quantized), Llama-13B (quantized), and Llama-7B (quantized), thus yielding a total of 36 untrained LLMs.

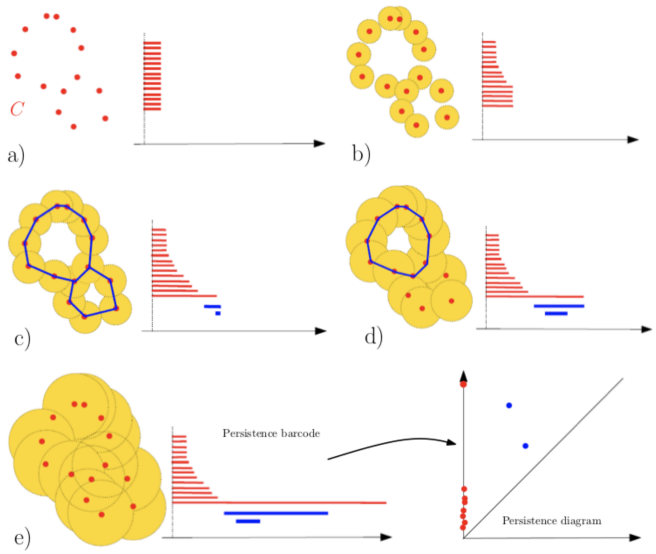

Figure 3: The original figures are from TDA with Python using the Gudhi Library - Persistent homology and persistence diagrams. a) For the radius $r = 0$, the union of balls is reduced to the initial finite set of point, each of them corresponding to a 0-dimensional feature, i.e. a connected component; an interval is created for the birth for each of these features at $r = 0$. b) Some of the balls started to overlap resulting in the death of some connected components that get merged together; the persistence diagram keeps track of these deaths, putting an end point to the corresponding intervals as they disappear. c) New components have merged giving rise to a single connected component and, so, all the intervals associated to a 0-dimensional feature have been ended, except the one corresponding to the remaining components; two new 1-dimensional features, have appeared resulting in two new intervals (in blue) starting at their birth scale. d) One of the two 1-dimensional cycles has been filled, resulting in its death in the filtration and the end of the corresponding blue interval. e) all the 1-dimensional features have died, it only remains the long (and never dying) red interval. The final barcode can also be equivalently represented as a persistence diagram where every interval (a, b) is represented by the the point of coordinate (a, b) in $\mathbb{R}$. Intuitively the longer is an interval in the barcode or, equivalently the farther from the diagonal is the corresponding point in the diagram, the more persistent, and thus relevant, is the corresponding homological feature across the filtration.

(2010); Boissonnat et al. (2018); Carter (2020); Carlsson & Vejdemo-Johansson (2021); Dey & Wang (2022); Virk (2022).

In our investigation, we utilized the potent tool PH to characterize the extracted data representations obtained from fMRI and LLMs embeddings, as elaborated in Appendix A.2.

2.2.3 An Intuitive Introduction to Persistence Diagram and q-Wasserstein Distance

Imagine the PH is applied to the data, identifying all the important holes and voids at different scales. Now, a persistence diagram is a way to represent this information visually.

Think of it as a scatter plot. On the x-axis, there is the birth time of each hole, and on the y-axis, presented is the death time. Each point on the plot represents a hole in the data, with its coordinates indicating when the hole was born and when it "died" as zoomed in or out (Carlsson, 2009; Edelsbrunner & Harer, 2010; Carter, 2020; Carlsson & Vejdemo-Johansson, 2021; Virk, 2022).

The beauty of persistence diagrams is that they capture the essential topological features of the data in a concise and informative way. By looking at the distribution of points and their positions relative to each other, one can gain insights into the structure and complexity of the data. The intuitive examples are shown in Figure 3, whereas examples of persistence diagrams for our current study can be found in Figure 34 and Figure 38.

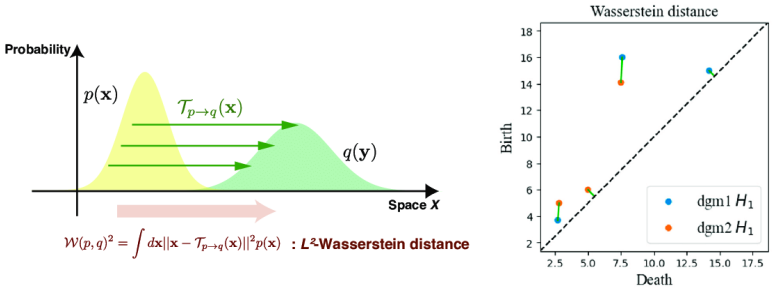

Figure 4: The figure on the left is originally from Nakazato & Ito (2021), which is a schematic of the L^2-Wasserstein distance. They here consider optimal transport from the probability distribution $p(\mathbf{x})$ to the probability distribution $p(\mathbf{y})$. The length of the green arrow shows the optimal transportation distance $||\mathbf{x} - \mathcal{T}_{p\to q}(\mathbf{x})||$, and the square of the L^2-Wasserstein distance is given by the expected value of the square of its optimal transportation distance. The plot on the right, originally from Zhang et al. (2021), shows that the Wasserstein distance is the sum of p-th power of the distance between all points matching two persistence diagrams. If no corresponding matching point is found, it will match to diagonal.

The q-Wasserstein distance is a way to measure the dissimilarity between two distributions. It's based on a concept from optimal transport theory (Mukherjee et al., 2021; Nietert et al., 2022), which is all about finding the most efficient way to "transport" one distribution of points to another.

In the context of persistence diagrams, the q-Wasserstein distance tells how much "work" is required to transform one set of holes into another. Here, "work" refers to the amount of "mass" (i.e., the persistence of each hole) that needs to be moved from one point to another.

Figure 4 illustrates the q-Wasserstein distances for two distributions (left) (Nakazato & Ito, 2021) and two persistence diagrams (right) (Zhang et al., 2021).

By calculating the q-Wasserstein distance between two persistence diagrams, one can quantitatively measure how similar or dissimilar their topological structures are. This is useful for tasks like comparing datasets, identifying outliers, or tracking changes in data over time (Carlsson, 2009; Edelsbrunner & Harer, 2010; Carter, 2020; Carlsson & Vejdemo-Johansson, 2021; Virk, 2022).

A concise overview of the q-Wasserstein Distance is provided in Rahman (2023), while more comprehensive and formal introductions can be found in Edelsbrunner & Harer (2010); Boissonnat et al. (2018); Carlsson & Vejdemo-Johansson (2021); Dey & Wang (2022).

In our study, series of q-Wasserstein Distances are calculated with the aim of deriving features that subsequently aid in the interpretation of the corresponding brainscores.

2.3 Implementation Details and Experiments

Expanding upon the introductory context, we further present the implementation details and experimental procedures, as outlined in Appendix A.

3 Results

We present our results in terms of the filtered-in reliable and valid features w.r.t q-Wasserstein Distances with specific q values for each ROI, and each hemisphere. Given the division of our training data into "Only Trained" and "Trained plus Untrained" LLMs, the aforementioned results are presented separately for each group.

Figure 50 in Appendix I illustrates the maximum q for each ROI and each hemisphere that achieves the highest averaged test R^2 score as our result for the First Pass (detailed in Appendix A.4.1).

The detailed results of the Second Pass are elucidated in Appendix J. In this phase, we refined features by scrutinizing whether each averaged p value met the criterion $p < 5\%$, as elaborated in Appendix A.4.2. Each table within Appendix J provides a summary of the ultimately filtered-

in features [3] [4] for each ROI (including the whole brain mask), and each hemisphere, along with their respective weights, including both lower and upper bounds of 95% confidence intervals (CI), standard error (SE) averaged across all 25 runs, feature importance t-statistic value $|t|$, and the p value [5].

Each table is then further accompanied with two figures in its following. The first figure depicts each feature's weight along with both the lower and upper bounds of the 95% CIs. While the second one illustrates the coefficient importance and its variability.

4 Discussions and Limitations

The brainscores, derived through the re-implementation of Caucheteux et al. (2023), served a dual purpose in our study. Not only did we employ them for learning Linear Regression Models, but they also played a crucial role in addressing our research questions. Simultaneously, we conducted various analyses to augment our investigation. The discussions and limitations outlined in Section 4.1, Section 4.2, Section 4.3, Section 4.4, and Section 4.5 are directly rooted in the computed brainscores. Subsequently, a follow-up discussion ensues, offering interpretations for these brainscores based on our original research findings, as detailed in Section 4.6.

4.1 Notes on Existing Brainscores

The novel meric brainscore (Schrimpf et al., 2018) has emerged as the primary one for mapping brains with LLMs (Karamolegkou et al., 2023). However, being a metric under ongoing development, divergent scores and even conflicting results have been observed across different studies (Schrimpf et al., 2018; Nonaka et al., 2021; Macpherson et al., 2021), thereby introducing significant uncertainty into interpretation studies, including our current one.

The absolute values of the calculated brainscores remain notably low, suggesting that our developed LLMs still exhibit considerable disparity from human brain functionality. Meanwhile, advancements in Brain–Computer Interface (BCI) technology [6] promise to yield higher-quality data, fostering further interdisciplinary investigations.

4.2 Comparisons of the Brainscores between LLMs: # Parameters Matter

We are interested in the brainscores for diverse LLMs incorporating innovations in architectures, representations, optimizations, generalizations, and the size i.e. # of parameters of the model, in light of the evolving landscape i.e. recent super giant LLMs (Beeching et al., 2023).

We pose a direct inquiry: *"Does the brainscore exhibit an increase with a larger number of parameters in the LLM?* The response aligns with our intuitive expectation, leaning towards an affirmative answer, but it is crucial to note that this correlation is not universal. While most instances in Figure 87 and Figure 88 within Appendix K support the affirmative correlation, counterexamples also exist. Specifically, the trained LLMs, namely Llama-2-7B-GPTQ, Lama-2-13B-GPTQ, and Lama-2-70B-GPTQ, as illustrated in Figure 87, challenge the notion that an increase in parameters consistently leads to higher brainscores.

These brainscores enable further statistical analyses, such as investigating the correlation between the number of parameters in each LLM and its associated brainscore. The outcomes of these analyses may vary based on the specific groups sampled, i.e. different ROIs, hemispheres, LLM layers, and task conditions, among others.

[3] The Intercept is consistently presented, although it is not utilized for interpretations, irrespective of its p value.

[4] Only the features meeting the criterion $p < 5\%$ are presented.

[5] All values are averaged across the 25 runs, except for SE.

[6] BCI implementations encompass a spectrum from non-invasive (EEG, MEG, MRI) and partially invasive (ECoG and endovascular) to invasive (microelectrode array), based on how close electrodes get to brain tissue (Martini et al., 2019).

4.3 DOES THE TRAINING HELP INCREASE THE LLM'S BRAINSCORE?

Researchers and practitioners have devoted extensive efforts to the training of the developed LLMs. The training process is instrumental in endowing the model with intelligence. The pivotal question arises: *Does the training contribute to an elevation in the LLM's brainscore?* Conversely, is there an inverse correlation, *does the training result in a decrease in the brainscore?*

Once again, the response to the aforementioned inquiries is predominantly *Yes*, though, *not always*. Notably, the instances in Figure 87 and Figure 88 within Appendix K exemplify both scenarios. In most examples, the brainscores of trained LLMs surpass those of their untrained counterparts. However, negative cases, such as bert-large-uncased and bert-base-cased in Figure 87, and bert-base-uncased and bert-large-uncased in Figure 88, deviate from this trend.

We, again, assert further analyses can unveil more patterns based on our computed brainscores, contributing to a deeper understanding of our created *intelligence* i.e. LLMs.

4.4 HOW DO THE COMPRESSION TECHNIQUES I.E. QUANTIZATION, PRUNING, DISTILLATION, LOW-RANK FACTORIZATION ETC. AFFECT THE BRAINSCORES?

Efforts have been undertaken to explore various compression techniques, i.e. Quantization, Pruning, Distillation, and Low-Rank Factorization (Zhu et al., 2023), as a response to the escalating training costs associated with the growing number of parameters in LLMs. Notably, some compression methods, such as quantization, impact the performance of LLMs (Frantar et al., 2023; Kawrakow, 2023). In light of this, we are interested in examining *How is the brainscore influenced when we compress the LLM compared to its full-resolution configuration?*

Regrettably, our sample size is insufficient to provide a conclusive answer, as our calculated brainscores only include Llama2-7b and its quantized counterpart relevant to this question. However, the situation is intriguingly nuanced and *depends* on specific instances. The comparison between the original and quantized LLMs, exemplified by Llama-2-7b-hf vs. Llama-2-7B-GPTQ in Figure 87 and Figure 88 in Appendix K, aligns with the aforementioned performance decline. While, an intriguing observation surfaces in Figure 89, where the quantized version in the pair exhibits an even higher brainscore. We posit this question as another avenue for future research.

4.5 ANALOGY OF BRAIN ROIS FOR DIFFERENT LLM LAYERS

Brain Regions of Interest (ROIs) are often used as the nodes of functional brain networks. ROIs consist of several fMRI measurement voxels that are assumed to exhibit functional homogeneity (Poldrack, 2007; Ryyppö et al., 2018). Concurrently, Transformer (Vaswani et al., 2017) based LLMs consist of varying numbers of Transformer-block layers. Drawing an analogy, our inquiry revolves around *Do different layers of LLMs play different functional roles compared to those delineated as Brain ROIs?*

Researches (Jawahar et al., 2019; Manning et al., 2020) in the field have indicated that the deep layers of some LLM encode higher-level and more contextualized representations than their initial layers. Additionally, interdisciplinary investigations involving Brain/Neural Science have sought to determine which LLM layers exhibit the most brain-like behavior (Tucker & Tuckute, 2023; Li et al., 2023a). However, previous studies such as Tucker & Tuckute (2023) only examined two LLMs, namely BERT-large-cased and GPT2-XL, for selected ROIs, while Li et al. (2023a) explored 14 LLMs without explicitly specifying any particular ROI. Consequently, this area remains significantly underexplored.

We present several heatmaps depicting the averaged brainscores across uniformly sampled eight layers [7] of the LLM: Llama-2-7b-hf, computed for all eight language-related brain ROIs. These visualizations are provided in Figure 90, Figure 91, and Figure 92 within Appendix K. Notably, the ROI labeled as PTL consistently exhibits elevated scores compared to others. The PTL is known for its significant involvement in lexicalized syntactic processing (Matchin et al., 2019; Matchin & Wood, 2020). This raises the question: *Do our developed LLMs primarily serve the functions associated with this specific ROI?*

[7] The sampling procedures are detailed in Appendix A.2.2

4.6 INTERPRETATIONS FOR THE LEARNED LINEAR REGRESSION MODELS

Finally, it comes to interpret our learned Linear Regression Models. This process unfolds in Section 3, with additional granularity provided in Appendix I and Appendix J. Within each table, the t-statistic $|t|$ values serve as indicators of the significance of individual features. These distinct feature combinations facilitate the interpretation of the current brainscores associated with different ROIs and hemispheres.

It is notable that while the brainscore serves as a metric denoting the *similarity* between fMRI and LLM data representations, the q-Wasserstein Distance quantifies the *dissimilarities* between two distributions. Therefore, particular attention should be directed towards features bearing negative weights.

Our learned Linear Regression Models and the subsequently filtered-in features are not yet in their definitive optimal state, as evidenced by Figure 50, where all the highest averaged test R^2 scores remain below the ideal threshold to varying degrees. Potential avenues for improvement includes considerations on removal of outliers on persistence diagrams (Appendix A.2), i.e. the q-Wasserstein Distance is reported to be sensitive to outliers (Mukherjee et al., 2021; Nietert et al., 2022), the exploration of different q and p values for constructing features through q-Wasserstein Distances (Appendix A.3), and the utilization of diverse Linear Regression Models (Appendix A.4), i.e. Ridge vs. OLS. It is evident that these considerations will lead to algorithm optimization and increased complexity.

Moreover, there has been limited comparison of the structural properties between human brain/neural systems and LLMs [8]. We thus propose it as another promising avenue for research.

5 CONCLUSIONS

We devoted our efforts in mining the meaning of the novel metric brainscores through the construction of topological features derived from both human fMRI data, encompassing 190 subjects, and 39 LLMs along with their untrained counterparts. Subsequently, we trained a total of 36 Linear Regression Models and conducted thorough statistical analyses to discern reliable and valid features from our constructed ones. Our findings reveal distinctive feature combinations conducive to interpreting existing brainscores across various ROIs and hemispheres, thereby contributing significantly to advancing iML studies. The study is enhanced by our further discussions and analyses regarding existing brainscores. To our knowledge, this study represents the first attempt to comprehend the novel metric brainscore within this interdisciplinary domain. Overall, we take incremental steps toward addressing fundamental inquiries: *"In what sense are the LLMs we create human-like?"*.

IMPACT STATEMENTS

Our study does not present direct negative societal impacts, as we did not generate any new fMRI data nor develop any new LLMs. All the data and tools utilized in our research are publicly accessible.

It is crucial to recognize that the biological brain and neural system do not represent the exclusive or ultimate objective for AI development and research. This perspective arises from a fundamental philosophical inquiry: *"Is the biological brain and neural system the epitome of purity and perfection? Can intelligence be created in forms distinct from our own?"* Consequently, future advancements in AI must factor in this perspective. However, comprehending the disparities between human intelligence and artificially created one is paramount. This understanding is propelled by our curiosity about the unknown and our societal responsibilities.

[8] For instance, with the exception of studies such as the Brain Hierarchy (BH) score, which quantifies the degree of hierarchical correspondence between DNN and human brain activity in image recognition (Nonaka et al., 2021)

REFERENCES

Cordell M Baker, Joshua D Burks, Robert G Briggs, Andrew K Conner, Chad A Glenn, Kathleen N Taylor, Goksel Sali, Tressie M McCoy, James D Battiste, Daniel L O'Donoghue, et al. A connectomic atlas of the human cerebrum—chapter 7: the lateral parietal lobe. *Operative Neurosurgery*, 15(Suppl 1):S295, 2018.

Edward Beeching, Clémentine Fourrier, Nathan Habib, Sheon Han, Nathan Lambert, Nazneen Rajani, Omar Sanseviero, Lewis Tunstall, and Thomas Wolf. Open llm leaderboard. `https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard`, 2023.

Melinda Zsofia Bognar. Prospects of ai in architecture: Symbolicism, connectionism, actionism. *Journal of Architectural Informatics Society*, 2022.

Jean-Daniel Boissonnat, Frédéric Chazal, and Mariette Yvinec. *Geometric and topological inference*, volume 57. Cambridge University Press, 2018.

Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*, 2023.

Sofia Bzovsky, Mark R Phillips, Robyn H Guymer, Charles C Wykoff, Lehana Thabane, Mohit Bhandari, Varun Chaudhary, and RETINA study group Sivaprasad Sobha 8 Kaiser Peter 9 Sarraf David 10 Bakri Sophie J. 11 Garg Sunir J. 12 Singh Rishi P. 13 14 Holz Frank G. 15 Wong Tien Y. 16 17. The clinician's guide to interpreting a regression analysis. *Eye*, 36(9):1715–1717, 2022.

Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

Gunnar Carlsson and Mikael Vejdemo-Johansson. *Topological data analysis with applications*. Cambridge University Press, 2021.

Nathan Carter. *Data Science for Mathematicians*. CRC Press, 2020.

Charlotte Caucheteux and Jean-Rémi King. Brains and algorithms partially converge in natural language processing. *Communications biology*, 5(1):134, 2022.

Charlotte Caucheteux, Alexandre Gramfort, and Jean-Rémi King. Evidence of a predictive coding hierarchy in the human brain listening to speech. *Nature human behaviour*, 7(3):430–441, 2023.

Davide Chicco, Matthijs J Warrens, and Giuseppe Jurman. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, 7:e623, 2021.

David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have l p -stable persistence. *Foundations of Computational Mathematics*, 10(2):127–139, Jan 2010. doi: https://doi.org/10.1007/s10208-010-9060-6.

Alexis Conneau and Guillaume Lample. Cross-lingual language model pretraining. *Advances in neural information processing systems*, 32, 2019.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*, 2019.

Rutvik H Desai, Usha Tadimeti, and Nicholas Riccardi. Proper and common names in the semantic system. *Brain Structure and Function*, 228(1):239–254, 2023.

Christophe Destrieux, Bruce Fischl, Anders Dale, and Eric Halgren. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, 53(1):1–15, 2010.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Tamal Krishna Dey and Yusu Wang. *Computational topology for data analysis*. Cambridge University Press, 2022.

Herbert Edelsbrunner and John L Harer. *Computational topology: an introduction*. American Mathematical Society, 2010.

Oscar Esteban, Christopher J Markiewicz, Ross W Blair, Craig A Moodie, A Ilkay Isik, Asier Erramuzpe, James D Kent, Mathias Goncalves, Elizabeth DuPre, Madeleine Snyder, et al. fmriprep: a robust preprocessing pipeline for functional mri. *Nature methods*, 16(1):111–116, 2019.

Evelina Fedorenko, Po-Jang Hsieh, Alfonso Nieto-Castañón, Susan Whitfield-Gabrieli, and Nancy Kanwisher. New method for fmri investigations of language: defining rois functionally in individual subjects. *Journal of neurophysiology*, 104(2):1177–1194, 2010.

Evelina Fedorenko, Alfonso Nieto-Castanon, and Nancy Kanwisher. Lexical and syntactic representations in the brain: an fmri investigation with multi-voxel pattern analyses. *Neuropsychologia*, 50(4):499–513, 2012.

Evelina Fedorenko, Terri L Scott, Peter Brunner, William G Coon, Brianna Pritchett, Gerwin Schalk, and Nancy Kanwisher. Neural correlate of the construction of sentence meaning. *Proceedings of the National Academy of Sciences*, 113(41):E6256–E6262, 2016.

Elias Frantar, Saleh Ashkboos, Torsten Hoefler, and Dan Alistarh. Gptq: Accurate post-training quantization for generative pre-trained transformers. *arXiv preprint arXiv:2210.17323*, 2022.

Elias Frantar, Saleh Ashkboos, Torsten Hoefler, and Dan Alistarh. OPTQ: accurate quantization for generative pre-trained transformers. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/pdf?id=tcbBPnfwxS.

Ulderico Fugacci, Sara Scaramuccia, Federico Iuricich, Leila De Floriani, et al. Persistent homology: a step-by-step introduction for newcomers. In *STAG*, pp. 1–10, 2016.

Matthew F Glasser, Timothy S Coalson, Emma C Robinson, Carl D Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F Beckmann, Mark Jenkinson, et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178, 2016.

Suriya Gunasekar, Yi Zhang, Jyoti Aneja, Caio César Teodoro Mendes, Allie Del Giorno, Sivakanth Gopi, Mojan Javaheripi, Piero Kauffmann, Gustavo de Rosa, Olli Saarikivi, et al. Textbooks are all you need. *arXiv preprint arXiv:2306.11644*, 2023.

Avital Hahamy, Haim Dubossarsky, and Timothy EJ Behrens. The human brain reactivates context-specific past information at event boundaries of naturalistic experiences. *Nature neuroscience*, 26 (6):1080–1089, 2023.

Alyssa Hughes. Phi-2: The surprising power of small language models, 12 2023. URL https://www.microsoft.com/en-us/research/blog/phi-2-the-surprising-power-of-small-language-models/.

Ganesh Jawahar, Benoît Sagot, and Djamé Seddah. What does bert learn about the structure of language? In *ACL 2019-57th Annual Meeting of the Association for Computational Linguistics*, 2019.

Antonia Karamolegkou, Mostafa Abdou, and Anders Søgaard. Mapping brains with language models: A survey. *arXiv preprint arXiv:2306.05126*, 2023.

Iwan Kawrakow. k-quants by ikawrakow · pull request #1684 · ggerganov/llama.cpp, Jun 2023. URL https://github.com/ggerganov/llama.cpp/pull/1684.

Nitish Shirish Keskar, Bryan McCann, Lav R Varshney, Caiming Xiong, and Richard Socher. Ctrl: A conditional transformer language model for controllable generation. *arXiv preprint arXiv:1909.05858*, 2019.

Gary Koplik. Persistent homology: A non-mathy introduction with examples, Jun 2022. URL https://towardsdatascience.com/persistent-homology-with-examples-1974d4b9c3d0.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*, 2019.

Jiaang Li, Antonia Karamolegkou, Yova Kementchedjhieva, Mostafa Abdou, Sune Lehmann, and Anders Søgaard. Structural similarities between language models and neural response measurements. In *NeurIPS 2023 Workshop on Symmetry and Geometry in Neural Representations*, 2023a.

Yuanzhi Li, Sébastien Bubeck, Ronen Eldan, Allie Del Giorno, Suriya Gunasekar, and Yin Tat Lee. Textbooks are all you need ii: phi-1.5 technical report. *arXiv preprint arXiv:2309.05463*, 2023b.

Xu Liu, Mengyue Zhou, Gaosheng Shi, Yu Du, Lin Zhao, Zihao Wu, David Liu, Tianming Liu, and Xintao Hu. Coupling artificial neurons in bert and biological neurons in the human brain. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(7):8888–8896, 2023.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

Tom Macpherson, Anne Churchland, Terry Sejnowski, James DiCarlo, Yukiyasu Kamitani, Hidehiko Takahashi, and Takatoshi Hikida. Natural and artificial intelligence: A brief introduction to the interplay between ai and neuroscience research. *Neural Networks*, 144:603–613, 2021.

Christopher D Manning, Kevin Clark, John Hewitt, Urvashi Khandelwal, and Omer Levy. Emergent linguistic structure in artificial neural networks trained by self-supervision. *Proceedings of the National Academy of Sciences*, 117(48):30046–30054, 2020.

Clément Maria, Jean-Daniel Boissonnat, Marc Glisse, and Mariette Yvinec. The gudhi library: Simplicial complexes and persistent homology. In *Mathematical Software–ICMS 2014: 4th International Congress, Seoul, South Korea, August 5-9, 2014. Proceedings 4*, pp. 167–174. Springer, 2014.

Michael L Martini, Eric Karl Oermann, Nicholas L Opie, Fedor Panov, Thomas Oxley, and Kurt Yaeger. Sensor modalities for brain-computer interface technology: A comprehensive literature review. *Neurosurgery*, 86(2), Jul 2019. doi: https://doi.org/10.1093/neuros/nyz286.

William Matchin and Emily Wood. Syntax-sensitive regions of the posterior inferior frontal gyrus and the posterior temporal lobe are differentially recruited by production and perception. *Cerebral Cortex Communications*, 1(1):tgaa029, 2020.

William Matchin, Christian Brodbeck, Christopher Hammerly, and Ellen Lau. The temporal dynamics of structure and content in sentence comprehension: Evidence from fmri-constrained meg. *Human Brain Mapping*, 40(2):663–678, 2019. doi: https://doi.org/10.1002/hbm.24403. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.24403.

Camille K Milton, Vukshitha Dhanaraj, Isabella M Young, Hugh M Taylor, Peter J Nicholas, Robert G Briggs, Michael Y Bai, Rannulu D Fonseka, Jorge Hormovas, Yueh-Hsin Lin, et al. Parcellation-based anatomic model of the semantic network. *Brain and behavior*, 11(4):e02065, 2021.

Christoph Molnar. *Interpretable machine learning : a guide for making black box models explainable*. Leanpub, 2022. ISBN 9798411463330.

Debarghya Mukherjee, Aritra Guha, Justin M Solomon, Yuekai Sun, and Mikhail Yurochkin. Outlier-robust optimal transport. In *International Conference on Machine Learning*, pp. 7850–7860. PMLR, 2021.

Muka Nakazato and Sosuke Ito. Geometrical aspects of entropy production in stochastic thermodynamics based on wasserstein distance. *Physical Review Research*, 3(4):043093, 2021.

Samuel A Nastase, Yun-Fei Liu, Hanna Hillman, Asieh Zadbood, Liat Hasenfratz, Neggin Keshavarzian, Janice Chen, Christopher J Honey, Yaara Yeshurun, Mor Regev, et al. The "narratives" fmri dataset for evaluating models of naturalistic language comprehension. *Scientific data*, 8(1):250, 2021.

Robert Nau. What's a good value for r-squared?, 2019. URL https://people.duke.edu/~rnau/rsquared.htm.

Sloan Nietert, Ziv Goldfeld, and Rachel Cummings. Outlier-robust optimal transport: Duality, structure, and statistical analysis. In *International Conference on Artificial Intelligence and Statistics*, pp. 11691–11719. PMLR, 2022.

Soma Nonaka, Kei Majima, Shuntaro C Aoki, and Yukiyasu Kamitani. Brain hierarchy score: Which deep neural networks are hierarchically brain-like? *IScience*, 24(9), 2021.

Subba Reddy Oota, Mounika Marreddy, Manish Gupta, and Raju Bapi. How does the brain process syntactic structure while listening? In *Findings of the Association for Computational Linguistics: ACL 2023*, pp. 6624–6647, 2023.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Guilherme Penedo, Quentin Malartic, Daniel Hesslow, Ruxandra Cojocaru, Alessandro Cappelli, Hamza Alobeidli, Baptiste Pannier, Ebtesam Almazrouei, and Julien Launay. The refinedweb dataset for falcon llm: outperforming curated corpora with web data, and web data only. *arXiv preprint arXiv:2306.01116*, 2023.

Russell A Poldrack. Region of interest analysis for fmri. *Social cognitive and affective neuroscience*, 2(1):67–70, 2007.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners, 2019.

Moklesur Rahman. Understanding wasserstein distance: A powerful metric in machine learning, Jun 2023. URL https://rmoklesur.medium.com/understanding-wasserstein-distance-a-powerful-metric-in-machine-learning-100a1ff46b66.

Stuart J Russell and Peter Norvig. *Artificial intelligence a modern approach*. London, 2010.

Elisa Ryyppö, Enrico Glerean, Elvira Brattico, Jari Saramäki, and Onerva Korhonen. Regions of interest as nodes of dynamic functional brain networks. *Network Neuroscience*, 2(4):513–535, 2018.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, 2019.

Martin Schrimpf, Jonas Kubilius, Ha Hong, Najib J Majaj, Rishi Rajalingham, Elias B Issa, Kohitij Kar, Pouya Bashivan, Jonathan Prescott-Roy, Franziska Geiger, et al. Brain-score: Which artificial neural network for object recognition is most brain-like? *BioRxiv*, pp. 407007, 2018.

Martin Schrimpf, Idan Asher Blank, Greta Tuckute, Carina Kauf, Eghbal A Hosseini, Nancy Kanwisher, Joshua B Tenenbaum, and Evelina Fedorenko. The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, 118(45):e2105646118, 2021.

Guillaume Tauzin, Umberto Lupo, Lewis Tunstall, Julian Burella Pérez, Matteo Caorsi, Anibal M Medina-Mardones, Alberto Dassatti, and Kathryn Hess. giotto-tda: A topological data analysis toolkit for machine learning and data exploration. *The Journal of Machine Learning Research*, 22(1):1834–1839, 2021.

MosaicML NLP Team. Introducing mpt-7b: A new standard for open-source, commercially usable llms, 2023. URL www.mosaicml.com/blog/mpt-7b. Accessed: 2023-05-05.

Matthew S Thiese, Brenden Ronna, and Ulrike Ott. P value interpretations and considerations. *Journal of thoracic disease*, 8(9):E928, 2016.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.

Mycal Tucker and Greta Tuckute. Increasing brain-llm alignment via information-theoretic compression. In *UniReps: the First Workshop on Unifying Representations in Neural Models*, 2023. URL https://openreview.net/forum?id=WcfVyzzJOS.

Katharine Turner. Katharine turner (12/3/19): Why should q=p in the wasserstein distance between persistence diagrams?, 12 2019. URL https://www.youtube.com/watch?v=d_iqovKui6k.

Stephan Ulmer and Olav Jansen. *fMRI : Basics and Clinical Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. ISBN 9783540681328.

Stephan Ulmer and Olav Jansen. *fMRI : Basics and Clinical Applications (Second Edition)*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 9783642343421.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Žiga Virk. Introduction to persistent homology, 2022. URL https://zalozba.fri.uni-lj.si/virk2022.pdf.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32, 2019.

Ziyi Ye, Jingtao Zhan, Qingyao Ai, Yiqun Liu, Maarten de Rijke, Christina Lioma, and Tuukka Ruotsalo. Query augmentation by decoding semantics from brain signals. *arXiv preprint arXiv:2402.15708*, 2024.

Yunsheng Zhang, Qingzhang Shi, Jiawei Zhu, Jian Peng, and Haifeng Li. Time series clustering with topological and geometric mixed distance. *Mathematics*, 9(9):1046, 2021.

Xunyu Zhu, Jian Li, Yong Liu, Can Ma, and Weiping Wang. A survey on model compression for large language models. *arXiv preprint arXiv:2308.07633*, 2023.

A IMPLEMENTATION DETAILS AND EXPERIMENTS

A.1 EXTRACTING DATA REPRESENTATIONS FROM HUMAN FMRI AND LLMS

A.1.1 EXTRACTING DATA REPRESENTATIONS FROM HUMAN FMRI

Consistent with the approach in Caucheteux et al. (2023), we used the preprocessed fMRI signals from the Narratives dataset (Nastase et al., 2021), without spatial smoothing (referred to as "afni-nosmooth" in the repository) and sampled a repetition time (TR) of 1.5s. The preprocessing steps were performed using fMRIPrep (Esteban et al., 2019); no temporal filtering was applied. The resulting preprocessing led to the analysis of cortical voxels projected onto the surface and morphed onto an "fsaverage" template brain; hereafter, they are referred to as voxels for simplicity.

We systematically derived data representations for *each task, subject, hemisphere, and Region of Interest (ROI)* from the aforementioned fMRI dataset. The term "hemisphere" denotes either the *L*(eft) or *R*(ight) half of the brain. We referred to the Glasser Atlas (Glasser et al., 2016) (consist of 180 ROIs in each hemisphere) to mask the Brain ROI, since the Narratives dataset (Nastase et al., 2021) contains annotations tied to this atlas. Consistent with the approach in Oota et al. (2023), eight language-related brain ROIs, subcategorized according to Fedorenko et al. (2010; 2012; 2016); Baker et al. (2018); Milton et al. (2021); Schrimpf et al. (2021); Desai et al. (2023), were selected: (i) angular gyrus (AG: PFm, PGs, PGi, TPOJ2, and TPOJ3); (ii) anterior temporal lobe (ATL: STSda, STSva, STGa, TE1a, TE2a, TGv, and TGd); (iii) posterior temporal lobe (PTL: A5, STSdp, STSvp, PSL, STV, TPOJ1); (iv) inferior frontal gyrus (IFG: 44, 45, IFJa, IFSp); (v) middle frontal gyrus (MFG: 55b); (vi) inferior frontal gyrus orbital (IFGOrb: a47r, p47r, a9-46v), (vii) posterior cingulate cortex (PCC: 31pv, 31pd, PCV, 7m, 23, RSC); and (viii) dorsal medial pre-frontal cortex (dmPFC: 9m, 10d, d32). Additionally, the whole brain mask was incorporated into the data processing pipeline [9]. We derived 3,366 matrices in total, each shape of those being like # TR $\times$ 40,962 voxels [10]. The noise, as suggested in the original paper (Nastase et al., 2021), led to the exclusion of specific individual–story pairs, resulting in reduced TRs and a diminished number of subjects for each of the selected 3 tasks [11]. To simplify the analysis, a subdivision of the Destrieux atlas (Destrieux et al., 2010) was employed, where regions with more than 500 vertices were subdivided, resulting in 142 regions per hemisphere, each containing fewer than 500 vertices, which reduced the number of voxels from 40,962 to 75 for each of the 3,366 matrices generated.

Similar to Caucheteux et al. (2023), formally, we denote:

- Y as the fMRI recordings elicited by a subject listening to a task, of size $T \times V$, with T as the number of fMRI time samples (TRs) and V as the number of voxels;

A.1.2 EXTRACTING DATA REPRESENTATIONS FROM LLMS

We examined and analyzed 39 LLMs, spanning from albert-base to Llama-70B (quantized), plus their untrained counterparts [12]. All of these LLMs are constructed with Transformer (Vaswani et al., 2017) blocks as their principal components, signifying their composition of varying numbers of Transformer-block layers.

Upon inputting sequences of words into each of the aforementioned LLMs, we sampled and extracted embeddings for *each LLM across each task*. This process yielded a total of 225 tensors, encompassing 39 trained LLMs and 36 untrained ones, each multiplied by 3 tasks. The tensor is three-dimensional, representing # LLM layers, # tokens in the task, and the LLM embedding dimension. [13] All the LLMs analysed are publicly accessible via Hugging Face.

Similar to Caucheteux et al. (2023), we denote:

- w as a sequence of M words (that is, our selected 3 tasks)

[9] See figures for the fMRI collected and averaged for each ROI and task in Appendix C.

[10] See Appendix D and Table 2 for details.

[11] See Appendix B and Table 1 for details.

[12] Untrained versions were unavailable for LLMs: Llama-70B (quantized), Llama-13B (quantized), and Llama-7B (quantized), thus yielding a total of 36 untrained LLMs.

[13] Refer to Appendix E and Table 4 for details.

- X as the embeddings of a LLM model input with w, of size $M \times U$, with U as the dimensionality of the embeddings (for a layer of i.e. GPT-2, U = 768). We explicitly denote X_k as the embeddings extracted from layer k.

Diverging from Caucheteux et al. (2023), we refrained from summing the LLM embeddings of words presented within the same TR to align with the sampling frequency of the fMRI and LLMs. This deviation is attributed to the nature of the LLMs under investigation, which process text (transcript) directly rather than audio input [14]. Our objective is to examine the intrinsic nature of the internal embeddings without introducing any supplementary interference.

A.2 Characterizing the Data Representations by Persistent Homology (PH)

A.2.1 Characterizing the fMRI Data Representations by PH

Continuing from Appendix A.1.1 and considering the temporal dimension (TR), we observe an array of time-series signals. We then averaged across all valid subjects for *each task, hemisphere, and ROI*, resulting in a diminished number of matrices while preserving consistent dimensions. Subsequently, we aggregated each of these matrices along the TR dimension, transforming each matrix into a 1-D time-series vector. The length of each vector corresponds to the number of TRs for the respective task (refer to Figure 32 in Appendix F for an illustrative example).

The TDA Toolkit Giotto-TDA (Tauzin et al., 2021) was employed to project the aforementioned time-series signals into a 3-D space as a point cloud. Following the tutorial provided in Topology of time series, we utilized the algorithm embedded in the Giotto-TDA toolkit to determine as optimal choices for the embedding dimension (d) and time delay (τ). The search started from relatively large values of (d, τ) to prevent convergence to sub-optimal minimum. However, we restricted the optimal embedding dimension d not to exceed 3. This restriction stems from the fact that surpassing $d = 3$ would necessitate truncating d to 3 through dimensionality reduction, thereby incurring information loss (refer to our employed Algorithm 1 in practice).

Algorithm 1 Transform 1-D Time-Series into 3-D Point Cloud

```
Input: max embedding dimension d, optimal time delay τ
Initialize d = 3, τ = # TR.
for i = 1 to # TR do
    pointCloud = searchAlgo(τ//i), d)
    if pointCloud.shape[1] < 3 then
        continue
    end if
    τ = τ//i
    break
end for
```

The shape of our projected point cloud is interpreted as optimal time delay $\tau\times$ restricted embedding dimension $d = 3$. All values of the point cloud were normalized to fall within the range of $[0, 1]$. This normalization ensures that the threshold "birth-death" time points, delineated on each persistence diagram [15], maintain controlled scales, since Our emphasis lies in the topological "shape" of these point clouds, and we deliberately mitigate the influence of their geometrical sizes (refer to Figure 33 in Appendix F for an illustrative example of a point cloud in 3-D space).

Finally, we employed the "VietorisRipsPersistence" transformer from Giotto-TDA to execute the computation for persistence. Considering computation complexity and cost, we restricted the homology dimensions to 0,1,2 (refer to Figure 34 and Figure 35 in Appendix F for the summary representations i.e. persistence diagram and persistence barcode respectively).

[14] They in fact lack the capability to process audio signals and are exclusively designed for text processing.

[15] We utilized the GUDHI library (Maria et al., 2014) to generate summary representations, i.e. persistence diagrams and persistence barcodes, from our computed PH. This library was also employed in computing q-Wasserstein Distances as outlined in Appendix A.3.

A.2.2 CHARACTERIZING THE LLMS EMBEDDINGS BY PH

Following the preceding step (Appendix A.1.2) and focusing on the token dimension for each LLM layer, we obtained again a set of time-series signals. For each specific LLM layer, we firstly aggregated the matrices (# tokens of the task $\times$ LLM embedding dimension) along the token dimension. This transformation converted each matrix into a 1-D time-series vector, with its length corresponding to # tokens for the respective task (refer to Figure 36 in Appendix G). While we analyzed all 39 LLMs [16], we did not process every individual layer of each LLM. Instead, we uniformly sampled eight layers, evenly spaced from the first one to the last (inclusive at both ends), from each LLM. These selected layers are referred to as "Used Layers" in Appendix E and Table 4, where the layer numbers marked with an underline indicate the intermediate-to-deep layers of each LLM ($l = \frac{2}{3}n_{\text{layers}}$) according to Caucheteux et al. (2023), warranting our special attention due to their superior predictive capabilities for brain activity, as reported in Schrimpf et al. (2021); Caucheteux & King (2022); Caucheteux et al. (2023).

In a manner akin to Appendix A.2.1, we projected the aforementioned time-series signals into 3-D space as point clouds. Employing the same Algorithm 1, we pursued the determination of the optimal time delay τ, constraining the optimal embedding dimension $d = 3$, where the value initialized for $\tau =$ # tokens in this instance.

Subsequently, we obtained the projected point cloud, where the size is interpreted as optimal time delay $\tau\times$ restricted embedding dimension $d = 3$. In alignment with the rationale elucidated in Appendix A.2.1, we normalized all its values within the range of $[0, 1]$ (refer to Figure 37 in Appendix G for a representative point cloud projected into the 3-D space). Homologously, the computation for homology dimensions was constrained to 0,1,2 for considerations of computational complexity and cost (see Figure 38 for corresponding summary representations as the persistence diagram, and Figure 39 in Appendix G for the persistence barcode, respectively).

A.3 COMPUTING q-WASSERSTEIN DISTANCES BETWEEN PERSISTENCE DIAGRAMS

A total of 34,416 combination pairs were generated encompassing both fMRI data and LLMs Embeddings. Each pair was associated with a specific task listened to by the human subject and processed by the LLM. More specifically, the identification of each pair involved a combination of a *particular ROI, hemisphere, LLM layer, and training status*, consuming the same *task*. The ensuing step required the computation of q-Wasserstein Distances [17] for each unique combination. It is noteworthy that the computation of averaged layer embeddings was restricted to the whole brain mask exclusively, as the averaged layer represents the entire LLM embedding, which is inherently incomparable to a specific ROI, while a distinct layer embedding better serves this purpose.

A.4 LEARNING LINEAR REGRESSION MODELS FROM q-WASSERSTEIN DISTANCES AND EXISTING BRAINSCORES

The aim of our study is to comprehend the meaning of brainscores through constructing interpretative features. To achieve this, we utilized the Linear Regression Model, chosen for its simplicity in estimation and interpretability of weights (Molnar, 2022).

The training data was partitioned to learn Linear Regression Models for *each ROI across each hemisphere* in two: "Only Trained" LLMs and "Trained plus Untrained" LLMs. The latter group involved the brainscores for each untrained LLM. Consequently, a total of 36 models were learned, covering (8+1) ROIs $\times$ 2 hemispheres $\times$ 2 training groups. We employed the scikit-learn toolkit (Pedregosa et al., 2011) to construct the models with cross validations.

Each of the 34,416 combination pairs is characterized by 903 features, representing the q-Wasserstein Distances, where $q = p$, ranging from 1 to 300, along with the special case $q = p = \infty$, across 3 PH dimensions. However, not all features are deemed reliable and valid; hence, a systematic filtration process is essential. Therefore firstly, we conducted an Exploratory Data Analysis (EDA) on the constructed features. The observed trend indicated a consistent "long tail" distribution

[16] Including their 36 untrained counterparts.

[17] As advocated by Turner (2019), we constrained the two parameters $q = p$ when computing the q-Wasserstein Distances.

followed by the q-Wasserstein Distances spanning from $q = p = 1$ to $q = p = 300$ and $q = p = \infty$ for each pair in the 34,416 combinations, with minor fluctuations in the tail (refer to Figure 40 to Figure 48 in Appendix H). Formally, the q-Wasserstein Distance is bounded by the Wasserstein Stability Theorem (Edelsbrunner & Harer, 2010; Cohen-Steiner et al., 2010).

We devised a two-pass process to systematically identify reliable and valid features capable of elucidating a substantial portion of the variability present in the brainscores while demonstrating statistical significance as well.

A.4.1 First Pass

In the First Pass, we incrementally introduced in features from $q = p = 1$ to $q = p = 300$ [18] given the observed distribution of our constructed ones. At each iteration, we implemented a 5-time repeated 5-fold cross-validation (CV) pipeline to train the Linear Regression Model. We assessed the model's performance using the averaged test R^2 score across the 25 outcomes to gauge the extent to which our learned model could elucidate variation (Chicco et al., 2021). The regular R^2 score falls within the [0,1] range (higher is better in our context), while negative values typically indicates overfitting (Nau, 2019) (refer to Figure 49, Table 5, and Table 6 in Appendix I for an illustrative example). Therefore, we employed a stopping criterion triggered by encountering a negative averaged test R^2 score. Prior to that, we recorded all q values along with their corresponding averaged test R^2 scores, and ultimately the q value yielding the highest averaged test R^2 score was determined.

Formally, we established a mapping $\phi : q \rightarrow s$ where $q \in [1, 300] \cup \{+\infty\}$ denotes the search space, and $s \in [0, 1]$ represents the corresponding averaged test R^2 score for each iteration. The q value is determined as $q_{\max} = \arg\max \phi$. (See the pseudocode for the First Pass in Algorithm 2).

Algorithm 2 $q_{\max}$ on Highest Averaged Test R^2 Score

```
Initialize φ-list = []
for q = 1 to 300 do
    X = featuresContainingMaxqWassersteinDistances
    PlusBottleneckDistance
    Y = brainscores
    regressor = linearRegressor()
    cvModel = crossValidate(regressor, X, Y, scoring = ("r2"))
    if mean(cvModel["test-score"]) ≤ 0 then
        break
    end if
    φ-list.append((q, mean(cvModel["test-score"])))
end for
q_max = arg max (φ-list)
```

A.4.2 Second Pass

Determining the optimal q alone does not suffice to identify truly reliable and valid features from our constructed multicollinear ones. Subsequently, we designated the p-value (Thiese et al., 2016; Bzovsky et al., 2022) with a threshold of $p < 5\%$ to selectively retain desired features given the determined q value. Specifically, we computed the p-value for each iteration, averaged the 25 p-values corresponding to each feature, and ultimately retained those with $p < 5\%$.

B Summaries for the Three Selected Tasks from Narratives Dataset

This section provides a concise overview of our three tasks selected from the Narratives dataset (Nastase et al., 2021).

[18] The special case where $q = p = \infty$ is consistently included in each iteration, as it lies outside the $[1, 300]$ range.

It is important to acknowledge certain discrepancies observed in the tables between Table 1 and those presented in Nastase et al. (2021). Here, we present the precise values for "TRs," "Valid TRs", "Subjects", and "Valid Subjects" based on the repository from OpenNeuro (`https://openneuro.org/datasets/ds002345/versions/1.1.4`).

Table 1: Summary for our three selected tasks from Narratives dataset (Nastase et al., 2021).

STORY	DURATION	TRS	WORDS	SUBJECTS	ONSET	VALID TRS	VALID SUBJECTS
"PIE MAN"	07:02	300	957	82	0	300	75
"SHAPES"	06:45	313	910	59	3	310	58
"IT'S NOT THE FALL THAT GETS YOU"	09:07	400	1,601	56	3	397	54

C AVERAGED FMRI FOR EACH TASK-ROI PAIR

This section presents the averaged fMRI data for each Task-ROI Pair, derived from varying numbers of subjects engaged in the three tasks: "Pie Man," "Shapes," and "It's Not the Fall That Gets You," as documented in Nastase et al. (2021). Summaries are available in Table 1. Here, we display all eight language-related ROIs along with the whole brain mask (refer to Appendix A.1.1 for specifics). Initially, for each subject's fMRI measurements across all valid TRs, we computed the average across TRs, resulting in 40,962 voxel-level blood-oxygen-level-dependent (BOLD) signal values. Subsequently, we calculated the mean across subjects participating in each task to obtain the final set of 40,962 values. Each figure comprises 12 sub-figures representing six viewing perspectives—lateral, medial, dorsal, ventral, anterior, posterior—for each hemisphere: (L)eft and (R)ight. The colorbar on the right side of each figure illustrates the BOLD value ranges, while the task-ROI information is denoted in the title of each one.

Please note that the averaging across TRs for each figure is solely for visualization purposes; we did not employ this step when calculating the "brainscores" based on Caucheteux et al. (2023) or processing fMRI data representations in the subsequent data processing pipelines.

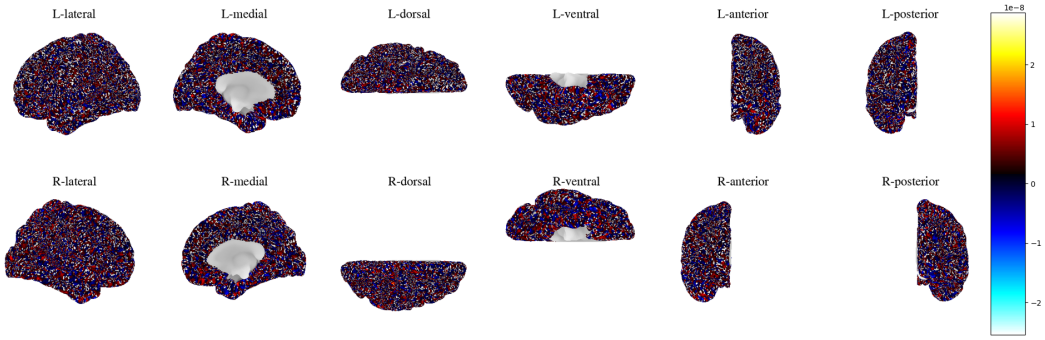

Figure 5: The averaged fMRI for the whole brain mask for the task: "Pie Man"

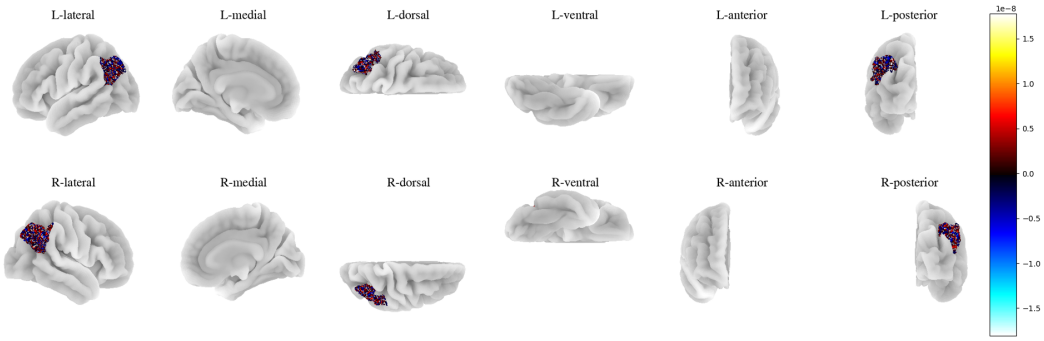

Figure 6: The averaged fMRI for the ROI: AG for the task: ”Pie Man”

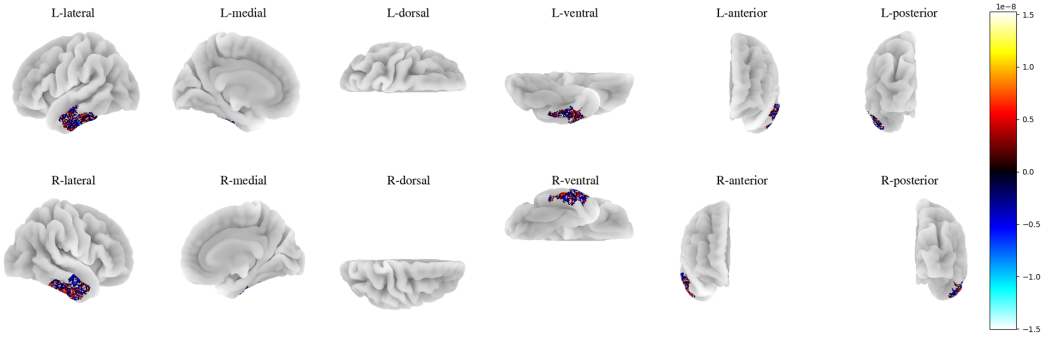

Figure 7: The averaged fMRI for the ROI: ATL for the task: ”Pie Man”

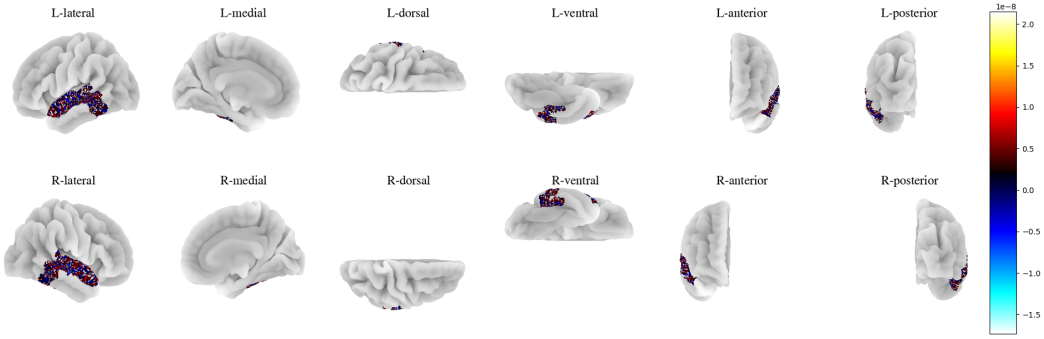

Figure 8: The averaged fMRI for the ROI: PTL for the task: ”Pie Man”

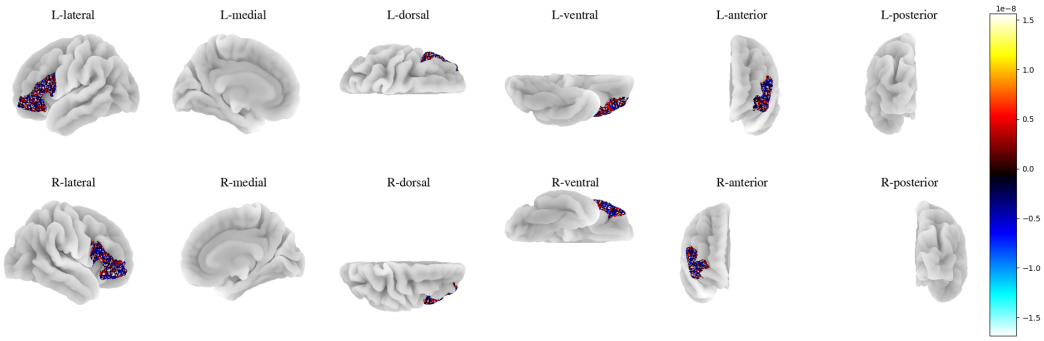

Figure 9: The averaged fMRI for the ROI: IFG for the task: ”Pie Man”

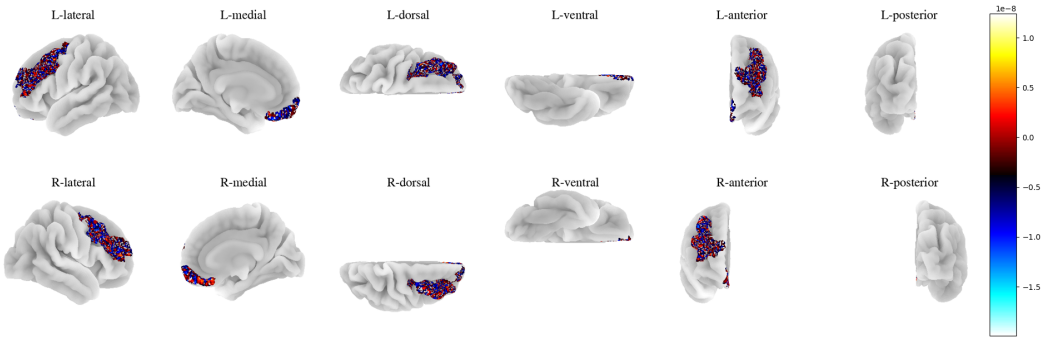

Figure 10: The averaged fMRI for the ROI: MFG for the task: ”Pie Man”

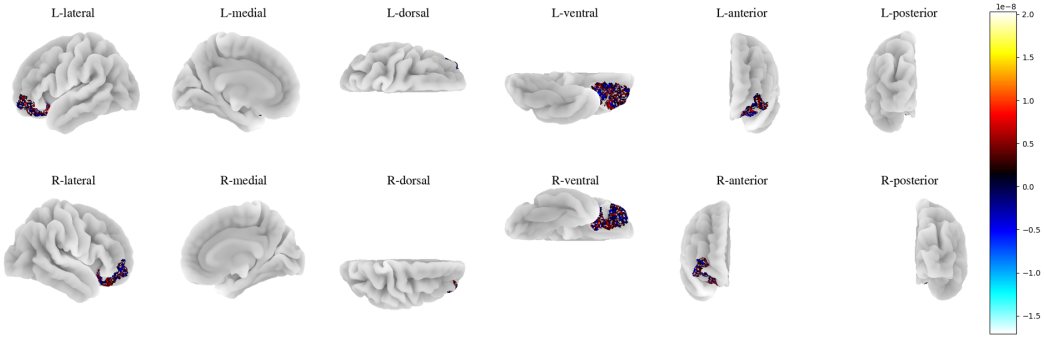

Figure 11: The averaged fMRI for the ROI: IFGorb for the task: ”Pie Man”

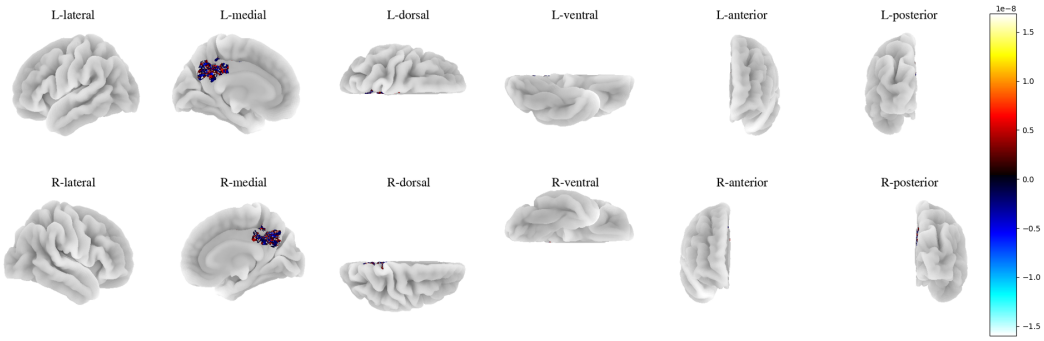

Figure 12: The averaged fMRI for the ROI: PCC for the task: ”Pie Man”

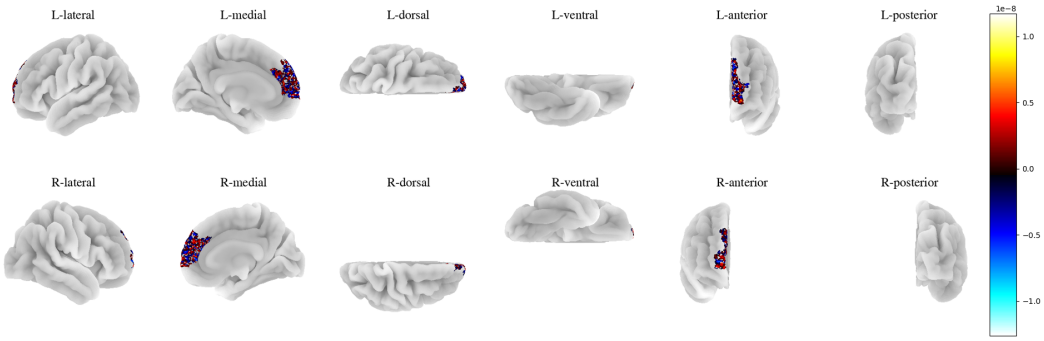

Figure 13: The averaged fMRI for the ROI: dmPFC for the task: ”Pie Man”

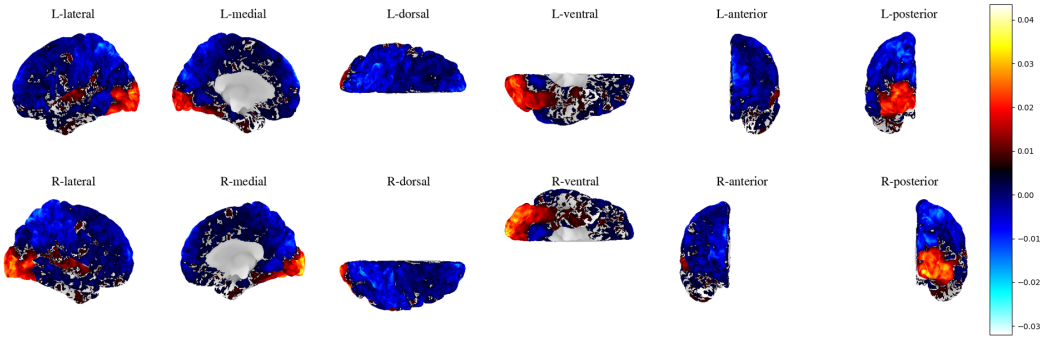

Figure 14: The averaged fMRI for the whole brain mask for the task: ”Shapes”

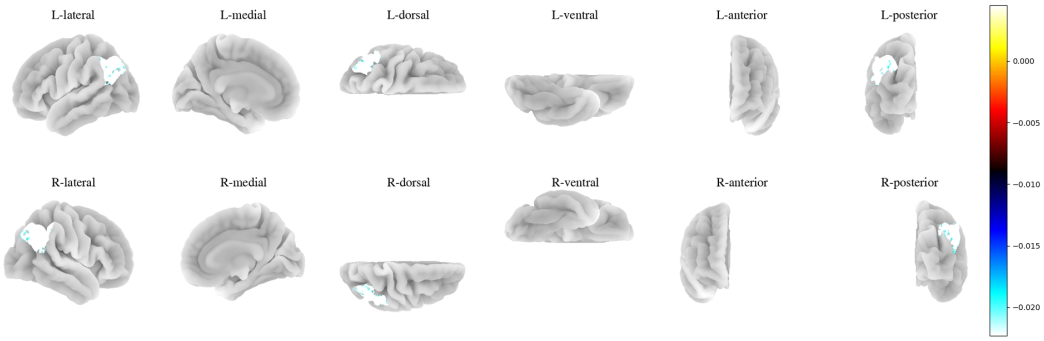

Figure 15: The averaged fMRI for the ROI: AG for the task: ”Shapes”

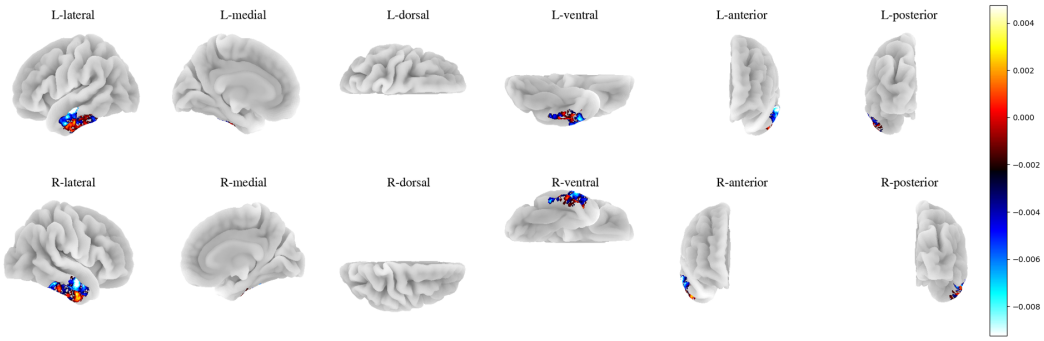

Figure 16: The averaged fMRI for the ROI: ATL for the task: ”Shapes”

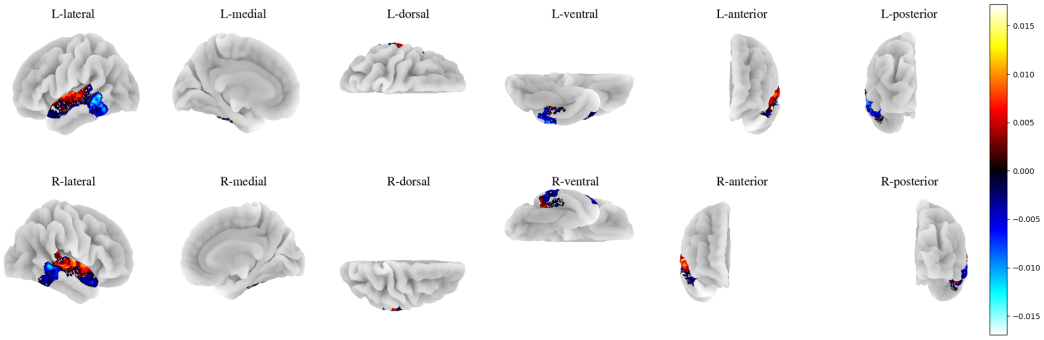

Figure 17: The averaged fMRI for the ROI: PTL for the task: ”Shapes”

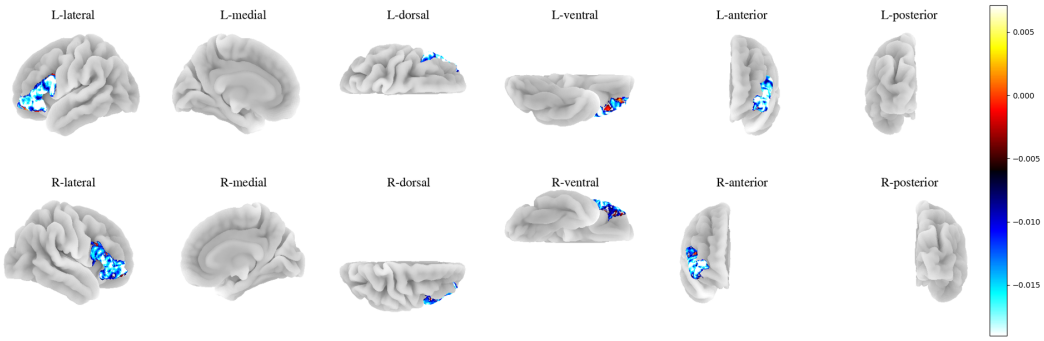

Figure 18: The averaged fMRI for the ROI: IFG for the task: ”Shapes”

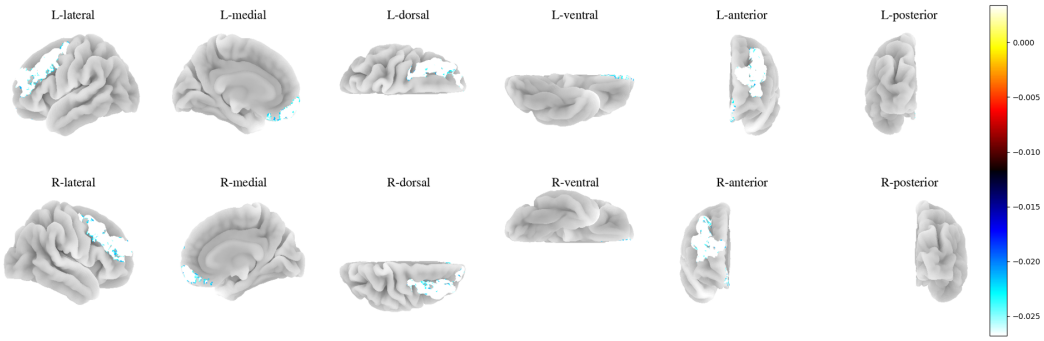

Figure 19: The averaged fMRI for the ROI: MFG for the task: ”Shapes”

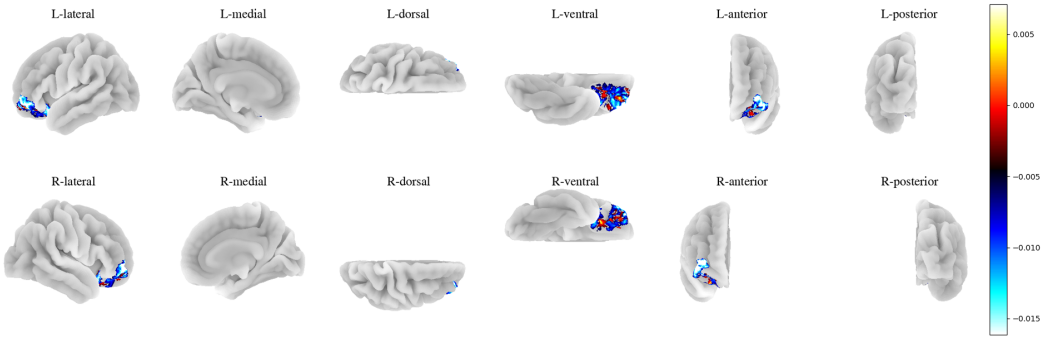

Figure 20: The averaged fMRI for the ROI: IFGorb for the task: ”Shapes”

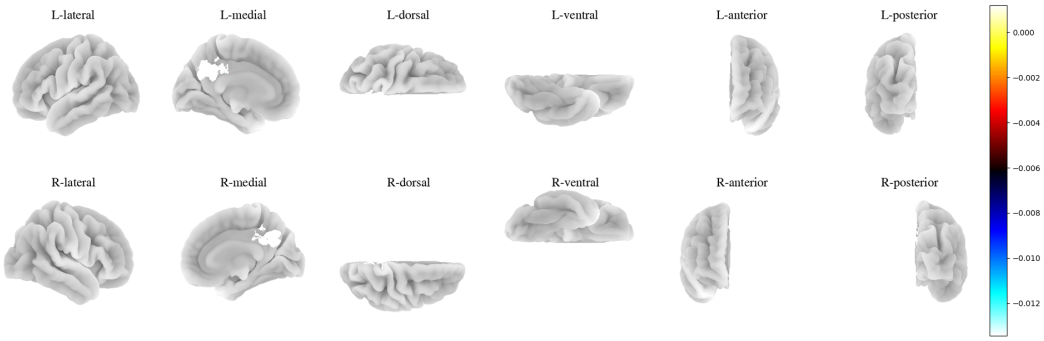

Figure 21: The averaged fMRI for the ROI: PCC for the task: ”Shapes”

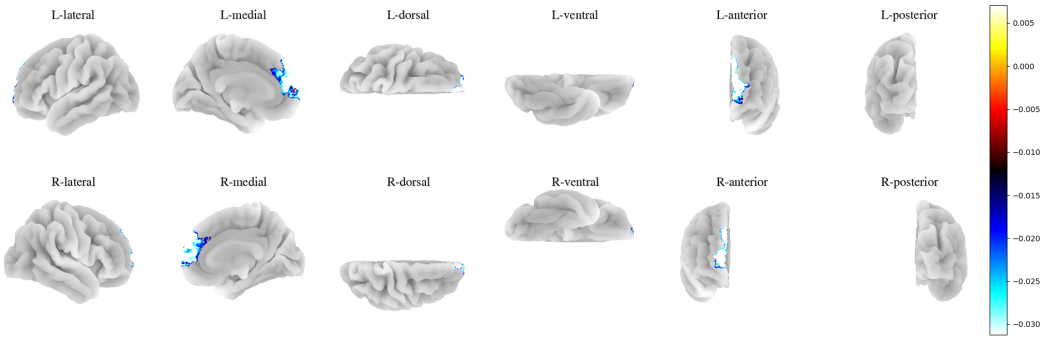

Figure 22: The averaged fMRI for the ROI: dmPFC for the task: ”Shapes”

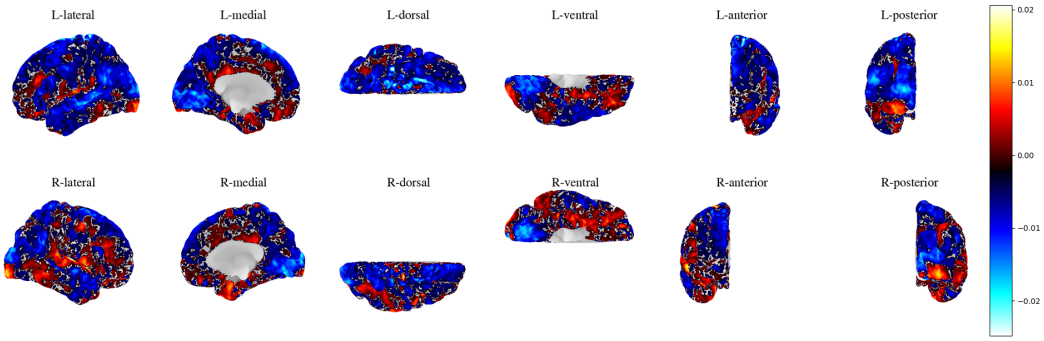

Figure 23: The averaged fMRI for the whole brain mask for the task: ”It’s Not the Fall That Gets You”

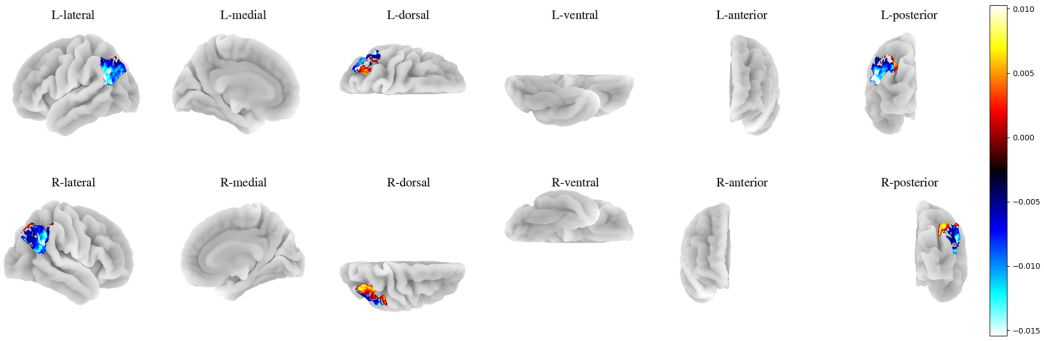

Figure 24: The averaged fMRI for the ROI: AG for the task: ”It’s Not the Fall That Gets You”

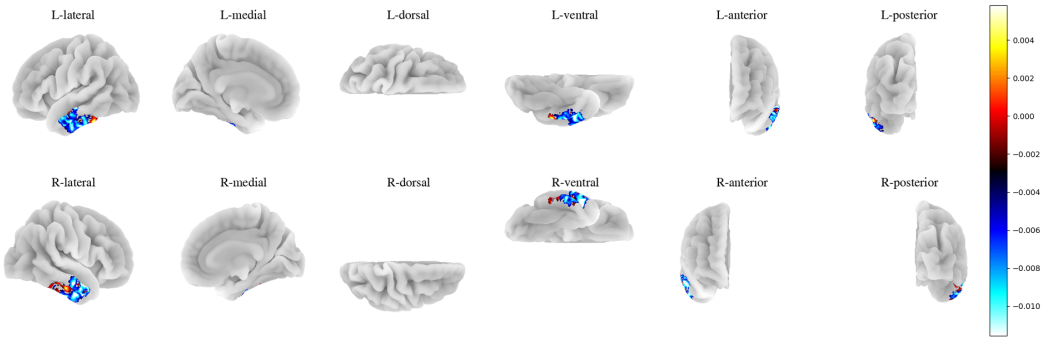

Figure 25: The averaged fMRI for the ROI: ATL for the task: ”It’s Not the Fall That Gets You”

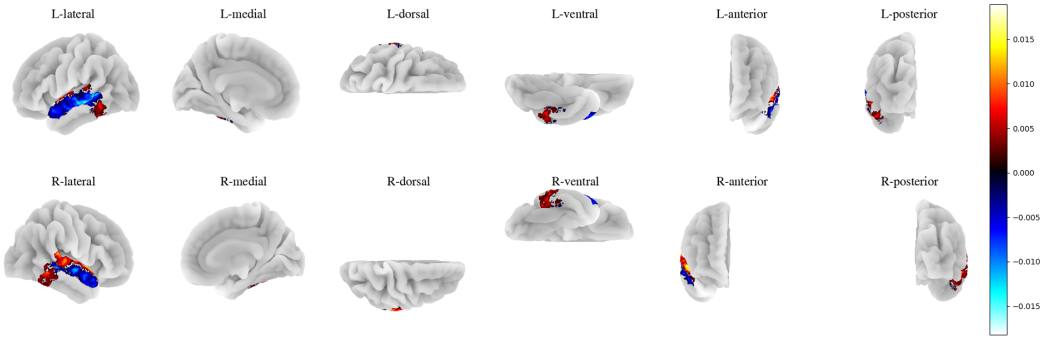

Figure 26: The averaged fMRI for the ROI: PTL for the task: ”It’s Not the Fall That Gets You”

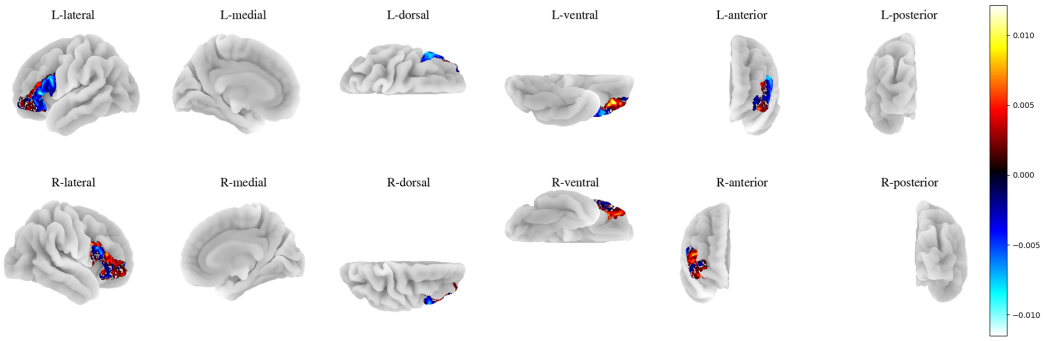

Figure 27: The averaged fMRI for the ROI: IFG for the task: ”It’s Not the Fall That Gets You”

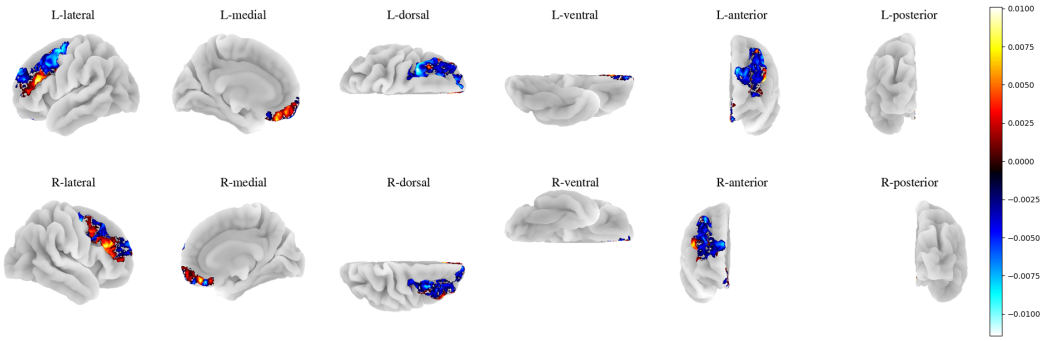

Figure 28: The averaged fMRI for the ROI: MFG for the task: ”It’s Not the Fall That Gets You”

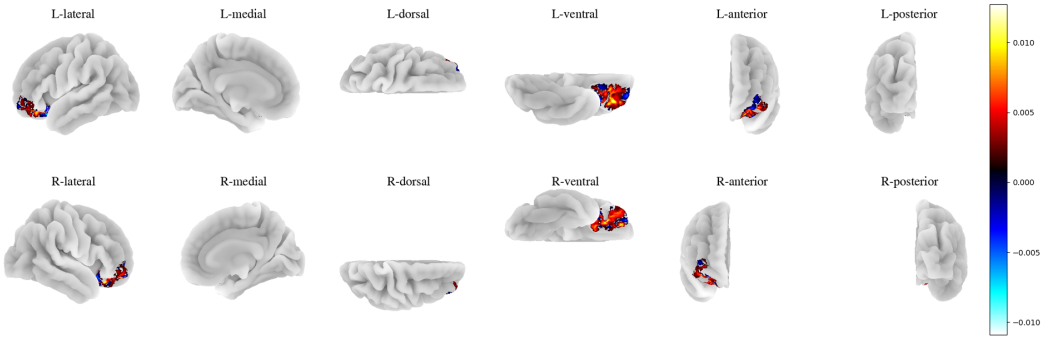

Figure 29: The averaged fMRI for the ROI: IFGorb for the task: ”It’s Not the Fall That Gets You”

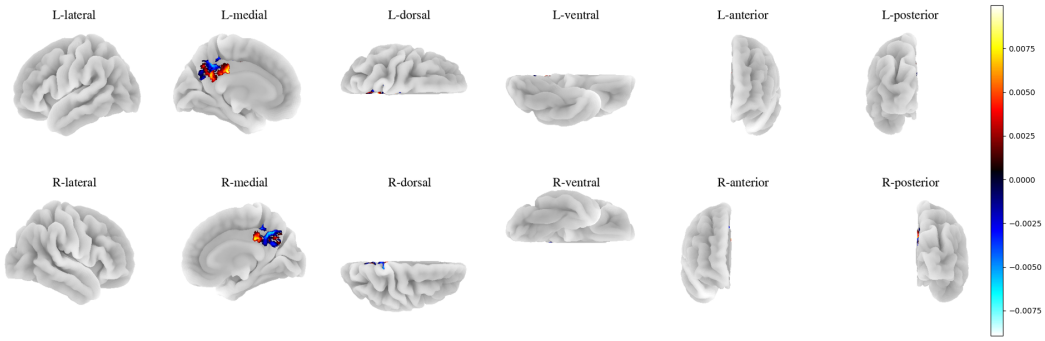

Figure 30: The averaged fMRI for the ROI: PCC for the task: "It's Not the Fall That Gets You"

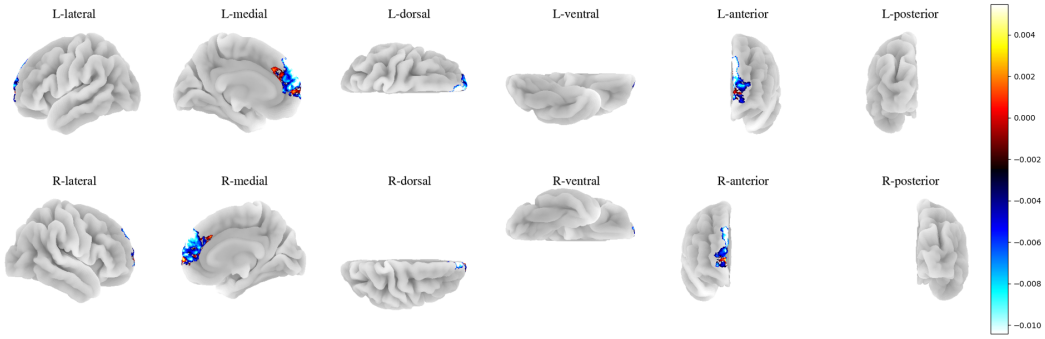

Figure 31: The averaged fMRI for the ROI: dmPFC for the task: "It's Not the Fall That Gets You"

D fMRI Data Representation Summaries

This section provides a concise overview of the fMRI data representations pertaining to our selected three tasks from the Narratives dataset (Nastase et al., 2021).

Table 2: Summary for fMRI data representations for selected three tasks from Narratives dataset (Nastase et al., 2021).

Story	# Subjects	# Hemispheres	# ROIs	Total
"Pie Man"	75	2	8+1	$75 \times 2 \times (8+1) = 1350$
"Shapes"	58	2	8+1	$58 \times 2 \times (8+1) = 1044$
"It's Not the Fall That Gets You"	54	2	8+1	$54 \times 2 \times (8+1) = 972$
Total				3,366

E LLMs Data Representation Summaries

Table 4 provides an overview of all the LLMs sampled and investigated in this study. The layer numbers underlined signify the intermediate-to-deep layers of each LLM ($l = \frac{2}{3}n_{\text{layers}}$), a detail we scrutinized closely due to its reported significance in predicting brain activity (Schrimpf et al., 2021; Caucheteux & King, 2022; Caucheteux et al., 2023).

We employed the quantized versions of Llama-2-13B and Llama-2-70B to accommodate the limitations of our NVIDIA A100 card with 40 GB GPU memory. However, the randomly initialized weights for the quantized versions of Llama-2-7B, Llama-2-13B, and Llama-2-70B were unavailable during our study. Consequently, we didn't have these three untrained LLMs for data processing.

The Hugging Face webpage URLs for each of the 39 LLMs are integrated into the model names presented in both tables: Table 3 [19] and Table 4. The references for these models can be accessed through the column "Reference" in Table 3.

Table 3: Number of Parameters, Embedding Dimensions and References for our selected 39 LLMs.

LLM	# Parameters	Embedding Dimension	Reference
distilbert-base-uncased	67M	768	(Sanh et al., 2019)
bert-base-uncased	110M	768	(Devlin et al., 2018)
bert-large-uncased	336M	1024	(Devlin et al., 2018)
bert-large-uncased-whole-word-masking	336M	1024	(Devlin et al., 2018)
distilbert-base-cased	65.8M	768	(Sanh et al., 2019)
bert-base-cased	109M	768	(Devlin et al., 2018)
bert-large-cased	335M	1024	(Devlin et al., 2018)
bert-large-cased-whole-word-masking	335M	1024	(Devlin et al., 2018)
distilroberta-base	82.8M	768	(Sanh et al., 2019)
roberta-base	125M	768	(Liu et al., 2019)
roberta-large	355M	1024	(Liu et al., 2019)
xlm-mlm-en-2048	668M	2048	(Conneau & Lample, 2019)
xlm-roberta-base	279M	768	(Conneau et al., 2019)
xlm-roberta-large	561M	1024	(Conneau et al., 2019)
xlnet-base-cased	117M	768	(Yang et al., 2019)
xlnet-large-cased	361M	1024	(Yang et al., 2019)
ctrl	1.64B	1280	(Keskar et al., 2019)
albert-base-v1	11.8M	768	(Lan et al., 2019)
albert-base-v2	11.8M	768	(Lan et al., 2019)
albert-large-v1	18.7M	1024	(Lan et al., 2019)
albert-large-v2	18.7M	1024	(Lan et al., 2019)
albert-xlarge-v1	59M	2048	(Lan et al., 2019)
albert-xlarge-v2	59M	2048	(Lan et al., 2019)
albert-xxlarge-v1	223M	4096	(Lan et al., 2019)
albert-xxlarge-v2	223M	4096	(Lan et al., 2019)
distilgpt2	88.2M	768	(Sanh et al., 2019)
gpt2	137M	768	(Radford et al., 2019)
gpt2-medium	380M	1024	(Radford et al., 2019)
gpt2-large	812M	1280	(Radford et al., 2019)
gpt2-xl	1.61B	1600	(Radford et al., 2019)
microsoft/phi-1	1.42B	2048	(Gunasekar et al., 2023)
microsoft/phi-1_5	1.42B	2048	(Li et al., 2023b)
microsoft/phi-2	2.78B	2560	(Hughes, 2023)
mosaicml/mpt-7b	6.65B	4096	(Team, 2023)
tiiuae/falcon-7b	6.93B	4544	(Penedo et al., 2023)
meta-llama/Llama-2-7b-hf	6.74B	4096	(Touvron et al., 2023)
TheBloke/Llama-2-7B-GPTQ	263M	4096	(Frantar et al., 2022) (Touvron et al., 2023)
TheBloke/Llama-2-13B-GPTQ	329M	5120	(Frantar et al., 2022) (Touvron et al., 2023)
TheBloke/Llama-2-70B-GPTQ	526M	8192	(Frantar et al., 2022) (Touvron et al., 2023)

[19]The number of parameters for each LLM is computed directly from the respective model configurations available on the Hugging Face platform.

Table 4: Summary for our selected 39 LLMs.

LLM	# Layers	Used Layers	Has Untrained?
distilbert-base-uncased	6	1,2,3,4,5,6,avg	Yes
bert-base-uncased	12	1, 2, 4, 5, 7, 8, 10, 12,avg	Yes
bert-large-uncased	24	1, 4, 7, 10, 14, 16, 17, 20, 24,avg	Yes
bert-large-uncased-whole-word-masking	24	1, 4, 7, 10, 14, 16, 17, 20, 24,avg	Yes
distilbert-base-cased	6	1,2,3,4,5,6,avg	Yes
bert-base-cased	12	1, 2, 4, 5, 7, 8, 10, 12,avg	Yes
bert-large-cased	24	1, 4, 7, 10, 14, 16, 17, 20, 24,avg	Yes
bert-large-cased-whole-word-masking	24	1, 4, 7, 10, 14, 16, 17, 20, 24,avg	Yes
distilroberta-base	6	1,2,3,4,5,6,avg	Yes
roberta-base	12	1, 2, 4, 5, 7, 8, 10, 12,avg	Yes
roberta-large	24	1, 4, 7, 10, 14, 16, 17, 20, 24,avg	Yes
xlm-mlm-en-2048	12	1, 2, 4, 5, 7, 8, 10, 12,avg	Yes
xlm-roberta-base	12	1, 2, 4, 5, 7, 8, 10, 12,avg	Yes
xlm-roberta-large	24	1, 4, 7, 10, 14, 16, 17, 20, 24,avg	Yes
xlnet-base-cased	24	1, 2, 4, 5, 7, 8, 10, 12,avg	Yes
xlnet-large-cased	24	1, 4, 7, 10, 14, 16, 17, 20, 24,avg	Yes
ctrl	48	1, 7, 14, 21, 27, 32, 34, 41, 48,avg	Yes
albert-base-v1	12	1, 2, 4, 5, 7, 8, 10, 12,avg	Yes
albert-base-v2	12	1, 2, 4, 5, 7, 8, 10, 12,avg	Yes
albert-large-v1	24	1, 4, 7, 10, 14, 16, 17, 20, 24,avg	Yes
albert-large-v2	24	1, 4, 7, 10, 14, 16, 17, 20, 24,avg	Yes
albert-xlarge-v1	24	1, 4, 7, 10, 14, 16, 17, 20, 24,avg	Yes
albert-xlarge-v2	24	1, 4, 7, 10, 14, 16, 17, 20, 24,avg	Yes
albert-xxlarge-v1	12	1, 2, 4, 5, 7, 8, 10, 12,avg	Yes
albert-xxlarge-v2	12	1, 2, 4, 5, 7, 8, 10, 12,avg	Yes
distilgpt2	6	1,2,3,4,5,6,avg	Yes
gpt2	12	1, 2, 4, 5, 7, 8, 10, 12,avg	Yes
gpt2-medium	24	1, 4, 7, 10, 14, 16, 17, 20, 24,avg	Yes
gpt2-large	36	1, 6, 11, 16, 21, 24, 26, 31, 36,avg	Yes
gpt2-xl	48	1, 7, 14, 21, 27, 32, 34, 41, 48,avg	Yes
microsoft/phi-1	24	1, 4, 7, 10, 14, 16, 17, 20, 24,avg	Yes
microsoft/phi-1_5	24	1, 4, 7, 10, 14, 16, 17, 20, 24,avg	Yes
microsoft/phi-2	32	1, 5, 9, 14, 18, 21, 23, 27, 32,avg	Yes
mosaicml/mpt-7b	32	1, 5, 9, 14, 18, 21, 23, 27, 32,avg	Yes
tiiuae/falcon-7b	32	1, 5, 9, 14, 18, 21, 23, 27, 32,avg	Yes
meta-llama/Llama-2-7b-hf	32	1, 5, 9, 14, 18, 21, 23, 27, 32,avg	Yes
TheBloke/Llama-2-7B-GPTQ	32	1, 5, 9, 14, 18, 21, 23, 27, 32,avg	No
TheBloke/Llama-2-13B-GPTQ	40	1, 6, 12, 17, 23, 26, 28, 34, 40,avg	No
TheBloke/Llama-2-70B-GPTQ	80	1, 12, 23, 34, 46, 53, 57, 68, 80,avg	No

F Illustrations for Characterizing the fMRI Data Representations by PH

This section elucidates the process by which we converted the fMRI Data Representations into 1-D time-series vectors, followed by projecting them into 3-D spaces as point clouds, and subsequently computed the Persistent Homology (PH) for these point clouds, depicted through summary representations such as persistence diagrams and persistence barcodes. Further elaboration on this process is provided in Appendix A.2.1.

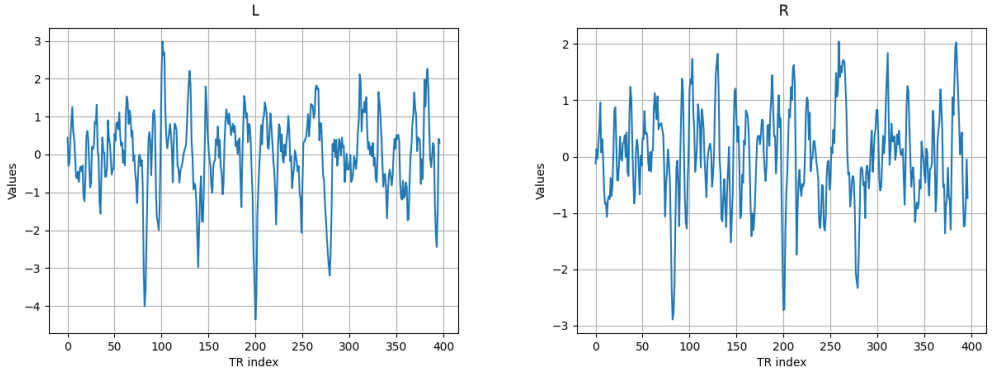

Figure 32: This figure depicts the time-series of the fMRI BOLD values for both the (L)eft and (R)ight hemispheres and the region of interest (ROI): PCC. These values are averaged across 54 out of 56 subjects while they listened to the task "It's Not the Fall that Gets You," and subsequently aggregated along the TR dimension.

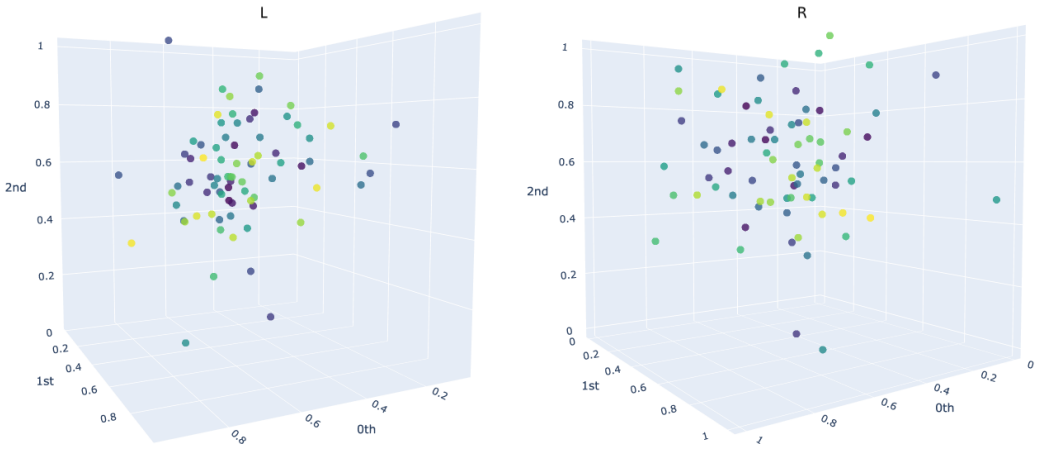

Figure 33: This figure presents the projected point cloud in 3-D space for the time-series signals illustrated in Figure 32. In this specific instance, the embedding dimension is constrained to 3, and the optimal time delay is determined to be $\tau = 78$ for the (L)eft hemisphere and $\tau = 76$ for the (R)ight one. This result pertains to the task "It's Not the Fall that Gets You," which comprised a total of 397 valid TRs.

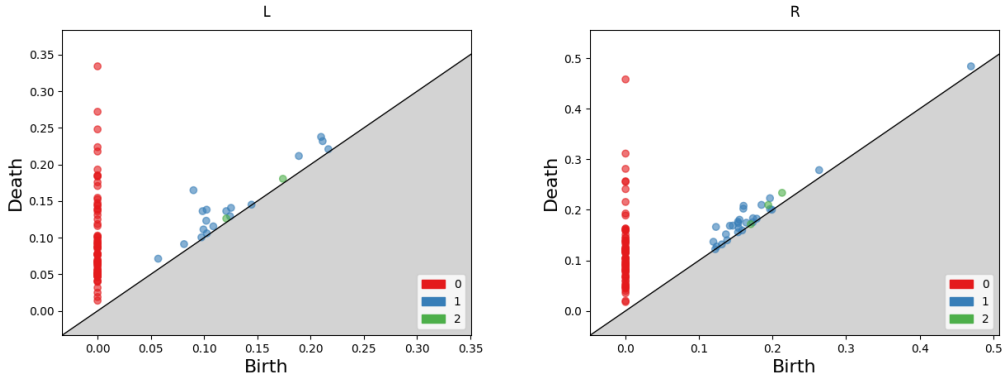

Figure 34: This figure depicts the persistence diagram derived from the Persistent Homology analysis conducted on the point cloud detailed in Figure 33.

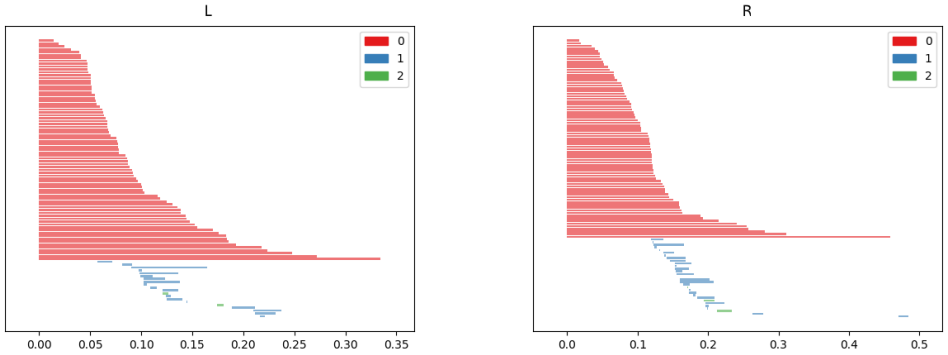

Figure 35: This figure depicts the persistence barcode derived from the Persistent Homology analysis conducted on the point cloud detailed in Figure 33.

G ILLUSTRATIONS FOR CHARACTERIZING THE LLMS EMBEDDINGS BY PH

This section elucidates the process by which we converted the LLM Embeddings into 1-D time-series vectors, then projected them into 3-D spaces as point clouds, and subsequently computed the Persistent Homology for these point clouds. The resultant summary representations, in the form of persistence diagrams and persistence barcodes, are illustrated herein. Further details regarding this process can be found in Appendix A.2.2.

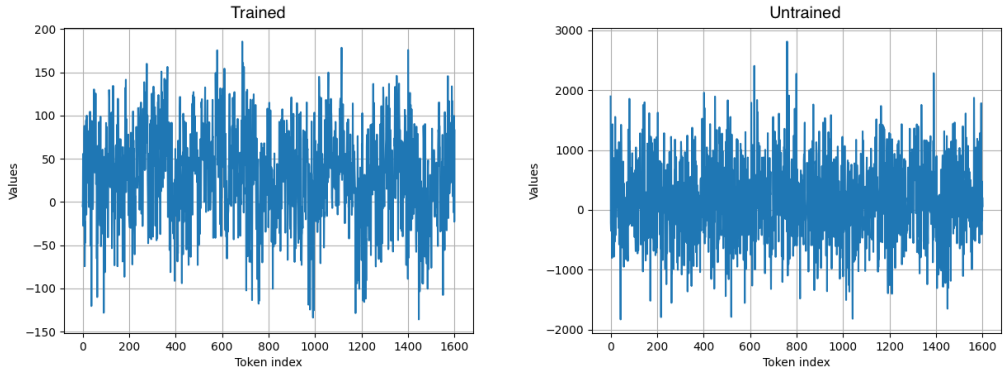

Figure 36: This figure presents the time-series for the embedding from Layer 21 of the LLM meta-llama/Llama-2-7b-hf, both trained and untrained, when provided with the task "It's not the Fall that Gets You". The embeddings are then aggregated along the token dimension. Layer 21 from meta-llama/Llama-2-7b-hf corresponds to the intermediate-to-deep layer of the LLM ($l = \frac{2}{3}n_{\text{layers}}$), as defined by Caucheteux et al. (2023).

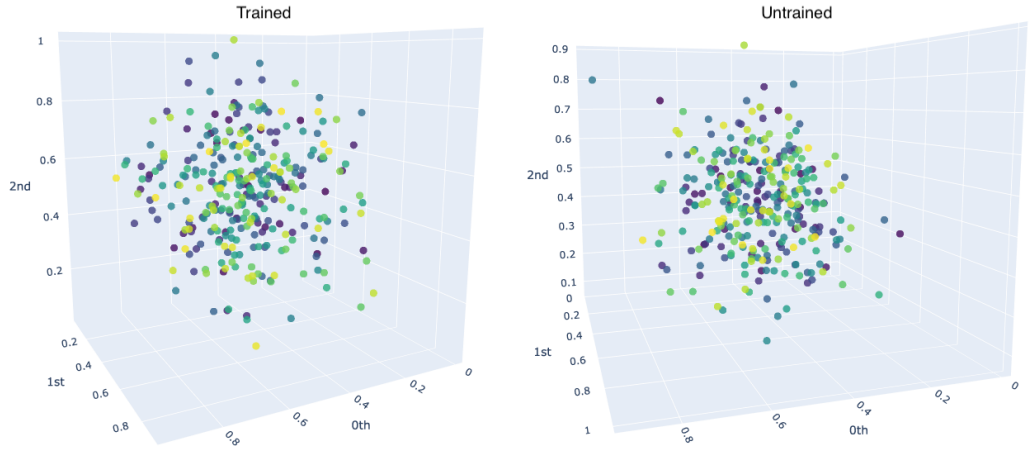

Figure 37: This figure depicts the projected point cloud in 3-D space for the time-series signals specified in Figure 36. In this specific instance, the embedding dimension is constrained to 3, with the optimal time delay searched being $\tau = 320$ for the trained model and $\tau = 318$ for the untrained one. The total number of tokens for the task "It's not the Fall that Gets You" is 1601.

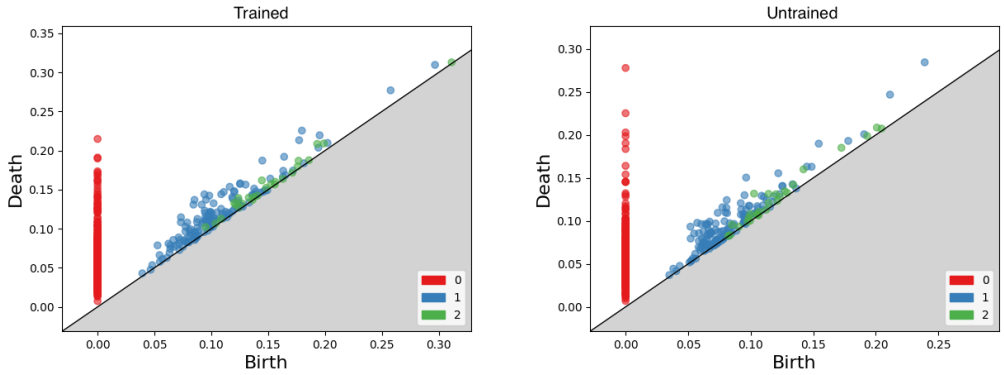

Figure 38: This figure presents the persistence diagram summarized from the Persistent Homology computed from the point cloud specified in Figure 37.

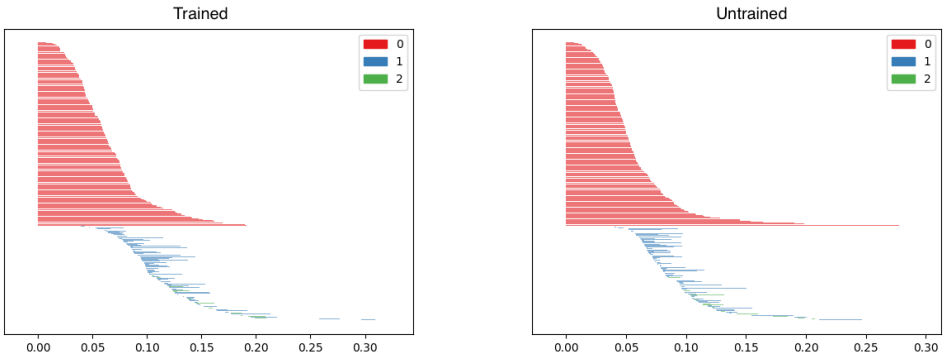

Figure 39: This figure presents the persistence barcode summarized from the Persistent Homology computed from the point cloud specified in Figure 37.

H Exploratory Data Analysis on Constructed Features: q-Wasserstein Distances

This section illustrates the q-Wasserstein Distances computed between the persistence diagrams for fMRI data from ROI: PCC, hemisphere: L, and embeddings from LLM: Llama-2-7b-hf, training status: trained, layer: 21, under the same task: "It's not the Fall That Gets You". Each figure displays the results for all three persistent homology dimensions: 0, 1, and 2, respectively, with the parameter ranges for q and p marked in each figure to denote the differences.

Across the q-Wasserstein Distances, we observe "long tail" distributions, where the values converge to approximately 0.30 for persistent homology dimension 0, and approximately 0.14 and 0.07 for persistent homology dimensions 1 and 2, respectively, in this specific example among all the 34,416 pairs in Appendix A.3. It is noteworthy that this distribution pattern is not unique to this particular example; rather, it is prevalent across all 34,416 pairs. The variations lie only in the specific values to which the q-Wasserstein Distances ultimately converge for each of the three persistent homology dimensions. Formally, the q-Wasserstein Distance is bounded by the Wasserstein Stability Theorem (Edelsbrunner & Harer, 2010; Cohen-Steiner et al., 2010).

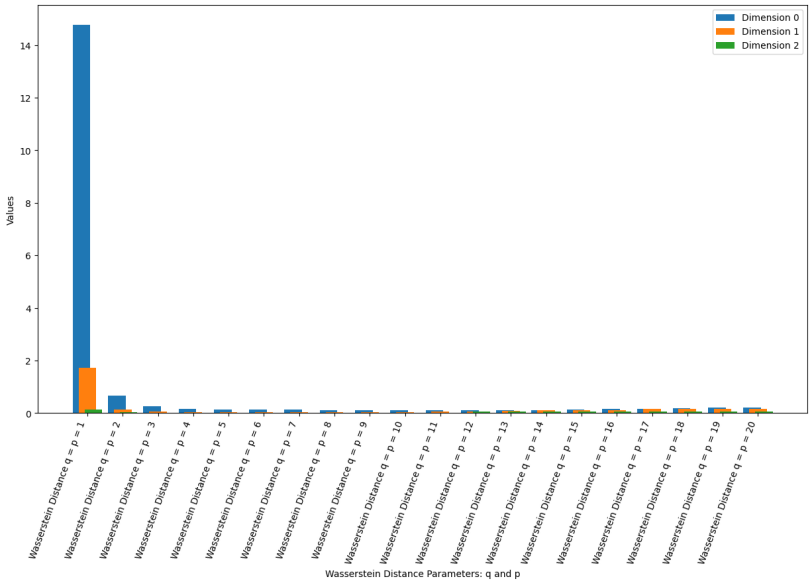

Figure 40: The figure depicts the q-Wasserstein Distances with parameters q and p spanning from $q = p = 1$ to $q = p = 20$.

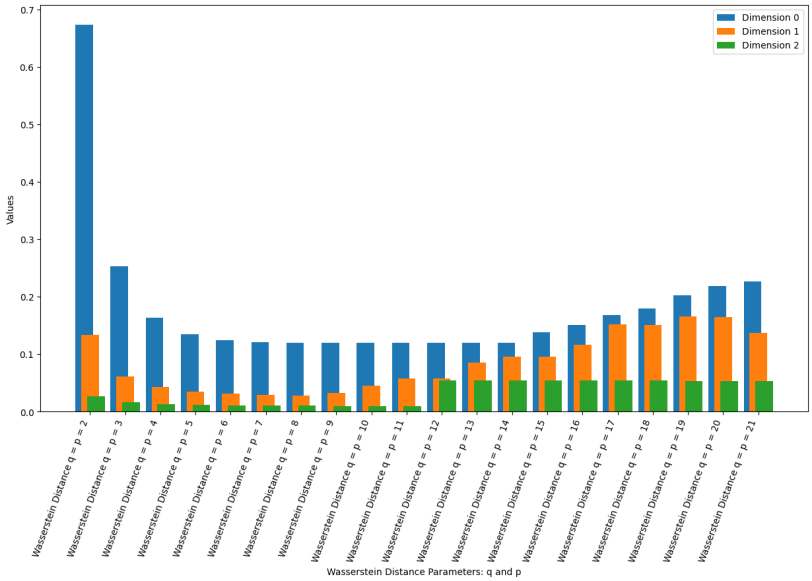

Figure 41: The figure depicts the q-Wasserstein Distances with parameters q and p spanning from $q = p = 2$ to $q = p = 21$.

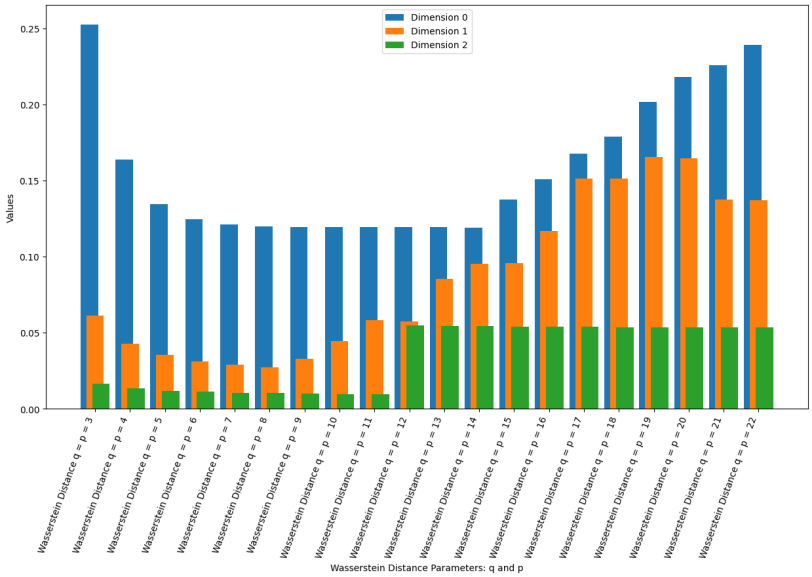

Figure 42: The figure depicts the q-Wasserstein Distances with parameters q and p spanning from $q = p = 3$ to $q = p = 22$.

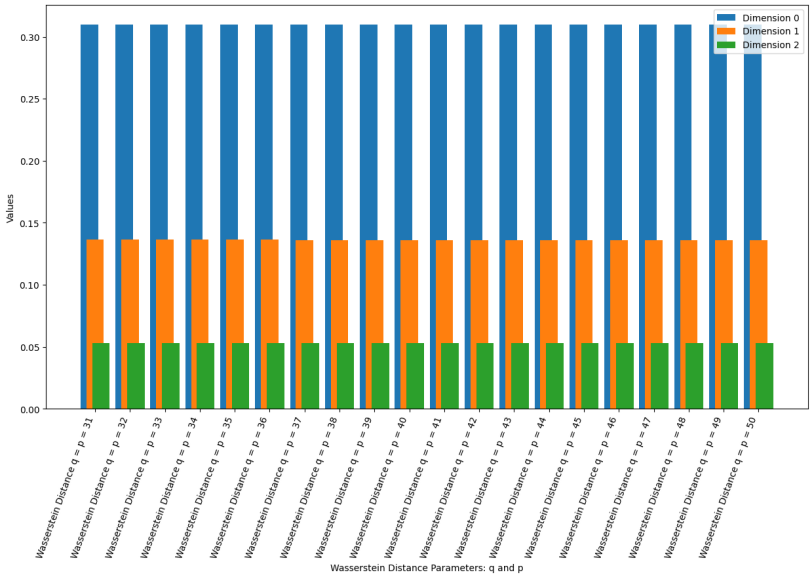

Figure 43: The figure depicts the q-Wasserstein Distances with parameters q and p spanning from $q = p = 31$ to $q = p = 50$.

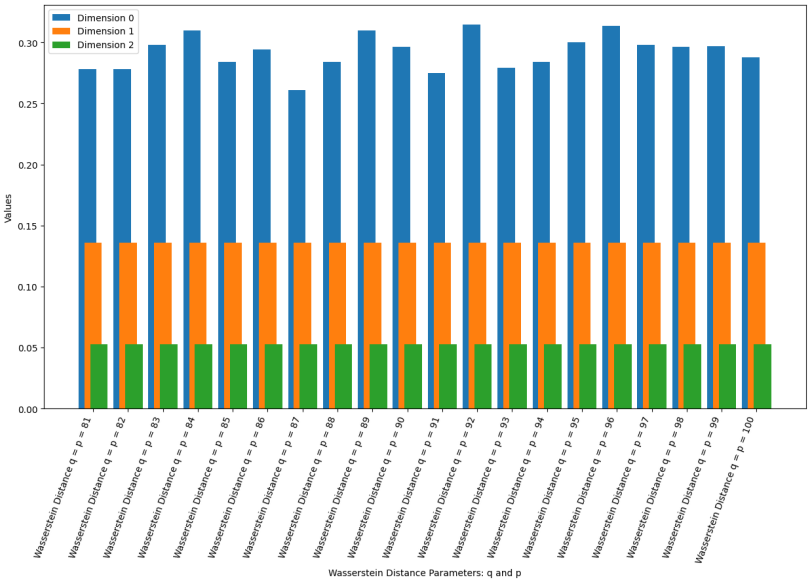

Figure 44: The figure depicts the q-Wasserstein Distances with parameters q and p spanning from $q = p = 81$ to $q = p = 100$.

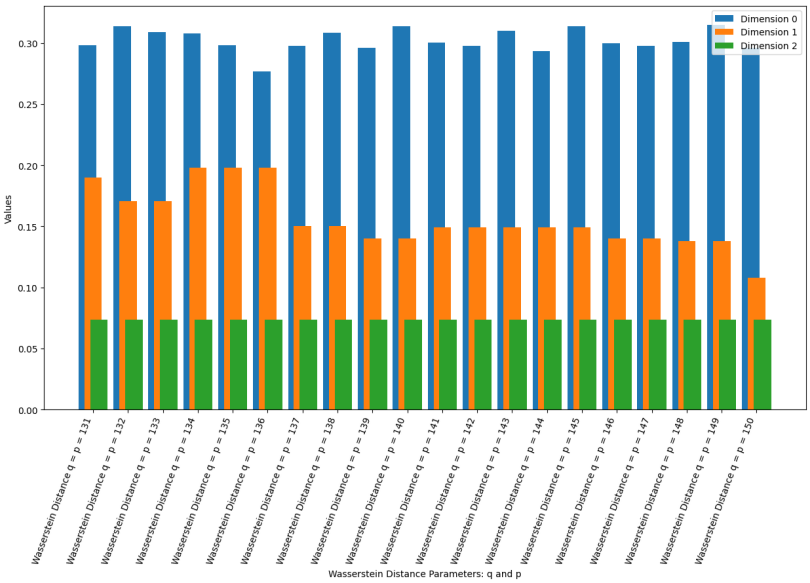

Figure 45: The figure depicts the q-Wasserstein Distances with parameters q and p spanning from $q = p = 131$ to $q = p = 150$.

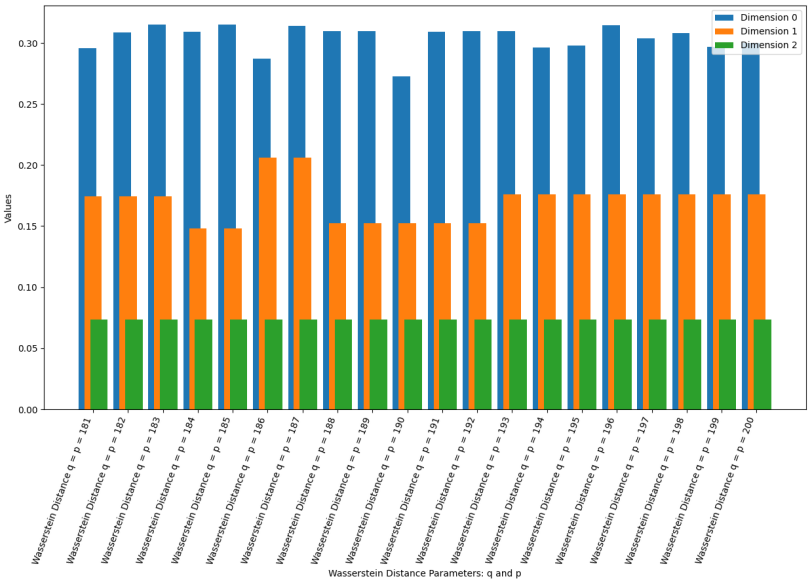

Figure 46: The figure depicts the q-Wasserstein Distances with parameters q and p spanning from $q = p = 181$ to $q = p = 200$.

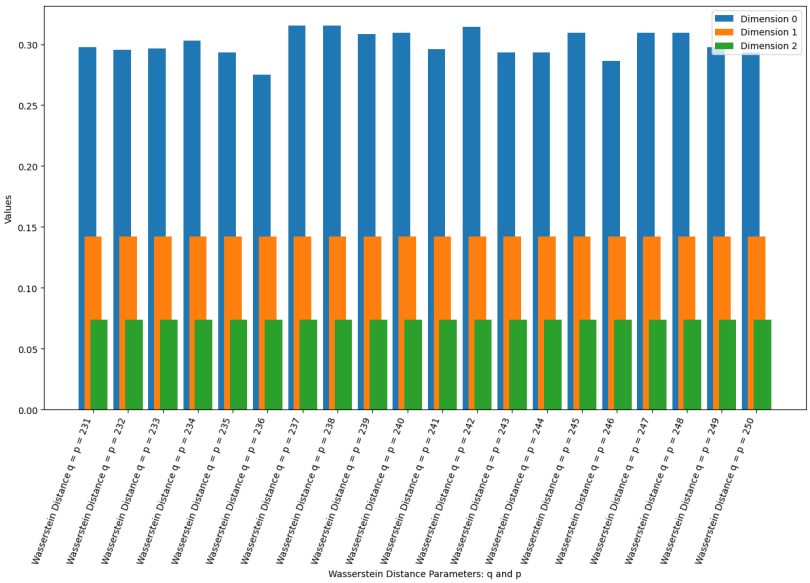

Figure 47: The figure depicts the q-Wasserstein Distances with parameters q and p spanning from $q = p = 231$ to $q = p = 250$.

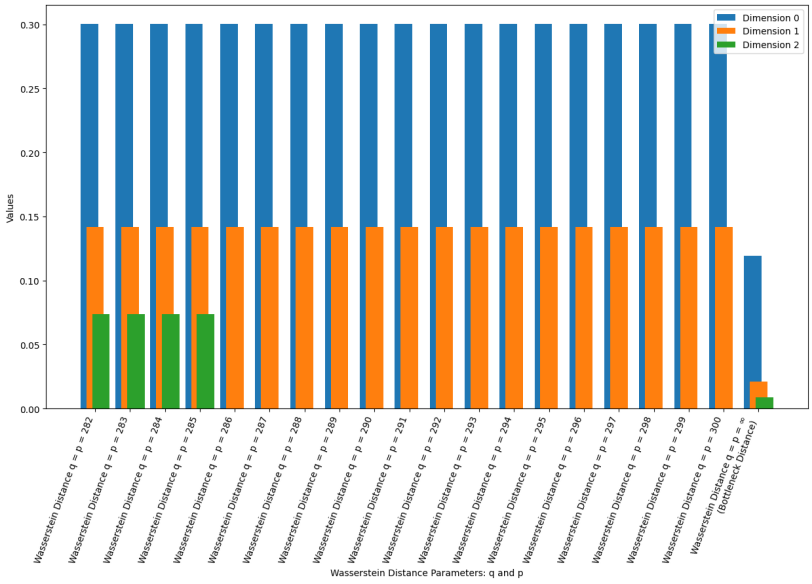

Figure 48: The figure presents the q-Wasserstein Distances with parameters q and p varying from $q = p = 282$ to $q = p = \infty$. In this specific instance, the distance is a.k.a Bottleneck Distance.

I FILTER IN RELIABLE AND VALID FEATURES: FIRST PASS

This section provides an illustrative example of our filtering process "First Pass" outlined in Appendix A.4. Specifically, the example focuses on the Region of Interest (ROI): PCC, considering both "Only Trained" LLMs and a combination of "Trained plus Untrained" ones.

At the end of this section, we present the maximum value of q for each Region of Interest (ROI) across each hemisphere, which corresponds to the highest averaged test R^2 score achieved during the "First Pass", as detailed in Appendix A.4.1.

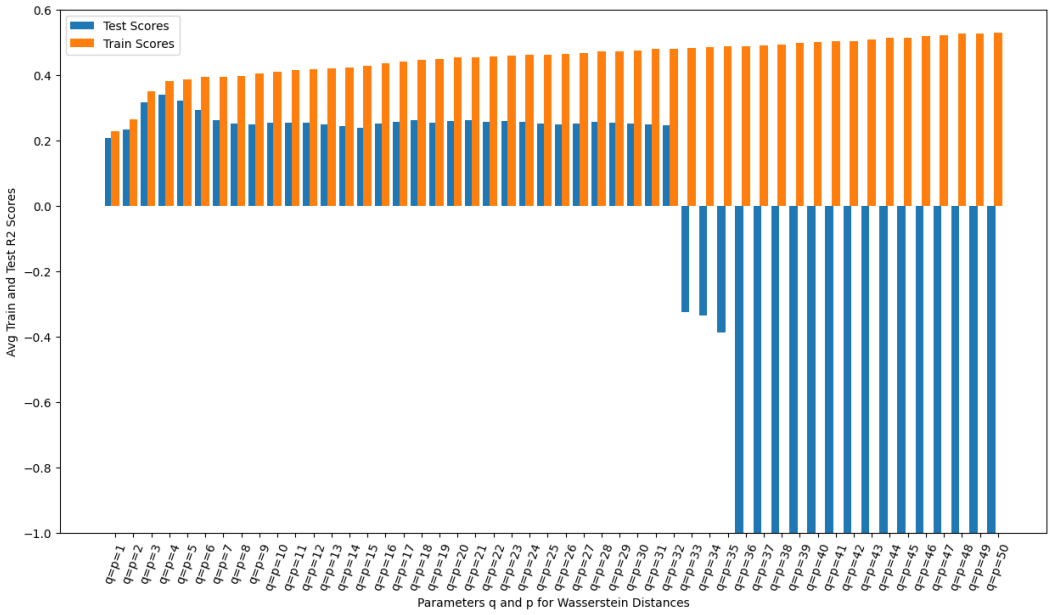

Figure 49: The figure depicts the averaged train and test R^2 scores acquired during the learning process of the Linear Regression model to correlate with the respective brainscores using "Only Trained" LLMs. The train scores exhibit a gradual increase as additional q-Wasserstein Distances are incorporated progressively. Conversely, the test scores deviate from the pattern, reaching a peak at $q = p = 4$ before sharply declining, with a significant downturn observed beyond $q = p = 35$ (we just hide those very large negative test scores below -1, refer to Table 5 and Table 6 for details). It is noteworthy that the trend observed in the test scores is not exclusive to "Only Trained" LLMs; it is pervasive across "Trained plus Untrained" LLMs, as well as both the L(eft) and R(ight) hemispheres and all ROIs. The critical distinction lies in the specific q values where the test score peaks and subsequently descends into negative territory.

Table 5: The table compiles the averaged **train** R^2 scores across four scenarios: the L(eft) and R(ight) Hemispheres, considering whether the training dataset comprises "Only Trained" or "Trained plus Untrained" LLMs. The q-Wasserstein Distances are progressively included as features for training the Linear Regression Model, ranging from $q = p = 1$ to $q = p = 50$.

$q(p)$ VALUE	L ONLY TRAINED	R ONLY TRAINED	L TRAINED AND UNTRAINED	R TRAINED AND UNTRAINED
$q = p = 1$	$2.2676e{-1}$	$2.8019e{-1}$	$1.9626e{-1}$	$2.1948e{-1}$
$q = p = 2$	$2.6528e{-1}$	$3.0415e{-1}$	$2.6011e{-1}$	$2.8302e{-1}$
$q = p = 3$	$3.5044e{-1}$	$3.7017e{-1}$	$3.3508e{-1}$	$3.4951e{-1}$
$q = p = 4$	$3.8208e{-1}$	$3.8145e{-1}$	$3.5319e{-1}$	$3.5761e{-1}$
$q = p = 5$	$3.8609e{-1}$	$3.8510e{-1}$	$3.5758e{-1}$	$3.6216e{-1}$
$q = p = 6$	$3.9301e{-1}$	$3.8735e{-1}$	$3.6052e{-1}$	$3.6619e{-1}$
$q = p = 7$	$3.9482e{-1}$	$3.9077e{-1}$	$3.6230e{-1}$	$3.6840e{-1}$
$q = p = 8$	$3.9646e{-1}$	$3.9167e{-1}$	$3.7142e{-1}$	$3.6907e{-1}$
$q = p = 9$	$4.0346e{-1}$	$3.9569e{-1}$	$3.8007e{-1}$	$3.7048e{-1}$
$q = p = 10$	$4.0933e{-1}$	$3.9800e{-1}$	$3.8601e{-1}$	$3.7439e{-1}$
$q = p = 11$	$4.1479e{-1}$	$4.0077e{-1}$	$3.9041e{-1}$	$3.7749e{-1}$
$q = p = 12$	$4.1723e{-1}$	$4.0173e{-1}$	$3.9141e{-1}$	$3.7866e{-1}$
$q = p = 13$	$4.1881e{-1}$	$4.0354e{-1}$	$3.9192e{-1}$	$3.8125e{-1}$
$q = p = 14$	$4.2224e{-1}$	$4.0605e{-1}$	$3.9326e{-1}$	$3.8477e{-1}$
$q = p = 15$	$4.2834e{-1}$	$4.0894e{-1}$	$3.9729e{-1}$	$3.8780e{-1}$
$q = p = 16$	$4.3450e{-1}$	$4.0994e{-1}$	$4.0117e{-1}$	$3.9079e{-1}$
$q = p = 17$	$4.4185e{-1}$	$4.1091e{-1}$	$4.0410e{-1}$	$3.9106e{-1}$
$q = p = 18$	$4.4598e{-1}$	$4.1226e{-1}$	$4.0634e{-1}$	$3.9196e{-1}$
$q = p = 19$	$4.4742e{-1}$	$4.1395e{-1}$	$4.0721e{-1}$	$3.9242e{-1}$
$q = p = 20$	$4.5304e{-1}$	$4.2334e{-1}$	$4.0868e{-1}$	$3.9586e{-1}$
$q = p = 21$	$4.5470e{-1}$	$4.2466e{-1}$	$4.0905e{-1}$	$3.9676e{-1}$
$q = p = 22$	$4.5631e{-1}$	$4.2585e{-1}$	$4.0950e{-1}$	$3.9750e{-1}$
$q = p = 23$	$4.5956e{-1}$	$4.2699e{-1}$	$4.1044e{-1}$	$3.9854e{-1}$
$q = p = 24$	$4.6116e{-1}$	$4.3547e{-1}$	$4.1130e{-1}$	$4.0235e{-1}$
$q = p = 25$	$4.6171e{-1}$	$4.4021e{-1}$	$4.1180e{-1}$	$4.0539e{-1}$
$q = p = 26$	$4.6372e{-1}$	$4.4153e{-1}$	$4.1297e{-1}$	$4.0669e{-1}$
$q = p = 27$	$4.6627e{-1}$	$4.4698e{-1}$	$4.1423e{-1}$	$4.0975e{-1}$
$q = p = 28$	$4.7133e{-1}$	$4.4992e{-1}$	$4.1742e{-1}$	$4.1135e{-1}$
$q = p = 29$	$4.7193e{-1}$	$4.5163e{-1}$	$4.1820e{-1}$	$4.1225e{-1}$
$q = p = 30$	$4.7406e{-1}$	$4.5370e{-1}$	$4.1925e{-1}$	$4.1628e{-1}$
$q = p = 31$	$4.7962e{-1}$	$4.6614e{-1}$	$4.2364e{-1}$	$4.2017e{-1}$
$q = p = 32$	$4.8062e{-1}$	$4.6891e{-1}$	$4.2486e{-1}$	$4.2158e{-1}$
$q = p = 33$	$4.8186e{-1}$	$4.7294e{-1}$	$4.2567e{-1}$	$4.2331e{-1}$
$q = p = 34$	$4.8566e{-1}$	$4.7470e{-1}$	$4.2822e{-1}$	$4.2413e{-1}$
$q = p = 35$	$4.8657e{-1}$	$4.7610e{-1}$	$4.2857e{-1}$	$4.2449e{-1}$
$q = p = 36$	$4.8875e{-1}$	$4.7986e{-1}$	$4.2935e{-1}$	$4.2619e{-1}$
$q = p = 37$	$4.9102e{-1}$	$4.8264e{-1}$	$4.3021e{-1}$	$4.2736e{-1}$
$q = p = 38$	$4.9382e{-1}$	$4.9000e{-1}$	$4.3099e{-1}$	$4.3422e{-1}$
$q = p = 39$	$4.9731e{-1}$	$4.9332e{-1}$	$4.3311e{-1}$	$4.3571e{-1}$
$q = p = 40$	$5.0121e{-1}$	$5.0561e{-1}$	$4.3550e{-1}$	$4.4134e{-1}$
$q = p = 41$	$5.0265e{-1}$	$5.0855e{-1}$	$4.3625e{-1}$	$4.4330e{-1}$
$q = p = 42$	$5.0424e{-1}$	$5.0902e{-1}$	$4.3733e{-1}$	$4.4418e{-1}$
$q = p = 43$	$5.0766e{-1}$	$5.1294e{-1}$	$4.3981e{-1}$	$4.4527e{-1}$
$q = p = 44$	$5.1244e{-1}$	$5.1654e{-1}$	$4.4163e{-1}$	$4.4744e{-1}$
$q = p = 45$	$5.1410e{-1}$	$5.1992e{-1}$	$4.4265e{-1}$	$4.4914e{-1}$
$q = p = 46$	$5.1848e{-1}$	$5.2137e{-1}$	$4.4355e{-1}$	$4.5063e{-1}$
$q = p = 47$	$5.2083e{-1}$	$5.2489e{-1}$	$4.4484e{-1}$	$4.5307e{-1}$
$q = p = 48$	$5.2561e{-1}$	$5.2628e{-1}$	$4.4699e{-1}$	$4.5348e{-1}$
$q = p = 49$	$5.2772e{-1}$	$5.2814e{-1}$	$4.4831e{-1}$	$4.5483e{-1}$
$q = p = 50$	$5.2951e{-1}$	$5.3148e{-1}$	$4.4872e{-1}$	$4.5604e{-1}$

Table 6: The table compiles the averaged **test** R^2 scores across four scenarios: the L(eft) and R(ight) Hemispheres, considering whether the training dataset comprises "Only Trained" or "Trained plus Untrained" LLMs. The q-Wasserstein Distances are progressively included as features for training the Linear Regression Model, ranging from $q = p = 1$ to $q = p = 50$.

$q(p)$ VALUE	L ONLY TRAINED	R ONLY TRAINED	L TRAINED AND UNTRAINED	R TRAINED AND UNTRAINED
$q = p = 1$	$2.0603e{-1}$	$2.5393e{-1}$	$1.8616e{-1}$	$2.0459e{-1}$
$q = p = 2$	$2.3421e{-1}$	$2.6107e{-1}$	$2.4498e{-1}$	$2.5628e{-1}$
$q = p = 3$	$3.1531e{-1}$	$3.3301e{-1}$	$3.1849e{-1}$	$3.2631e{-1}$
$q = p = 4$	$3.3954e{-1}$	$3.4015e{-1}$	$3.3480e{-1}$	$3.3212e{-1}$
$q = p = 5$	$3.2199e{-1}$	$3.3588e{-1}$	$3.3036e{-1}$	$3.3411e{-1}$
$q = p = 6$	$2.9304e{-1}$	$3.3314e{-1}$	$3.1853e{-1}$	$3.3560e{-1}$
$q = p = 7$	$2.6157e{-1}$	$3.2523e{-1}$	$3.0384e{-1}$	$3.3400e{-1}$
$q = p = 8$	$2.5110e{-1}$	$3.1828e{-1}$	$3.0736e{-1}$	$3.3102e{-1}$
$q = p = 9$	$2.4856e{-1}$	$3.1503e{-1}$	$3.1189e{-1}$	$3.2760e{-1}$
$q = p = 10$	$2.5271e{-1}$	$3.0938e{-1}$	$3.1651e{-1}$	$3.2852e{-1}$
$q = p = 11$	$2.5458e{-1}$	$3.0568e{-1}$	$3.1946e{-1}$	$3.2645e{-1}$
$q = p = 12$	$2.5511e{-1}$	$2.9922e{-1}$	$3.1697e{-1}$	$3.2432e{-1}$
$q = p = 13$	$2.4993e{-1}$	$2.9481e{-1}$	$3.1485e{-1}$	$3.2252e{-1}$
$q = p = 14$	$2.4366e{-1}$	$2.9236e{-1}$	$3.1423e{-1}$	$3.2412e{-1}$
$q = p = 15$	$2.3893e{-1}$	$2.8742e{-1}$	$3.1676e{-1}$	$3.2639e{-1}$
$q = p = 16$	$2.5084e{-1}$	$2.8472e{-1}$	$3.2079e{-1}$	$3.2775e{-1}$
$q = p = 17$	$2.5561e{-1}$	$2.8285e{-1}$	$3.2194e{-1}$	$3.2608e{-1}$
$q = p = 18$	$2.6034e{-1}$	$2.7908e{-1}$	$3.2266e{-1}$	$3.2417e{-1}$
$q = p = 19$	$2.5477e{-1}$	$2.7730e{-1}$	$3.2078e{-1}$	$3.2158e{-1}$
$q = p = 20$	$2.6030e{-1}$	$2.8381e{-1}$	$3.2033e{-1}$	$3.2274e{-1}$
$q = p = 21$	$2.6076e{-1}$	$2.8038e{-1}$	$3.1821e{-1}$	$3.2089e{-1}$
$q = p = 22$	$2.5601e{-1}$	$2.7905e{-1}$	$3.1649e{-1}$	$3.1885e{-1}$
$q = p = 23$	$2.5837e{-1}$	$2.7532e{-1}$	$3.1515e{-1}$	$3.1664e{-1}$
$q = p = 24$	$2.5548e{-1}$	$2.8413e{-1}$	$3.1456e{-1}$	$3.1879e{-1}$
$q = p = 25$	$2.5206e{-1}$	$2.8468e{-1}$	$3.1260e{-1}$	$3.2021e{-1}$
$q = p = 26$	$2.4848e{-1}$	$2.7970e{-1}$	$3.1212e{-1}$	$3.1858e{-1}$
$q = p = 27$	$2.5065e{-1}$	$2.8108e{-1}$	$3.1118e{-1}$	$3.1892e{-1}$
$q = p = 28$	$2.5681e{-1}$	$2.7902e{-1}$	$3.1319e{-1}$	$3.1824e{-1}$
$q = p = 29$	$2.5490e{-1}$	$2.7208e{-1}$	$3.1301e{-1}$	$3.1640e{-1}$
$q = p = 30$	$2.5025e{-1}$	$2.6217e{-1}$	$3.1275e{-1}$	$3.1849e{-1}$
$q = p = 31$	$2.4757e{-1}$	$2.7185e{-1}$	$3.1357e{-1}$	$3.1888e{-1}$
$q = p = 32$	$2.4519e{-1}$	$2.6816e{-1}$	$3.1397e{-1}$	$3.1772e{-1}$
$q = p = 33$	$-3.2415e{-1}$	$2.6636e{-1}$	$2.5329e{-1}$	$3.1610e{-1}$
$q = p = 34$	$-3.3517e{-1}$	$-2.2066e1$	$2.5427e{-1}$	$-1.0702e1$
$q = p = 35$	$-3.8890e{-1}$	$-1.9224e1$	$2.2239e{-1}$	$-1.0613e1$
$q = p = 36$	$-1.5627e1$	$-2.9217e1$	$-7.6096e{-1}$	$-1.1058e1$
$q = p = 37$	$-1.3481e1$	$-8.9582e3$	$-9.2275e1$	$-5.4787e1$
$q = p = 38$	$-1.5631e1$	$-6.8164e4$	$-9.2302e1$	$-7.1104e1$
$q = p = 39$	$-3.6913e1$	$-1.0058e5$	$-1.5284e2$	$-7.2362e1$
$q = p = 40$	$-3.4904e1$	$-1.1385e6$	$-1.3840e2$	$-8.0620e3$
$q = p = 41$	$-5.7584e3$	$-1.5004e6$	$-1.4184e2$	$-1.0231e4$
$q = p = 42$	$-3.6511e8$	$-1.4485e6$	$-2.1095e3$	$-6.3590e4$
$q = p = 43$	$-2.0746e8$	$-4.7278e8$	$-4.9774e3$	$-4.7846e7$
$q = p = 44$	$-2.5970e8$	$-5.8001e10$	$-3.7748e3$	$-1.5573e10$
$q = p = 45$	$-1.6952e8$	$-5.4524e10$	$-3.3020e4$	$-1.3105e10$
$q = p = 46$	$-1.3855e8$	$-5.9515e10$	$-1.2126e5$	$-1.5097e10$
$q = p = 47$	$-1.3019e12$	$-7.1320e10$	$-6.4879e11$	$-1.9357e10$
$q = p = 48$	$-9.4337e15$	$-4.9336e12$	$-3.6670e15$	$-7.1362e10$
$q = p = 49$	$-9.5105e15$	$-1.9473e16$	$-3.8326e15$	$-6.7940e10$
$q = p = 50$	$-2.8495e15$	$-1.0416e18$	$-8.7427e14$	$-1.1248e13$

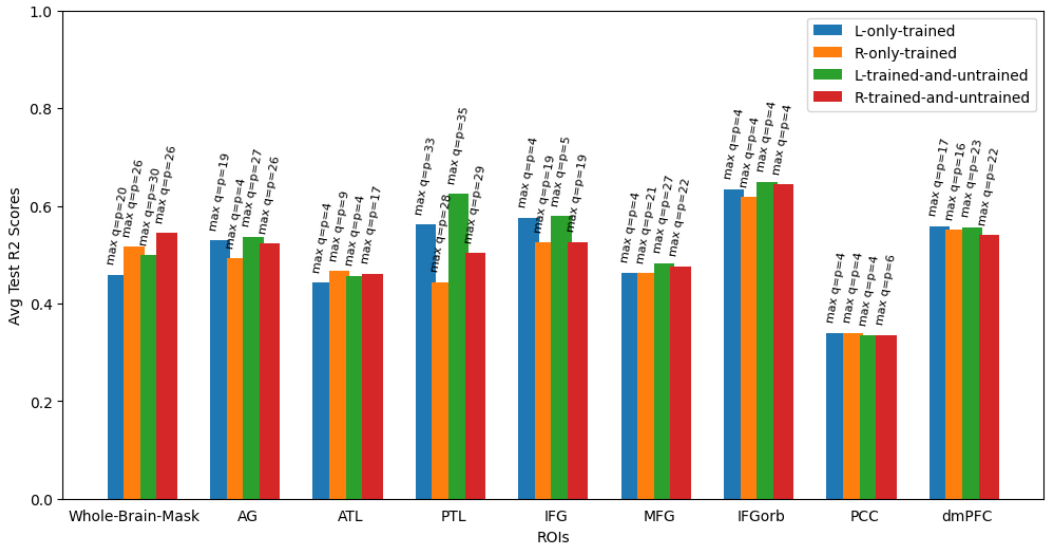

Figure 50: The figure depicts the maximum q value corresponding to each ROI and hemisphere, which yields the highest averaged test R^2 score, as elucidated in our results for the First Pass detailed in Appendix A.4.1. Additionally, it employs distinct colors to denote whether the training dataset encompasses untrained LLMs.

J FILTER IN RELIABLE AND VALID FEATURES: SECOND PASS

This section provides the detailed results stemming from the Second Pass, as delineated in Section 3. Specifically, we conducted further filtration to eliminate insignificant features by scrutinizing whether each averaged p value satisfies $p < 5\%$, as detailed in Appendix A.4.2. Within this section, each table encapsulates the features that ultimately passed through the filtering process [20] [21] for each ROI (including the whole brain mask) and hemisphere (L and R) with each feature's Weight (plus both Lower and Upper Bound of 95% Confidence Intervals), the Standard Error (SE) across all 25 runs, the Feature Importance t-statistic value $|t|$, and the corresponding p value [22].

Each table is then further accompanied with two figures in its following, illustrating each feature's Weight across both Lower and Upper Bound of 95% Confidence Intervals, and its Coefficient Importance and Variability respectively.

[20] The feature "Intercept" is consistently included, albeit not utilized for interpretations, and is listed regardless of its p value.

[21] We exclusively present features derived from the Second Pass in Appendix A.4.2, ensuring their compliance with $p < 5\%$.

[22] All the above values are averaged across all the 25 runs except for the Standard Error (SE).

Table 7: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: Whole Brain Mask and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$1.9346e{-2}$	$1.8800e{-4}$	$3.8504e{-2}$	$5.8683e{-3}$	$3.4248e0$	$8.1760e{-2}$
DIM 0 $q = p = 1$	$9.4292e{-3}$	$6.6046e{-3}$	$1.2254e{-2}$	$1.0329e{-3}$	$8.0441e0$	$0.0000e0$
DIM 0 $q = p = 2$	$-3.5147e{-1}$	$-4.8224e{-1}$	$-2.2070e{-1}$	$3.4591e{-2}$	$9.3502e0$	$0.0000e0$
DIM 0 $q = p = 3$	$1.0834e0$	$5.0134e{-1}$	$1.6655e0$	$1.7312e{-1}$	$5.3600e0$	$7.4000e{-3}$
DIM 0 $q = p = 20$	$8.3403e{-2}$	$-4.1044e{-3}$	$1.7091e{-1}$	$2.4431e{-2}$	$4.3134e0$	$3.7840e{-2}$
DIM 0 $q = p = \infty$	$5.2090e{-1}$	$2.0966e{-1}$	$8.3214e{-1}$	$5.4023e{-2}$	$1.0681e1$	$4.8000e{-4}$
DIM 1 $q = p = 1$	$3.5372e{-2}$	$7.2709e{-3}$	$6.3473e{-2}$	$1.0270e{-2}$	$3.4113e0$	$4.5280e{-2}$
DIM 1 $q = p = 2$	$-2.1050e0$	$-3.5764e0$	$-6.3351e{-1}$	$4.6105e{-1}$	$4.0441e0$	$3.9760e{-2}$
DIM 1 $q = p = \infty$	$-1.3601e0$	$-2.5898e0$	$-1.3045e{-1}$	$3.1620e{-1}$	$5.8714e0$	$7.4800e{-3}$
DIM 2 $q = p = 11$	$-1.9798e{-1}$	$-3.3485e{-1}$	$-6.1120e{-2}$	$3.0268e{-2}$	$6.4753e0$	$8.9200e{-3}$
DIM 2 $q = p = 12$	$-1.5506e{-1}$	$-2.5847e{-1}$	$-5.1657e{-2}$	$1.9166e{-2}$	$8.0050e0$	$6.8000e{-3}$
DIM 2 $q = p = 16$	$-7.5968e{-2}$	$-1.2718e{-1}$	$-2.4759e{-2}$	$1.3664e{-2}$	$5.5662e0$	$1.0080e{-2}$
DIM 2 $q = p = 20$	$5.2582e{-2}$	$1.8756e{-2}$	$8.6409e{-2}$	$9.2219e{-3}$	$4.5193e0$	$3.6480e{-2}$

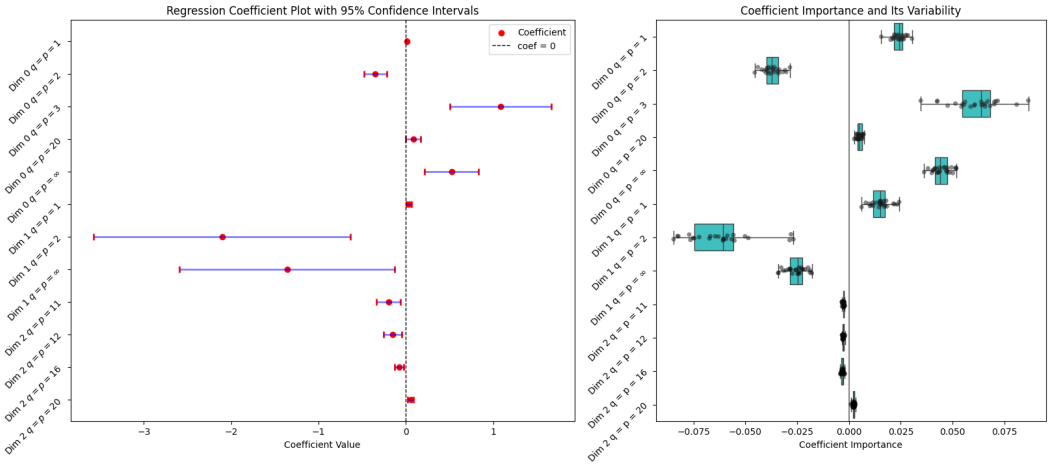

Figure 51: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: Whole Brain Mask and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 8: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: Whole Brain Mask and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$1.2627e{-2}$	$-3.8263e{-3}$	$2.9080e{-2}$	$5.2724e{-3}$	$3.5872e0$	$5.1160e{-2}$
DIM 0 $q = p = 1$	$7.3745e{-3}$	$5.4757e{-3}$	$9.2733e{-3}$	$5.3823e{-4}$	$1.4333e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-3.1132e{-1}$	$-4.0518e{-1}$	$-2.1746e{-1}$	$2.6914e{-2}$	$1.1603e1$	$0.0000e0$
DIM 0 $q = p = 3$	$9.6782e{-1}$	$5.5116e{-1}$	$1.3845e0$	$1.2708e{-1}$	$7.4729e0$	$8.0000e{-5}$
DIM 0 $q = p = 4$	$-1.4886e0$	$-2.5840e0$	$-3.9314e{-1}$	$3.0016e{-1}$	$5.3065e0$	$1.2000e{-2}$
DIM 0 $q = p = 20$	$8.0842e{-2}$	$1.7004e{-2}$	$1.4468e{-1}$	$1.7817e{-2}$	$4.3624e0$	$3.4440e{-2}$
DIM 0 $q = p = 30$	$-7.9946e{-2}$	$-1.1570e{-1}$	$-4.4192e{-2}$	$1.0071e{-2}$	$8.6739e0$	$4.0000e{-5}$
DIM 1 $q = p = 1$	$4.2343e{-2}$	$2.2814e{-2}$	$6.1872e{-2}$	$5.7902e{-3}$	$7.3898e0$	$3.6000e{-4}$
DIM 1 $q = p = 2$	$-2.0656e0$	$-3.1950e0$	$-9.3612e{-1}$	$3.1271e{-1}$	$7.5218e0$	$2.8000e{-4}$
DIM 1 $q = p = 3$	$7.2762e0$	$9.4592e{-1}$	$1.3607e1$	$1.8620e0$	$5.0218e0$	$1.0920e{-2}$
DIM 1 $q = p = 18$	$-9.5618e{-2}$	$-1.5525e{-1}$	$-3.5987e{-2}$	$1.5532e{-2}$	$5.1992e0$	$1.6440e{-2}$
DIM 1 $q = p = \infty$	$-1.1059e0$	$-1.9578e0$	$-2.5407e{-1}$	$2.6410e{-1}$	$4.5394e0$	$2.0000e{-2}$
DIM 2 $q = p = 12$	$-1.5540e{-1}$	$-2.3276e{-1}$	$-7.8033e{-2}$	$2.3939e{-2}$	$6.3398e0$	$1.0400e{-3}$
DIM 2 $q = p = 16$	$-3.2496e{-2}$	$-7.2192e{-2}$	$7.1994e{-3}$	$1.1286e{-2}$	$5.1708e0$	$1.1600e{-2}$
DIM 2 $q = p = 23$	$5.2076e{-2}$	$2.0964e{-2}$	$8.3189e{-2}$	$6.5853e{-3}$	$6.8184e0$	$8.6000e{-3}$
DIM 2 $q = p = \infty$	$-4.7243e0$	$-8.2617e0$	$-1.1869e0$	$8.5644e{-1}$	$5.6357e0$	$1.6480e{-2}$

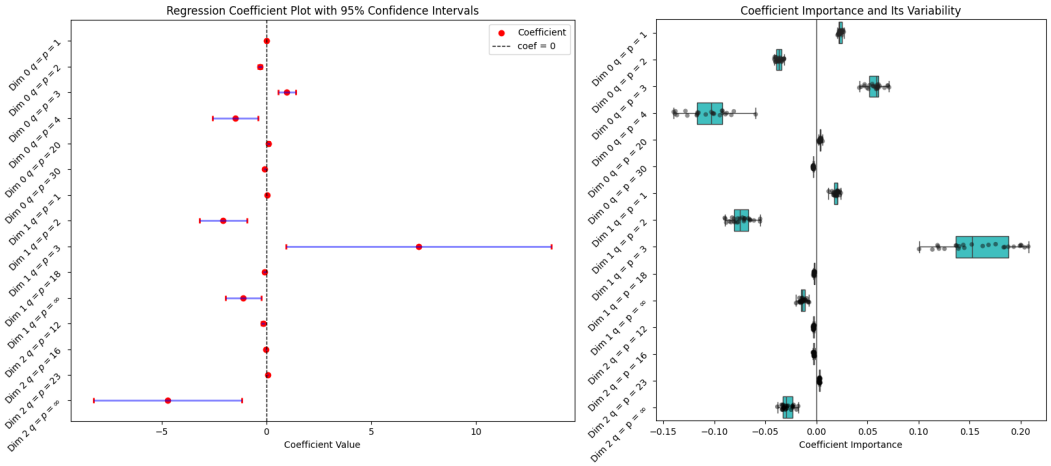

Figure 52: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: Whole Brain Mask and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 9: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: Whole Brain Mask and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$2.2858e{-3}$	$-2.1022e{-2}$	$2.5594e{-2}$	$6.4020e{-3}$	$1.7450e0$	$4.0984e{-1}$
DIM 0 $q = p = 1$	$-3.5462e{-3}$	$-6.4684e{-3}$	$-6.2392e{-4}$	$7.9883e{-4}$	$4.7242e0$	$2.4320e{-2}$
DIM 0 $q = p = 11$	$-3.8843e0$	$-5.9861e0$	$-1.7826e0$	$5.7617e{-1}$	$6.9440e0$	$2.4400e{-3}$
DIM 0 $q = p = 12$	$6.4340e{-1}$	$-2.5351e{-1}$	$1.5403e0$	$2.0817e{-1}$	$4.9566e0$	$4.0440e{-2}$
DIM 0 $q = p = 14$	$4.3350e{-1}$	$1.5303e{-2}$	$8.5169e{-1}$	$9.7953e{-2}$	$5.1920e0$	$2.9000e{-2}$
DIM 0 $q = p = 23$	$8.1616e{-2}$	$2.5986e{-3}$	$1.6063e{-1}$	$1.8969e{-2}$	$4.5890e0$	$4.8920e{-2}$
DIM 0 $q = p = 26$	$-9.9791e{-2}$	$-1.4632e{-1}$	$-5.3259e{-2}$	$1.5084e{-2}$	$5.6716e0$	$2.1600e{-3}$
DIM 1 $q = p = 11$	$-3.6901e{-1}$	$-6.1385e{-1}$	$-1.2417e{-1}$	$6.3202e{-2}$	$5.3309e0$	$1.7080e{-2}$
DIM 1 $q = p = 26$	$1.4624e{-1}$	$5.9673e{-2}$	$2.3281e{-1}$	$2.3238e{-2}$	$4.1770e0$	$4.8880e{-2}$
DIM 2 $q = p = 22$	$7.4633e{-2}$	$2.8737e{-2}$	$1.2053e{-1}$	$1.2754e{-2}$	$4.3268e0$	$3.6320e{-2}$
DIM 2 $q = p = 26$	$-6.6502e{-2}$	$-1.2655e{-1}$	$-6.4532e{-3}$	$1.2490e{-2}$	$5.9284e0$	$2.6280e{-2}$

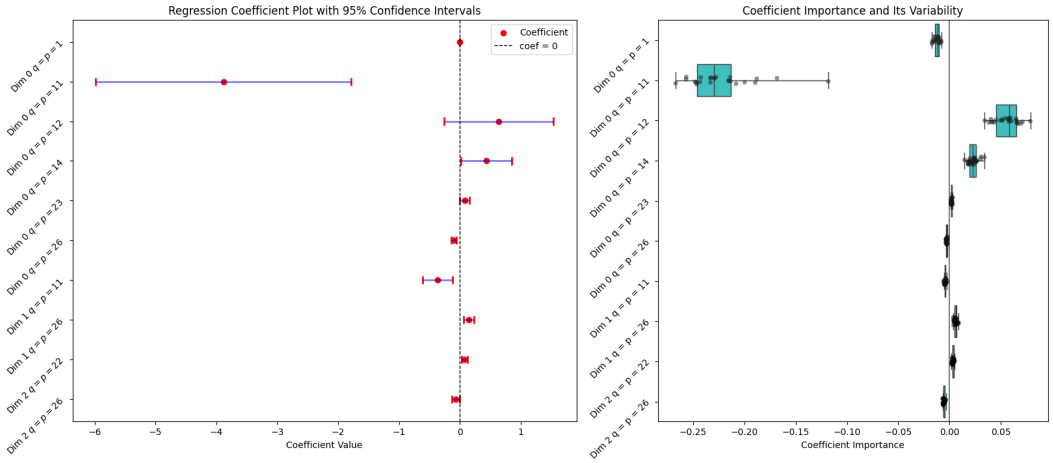

Figure 53: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: Whole Brain Mask and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 10: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: Whole Brain Mask and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$1.5559e{-2}$	$-2.6929e{-3}$	$3.3811e{-2}$	$4.7705e{-3}$	$2.2174e0$	$3.1344e{-1}$
DIM 0 $q = p = 2$	$-1.8406e{-1}$	$-3.1568e{-1}$	$-5.2445e{-2}$	$3.8945e{-2}$	$3.8918e0$	$5.0000e{-2}$
DIM 0 $q = p = 3$	$1.3628e0$	$5.3039e{-1}$	$2.1953e0$	$2.2354e{-1}$	$5.6215e0$	$1.0760e{-2}$
DIM 0 $q = p = 4$	$-3.6328e0$	$-6.2187e0$	$-1.0470e0$	$6.7469e{-1}$	$5.0968e0$	$2.5160e{-2}$
DIM 0 $q = p = 20$	$1.0482e{-1}$	$1.6801e{-2}$	$1.9284e{-1}$	$2.1982e{-2}$	$5.2974e0$	$1.7800e{-2}$
DIM 0 $q = p = 21$	$1.0367e{-1}$	$3.1106e{-2}$	$1.7624e{-1}$	$1.4984e{-2}$	$6.1088e0$	$2.2720e{-2}$
DIM 0 $q = p = 26$	$-5.8867e{-2}$	$-9.2287e{-2}$	$-2.5447e{-2}$	$8.2864e{-3}$	$7.5092e0$	$8.8000e{-4}$
DIM 1 $q = p = 1$	$1.4654e{-2}$	$-1.9281e{-3}$	$3.1236e{-2}$	$4.3709e{-3}$	$5.3080e0$	$1.4680e{-2}$
DIM 1 $q = p = 11$	$-3.1760e{-1}$	$-5.0292e{-1}$	$-1.3228e{-1}$	$3.6330e{-2}$	$8.6244e0$	$1.9200e{-3}$
DIM 1 $q = p = 13$	$-9.4151e{-2}$	$-1.7372e{-1}$	$-1.4584e{-2}$	$2.0149e{-2}$	$5.3008e0$	$1.8200e{-2}$
DIM 2 $q = p = 13$	$6.6123e{-2}$	$1.4182e{-2}$	$1.1807e{-1}$	$1.4526e{-2}$	$4.9253e0$	$1.8240e{-2}$
DIM 2 $q = p = 26$	$-4.6102e{-2}$	$-8.5844e{-2}$	$-6.3609e{-3}$	$7.3413e{-3}$	$7.7983e0$	$7.7200e{-3}$

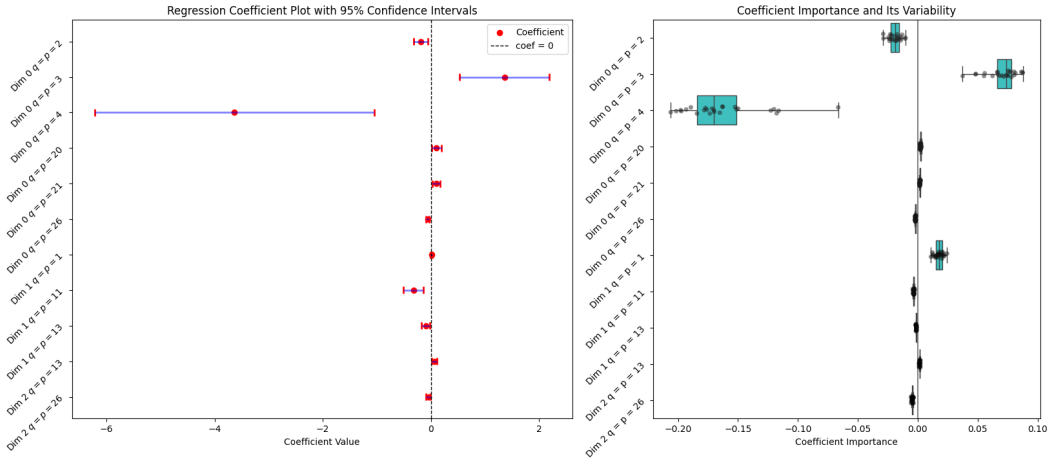

Figure 54: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: Whole Brain Mask and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 11: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: AG and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$7.7569e{-3}$	$5.2319e{-3}$	$1.0282e{-2}$	$8.9245e{-4}$	$8.0471e0$	$0.0000e0$
DIM 0 $q = p = 1$	$1.7418e{-3}$	$1.3126e{-3}$	$2.1710e{-3}$	$1.9325e{-4}$	$9.1925e0$	$0.0000e0$
DIM 0 $q = p = 2$	$-8.6721e{-2}$	$-1.1524e{-1}$	$-5.8203e{-2}$	$1.9219e{-2}$	$4.9603e0$	$0.0000e0$
DIM 0 $q = p = 19$	$1.7606e{-2}$	$2.5294e{-3}$	$3.2683e{-2}$	$4.8392e{-3}$	$5.4937e0$	$2.6400e{-3}$
DIM 1 $q = p = 1$	$6.3380e{-3}$	$2.4140e{-3}$	$1.0262e{-2}$	$1.4778e{-3}$	$4.2500e0$	$1.0000e{-2}$
DIM 1 $q = p = 2$	$-2.2954e{-1}$	$-4.2704e{-1}$	$-3.2041e{-2}$	$7.8173e{-2}$	$3.3021e0$	$3.3440e{-2}$
DIM 2 $q = p = 16$	$4.3482e{-3}$	$-1.7606e{-3}$	$1.0457e{-2}$	$1.1942e{-3}$	$6.2958e0$	$2.5680e{-2}$
DIM 2 $q = p = 17$	$-3.7191e{-3}$	$-9.1522e{-3}$	$1.7140e{-3}$	$1.2206e{-3}$	$4.9698e0$	$4.4200e{-2}$

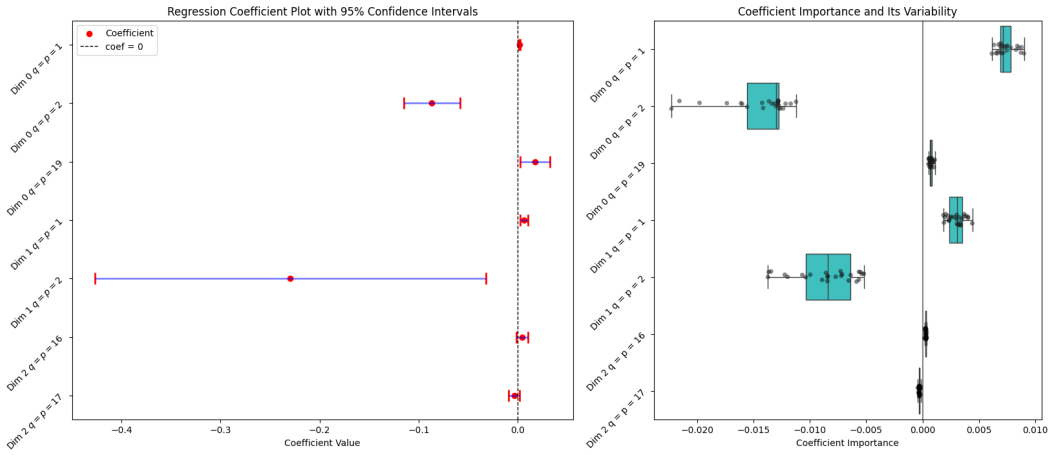

Figure 55: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: AG and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 12: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: AG and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$9.3740e{-3}$	$7.2599e{-3}$	$1.1488e{-2}$	$8.0076e{-4}$	$1.1223e1$	$0.0000e0$
DIM 0 $q = p = 1$	$1.8345e{-3}$	$1.4837e{-3}$	$2.1853e{-3}$	$1.6255e{-4}$	$1.1828e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-8.7200e{-2}$	$-1.1267e{-1}$	$-6.1728e{-2}$	$2.3745e{-2}$	$4.3938e0$	$0.0000e0$
DIM 0 $q = p = 3$	$1.2756e{-1}$	$-4.5414e{-2}$	$3.0054e{-1}$	$2.1260e{-1}$	$1.4209e0$	$2.6600e{-2}$
DIM 0 $q = p = 6$	$5.7342e0$	$2.7947e0$	$8.6736e0$	$1.3476e0$	$3.4930e0$	$4.8840e{-2}$
DIM 0 $q = p = 7$	$-4.8757e0$	$-7.6865e0$	$-2.0649e0$	$7.1900e{-1}$	$5.8813e0$	$1.2680e{-2}$
DIM 0 $q = p = 8$	$1.9270e0$	$6.1853e{-1}$	$3.2355e0$	$2.8687e{-1}$	$5.6867e0$	$2.5040e{-2}$
DIM 0 $q = p = 16$	$-1.8583e{-2}$	$-3.2215e{-2}$	$-4.9514e{-3}$	$3.7571e{-3}$	$4.5648e0$	$3.1000e{-2}$
DIM 0 $q = p = 18$	$-1.7607e{-2}$	$-2.8040e{-2}$	$-7.1743e{-3}$	$2.9363e{-3}$	$6.5962e0$	$1.5200e{-3}$
DIM 0 $q = p = 19$	$1.6683e{-2}$	$2.9630e{-3}$	$3.0403e{-2}$	$4.9227e{-3}$	$3.9165e0$	$2.1520e{-2}$
DIM 1 $q = p = 1$	$6.1574e{-3}$	$3.2007e{-3}$	$9.1141e{-3}$	$8.8558e{-4}$	$6.6509e0$	$5.6000e{-4}$
DIM 1 $q = p = 2$	$-3.3154e{-1}$	$-4.9844e{-1}$	$-1.6464e{-1}$	$6.6522e{-2}$	$4.6371e0$	$3.4400e{-3}$
DIM 1 $q = p = 10$	$6.3321e{-2}$	$1.9917e{-2}$	$1.0672e{-1}$	$1.2613e{-2}$	$4.7153e0$	$1.9000e{-2}$

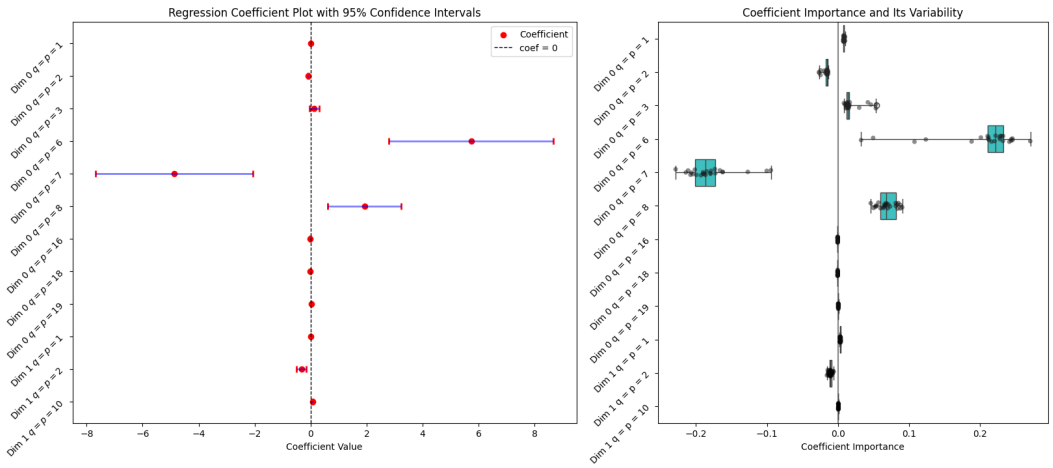

Figure 56: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: AG and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 13: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: AG and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$4.7500e{-3}$	$3.7984e{-3}$	$5.7016e{-3}$	$2.6542e{-4}$	$1.8713e1$	$0.0000e0$
DIM 0 $q = p = 1$	$1.0109e{-3}$	$5.9847e{-4}$	$1.4233e{-3}$	$1.8252e{-4}$	$5.4804e0$	$1.2000e{-4}$
DIM 0 $q = p = 2$	$-8.1555e{-2}$	$-1.0243e{-1}$	$-6.0683e{-2}$	$8.7707e{-3}$	$1.0170e1$	$0.0000e0$
DIM 0 $q = p = 3$	$2.5801e{-1}$	$1.8472e{-1}$	$3.3130e{-1}$	$2.8475e{-2}$	$1.0503e1$	$0.0000e0$
DIM 0 $q = p = 4$	$-1.8818e{-1}$	$-2.5087e{-1}$	$-1.2549e{-1}$	$2.1287e{-2}$	$1.0487e1$	$0.0000e0$
DIM 0 $q = p = \infty$	$2.0329e{-2}$	$7.9035e{-3}$	$3.2755e{-2}$	$2.7765e{-3}$	$8.4606e0$	$4.8000e{-4}$
DIM 1 $q = p = 1$	$1.5488e{-2}$	$1.1544e{-2}$	$1.9433e{-2}$	$1.1860e{-3}$	$1.2535e1$	$0.0000e0$
DIM 1 $q = p = 2$	$-4.2374e{-1}$	$-5.8951e{-1}$	$-2.5796e{-1}$	$4.9491e{-2}$	$7.6941e0$	$8.0000e{-5}$
DIM 1 $q = p = 3$	$1.0854e0$	$5.0985e{-1}$	$1.6610e0$	$1.7182e{-1}$	$5.3370e0$	$6.7200e{-3}$

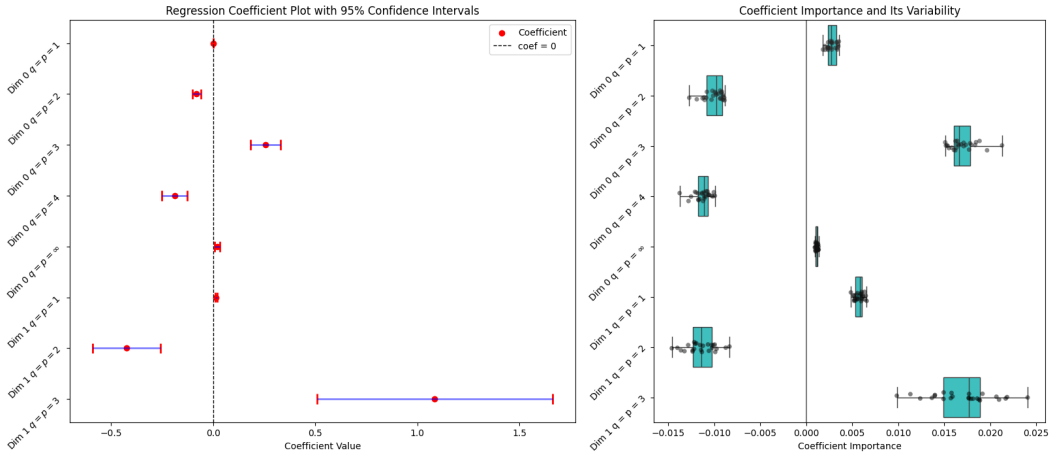

Figure 57: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: AG and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 14: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: AG and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\|t\|$	p
(INTERCEPT)	$-8.2385e{-4}$	$-3.4681e{-3}$	$1.8204e{-3}$	$6.6444e{-4}$	$8.8795e{-1}$	$6.7984e{-1}$
DIM 0 $q=p=1$	$1.4035e{-3}$	$1.0627e{-3}$	$1.7444e{-3}$	$9.3760e{-5}$	$1.6129e1$	$0.0000e0$
DIM 0 $q=p=2$	$-9.0451e{-2}$	$-1.1485e{-1}$	$-6.6048e{-2}$	$7.8371e{-3}$	$1.1291e1$	$0.0000e0$
DIM 0 $q=p=3$	$3.5668e{-1}$	$1.7908e{-1}$	$5.3428e{-1}$	$5.1019e{-2}$	$5.9654e0$	$2.7200e{-3}$
DIM 1 $q=p=1$	$6.2163e{-3}$	$2.9964e{-3}$	$9.4362e{-3}$	$1.0081e{-3}$	$5.4609e0$	$4.0800e{-3}$
DIM 1 $q=p=2$	$-2.3875e{-1}$	$-4.1399e{-1}$	$-6.3505e{-2}$	$5.5314e{-2}$	$3.9595e0$	$3.4360e{-2}$
DIM 1 $q=p=14$	$1.5117e{-2}$	$2.9759e{-3}$	$2.7258e{-2}$	$3.9668e{-3}$	$3.9243e0$	$3.4200e{-2}$
DIM 2 $q=p=10$	$4.0770e{-2}$	$1.5110e{-2}$	$6.6430e{-2}$	$6.6048e{-3}$	$5.1724e0$	$1.8000e{-2}$
DIM 2 $q=p=\infty$	$-7.4708e{-1}$	$-1.2996e0$	$-1.9459e{-1}$	$1.2706e{-1}$	$6.0437e0$	$1.2680e{-2}$

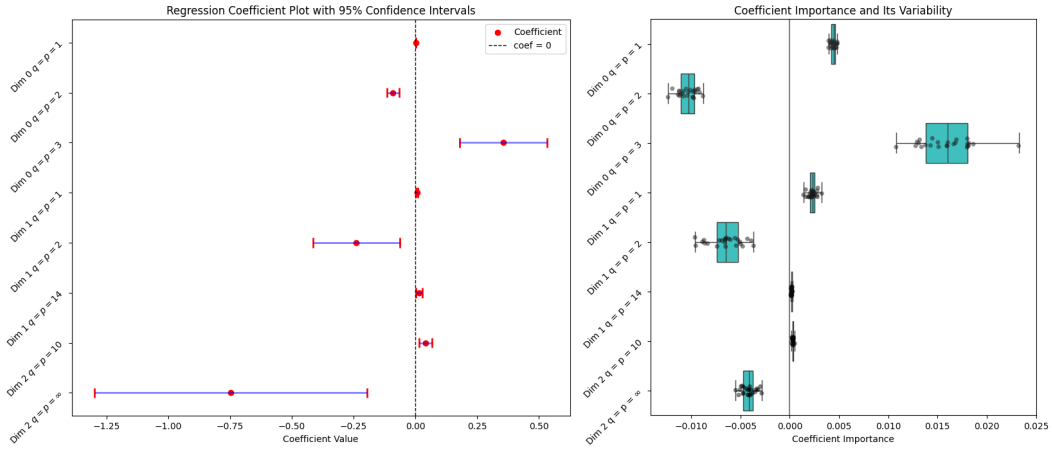

Figure 58: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: AG and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 15: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: ATL and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\|t\|$	p
(INTERCEPT)	$2.6347e{-3}$	$2.0549e{-3}$	$3.2144e{-3}$	$1.6393e{-4}$	$1.7366e1$	$0.0000e0$
DIM 0 $q=p=1$	$2.1014e{-4}$	$2.7715e{-5}$	$3.9257e{-4}$	$4.1501e{-5}$	$6.0691e0$	$1.3320e{-2}$
DIM 0 $q=p=2$	$-2.0689e{-2}$	$-2.9585e{-2}$	$-1.1793e{-2}$	$1.9951e{-3}$	$1.0714e1$	$0.0000e0$
DIM 0 $q=p=3$	$6.7637e{-2}$	$4.0634e{-2}$	$9.4640e{-2}$	$6.5712e{-3}$	$1.0587e1$	$0.0000e0$
DIM 0 $q=p=4$	$-4.1862e{-2}$	$-6.2291e{-2}$	$-2.1433e{-2}$	$5.6908e{-3}$	$7.5298e0$	$2.4000e{-4}$
DIM 0 $q=p=\infty$	$-7.4256e{-3}$	$-1.1949e{-2}$	$-2.9021e{-3}$	$1.4971e{-3}$	$5.3081e0$	$3.3600e{-3}$

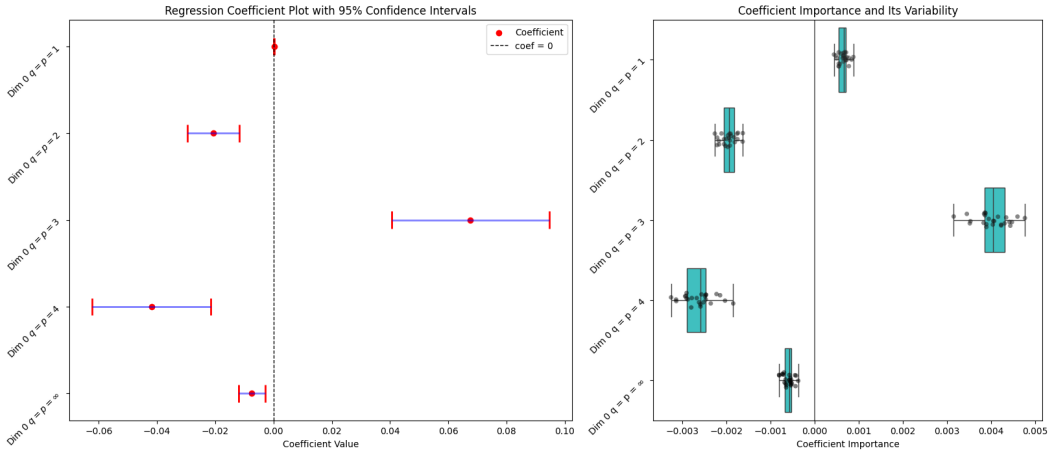

Figure 59: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: ATL and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 16: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: ATL and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$3.5153e{-3}$	$3.0754e{-3}$	$3.9552e{-3}$	$1.1921e{-4}$	$3.0574e1$	$0.0000e0$
DIM 0 $q = p = 1$	$4.1029e{-4}$	$2.7661e{-4}$	$5.4397e{-4}$	$3.5847e{-5}$	$1.1788e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-2.6809e{-2}$	$-3.3389e{-2}$	$-2.0228e{-2}$	$2.0654e{-3}$	$1.2977e1$	$0.0000e0$
DIM 0 $q = p = 3$	$7.7520e{-2}$	$5.7995e{-2}$	$9.7044e{-2}$	$5.5108e{-3}$	$1.3990e1$	$0.0000e0$
DIM 0 $q = p = 4$	$-4.7567e{-2}$	$-6.2149e{-2}$	$-3.2984e{-2}$	$3.6455e{-3}$	$1.2877e1$	$0.0000e0$
DIM 0 $q = p = \infty$	$-6.3964e{-3}$	$-9.5608e{-3}$	$-3.2319e{-3}$	$6.0558e{-4}$	$1.0680e1$	$1.2000e{-4}$

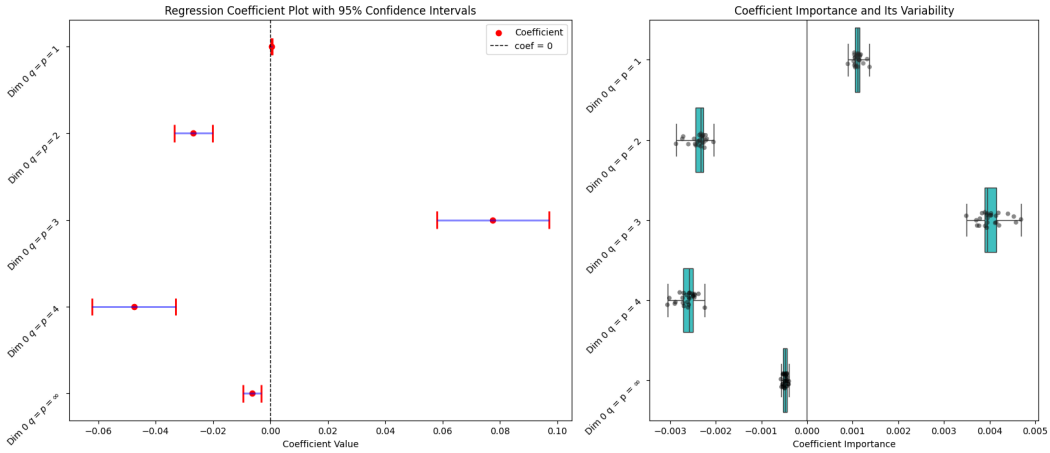

Figure 60: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: ATL and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 17: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: ATL and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$3.7035e{-3}$	$3.1671e{-3}$	$4.2399e{-3}$	$1.1697e{-4}$	$3.1591e1$	$0.0000e0$
DIM 0 $q = p = 1$	$4.8066e{-4}$	$2.8894e{-4}$	$6.7239e{-4}$	$5.8894e{-5}$	$7.6347e0$	$4.0000e{-5}$
DIM 0 $q = p = 2$	$-4.4705e{-2}$	$-5.9075e{-2}$	$-3.0335e{-2}$	$3.9365e{-3}$	$1.0551e1$	$0.0000e0$
DIM 0 $q = p = 3$	$2.5911e{-1}$	$1.6485e{-1}$	$3.5338e{-1}$	$2.1661e{-2}$	$1.1428e1$	$0.0000e0$
DIM 0 $q = p = 4$	$-3.8620e{-1}$	$-6.8443e{-1}$	$-8.7968e{-2}$	$6.1739e{-2}$	$6.7995e0$	$9.8400e{-3}$
DIM 1 $q = p = 2$	$-7.7284e{-2}$	$-1.5978e{-1}$	$5.2087e{-3}$	$1.9573e{-2}$	$4.6817e0$	$4.7920e{-2}$
DIM 2 $q = p = 8$	$-1.6718e{-1}$	$-2.4964e{-1}$	$-8.4720e{-2}$	$2.4815e{-2}$	$6.1501e0$	$1.4400e{-3}$

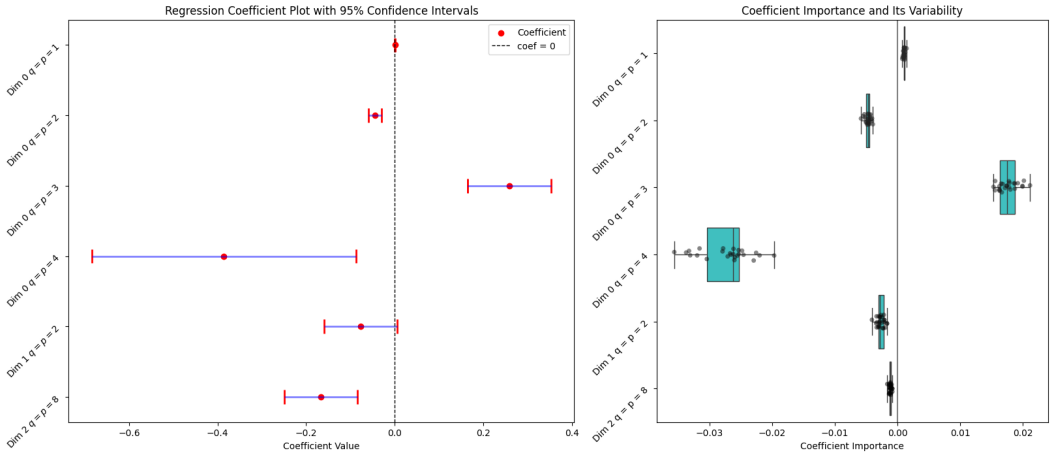

Figure 61: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: ATL and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 18: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: ATL and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$3.7193e{-3}$	$2.8529e{-3}$	$4.5856e{-3}$	$2.4718e{-4}$	$1.5594e1$	$0.0000e0$
DIM 0 $q = p = 1$	$5.5864e{-4}$	$4.0255e{-4}$	$7.1474e{-4}$	$2.8772e{-5}$	$1.9914e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-4.3459e{-2}$	$-5.5079e{-2}$	$-3.1838e{-2}$	$2.3736e{-3}$	$1.8322e1$	$0.0000e0$
DIM 0 $q = p = 3$	$2.2427e{-1}$	$1.5096e{-1}$	$2.9757e{-1}$	$1.8438e{-2}$	$1.2110e1$	$0.0000e0$
DIM 0 $q = p = 4$	$-3.2287e{-1}$	$-5.5225e{-1}$	$-9.3487e{-2}$	$5.8378e{-2}$	$6.0121e0$	$8.6400e{-3}$
DIM 0 $q = p = 13$	$-3.9876e{-2}$	$-6.4446e{-2}$	$-1.5307e{-2}$	$7.6197e{-3}$	$5.7064e0$	$2.6800e{-3}$
DIM 0 $q = p = 14$	$2.7933e{-2}$	$4.5715e{-3}$	$5.1295e{-2}$	$7.6033e{-3}$	$4.0033e0$	$3.4480e{-2}$
DIM 1 $q = p = 17$	$-3.4450e{-3}$	$-6.4170e{-3}$	$-4.7290e{-4}$	$6.1856e{-4}$	$6.1534e0$	$2.0560e{-2}$
DIM 2 $q = p = 8$	$-8.6516e{-2}$	$-1.5178e{-1}$	$-2.1253e{-2}$	$1.7083e{-2}$	$7.1545e0$	$1.2000e{-3}$

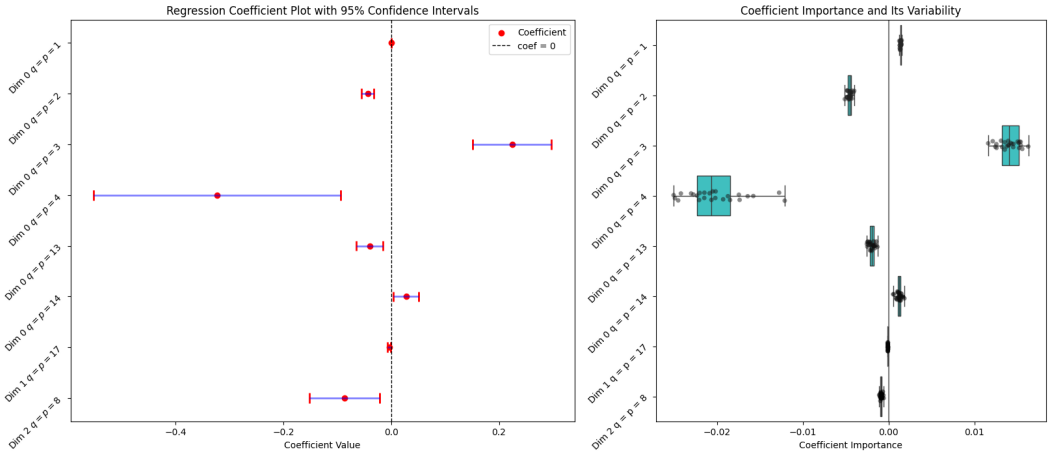

Figure 62: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: ATL and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 19: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: PTL and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$1.5585e{-2}$	$5.7167e{-3}$	$2.5453e{-2}$	$3.4389e{-3}$	$2.3243e0$	$1.8824e{-1}$
DIM 0 $q = p = 1$	$7.5348e{-4}$	$-2.5819e{-4}$	$1.7652e{-3}$	$3.2716e{-4}$	$3.9641e0$	$3.3800e{-2}$
DIM 0 $q = p = 13$	$2.6580e{-1}$	$3.7613e{-3}$	$5.2784e{-1}$	$5.8176e{-2}$	$5.0569e0$	$4.2080e{-2}$
DIM 0 $q = p = 18$	$7.6117e{-2}$	$9.9656e{-3}$	$1.4227e{-1}$	$1.3776e{-2}$	$5.1739e0$	$4.8600e{-2}$
DIM 0 $q = p = 23$	$3.5486e{-2}$	$1.0157e{-2}$	$6.0816e{-2}$	$6.3124e{-3}$	$5.5113e0$	$1.6000e{-2}$
DIM 1 $q = p = 9$	$-5.2947e{-1}$	$-8.6968e{-1}$	$-1.8927e{-1}$	$5.6995e{-2}$	$8.8984e0$	$4.9200e{-3}$
DIM 1 $q = p = 10$	$-1.2356e{-1}$	$-2.5348e{-1}$	$6.3622e{-3}$	$3.1852e{-2}$	$5.1894e0$	$2.4320e{-2}$
DIM 1 $q = p = 19$	$-3.1893e{-2}$	$-6.8747e{-2}$	$4.9612e{-3}$	$9.3464e{-3}$	$4.4091e0$	$4.6160e{-2}$
DIM 1 $q = p = 33$	$3.6272e{-2}$	$1.9693e{-2}$	$5.2852e{-2}$	$6.0548e{-3}$	$6.1350e0$	$5.2000e{-4}$
DIM 2 $q = p = 12$	$-2.5913e{-2}$	$-5.9565e{-2}$	$7.7385e{-3}$	$7.6498e{-3}$	$4.9852e0$	$4.1120e{-2}$
DIM 2 $q = p = 22$	$-2.9233e{-2}$	$-5.2112e{-2}$	$-6.3550e{-3}$	$5.8999e{-3}$	$5.5086e0$	$1.3320e{-2}$

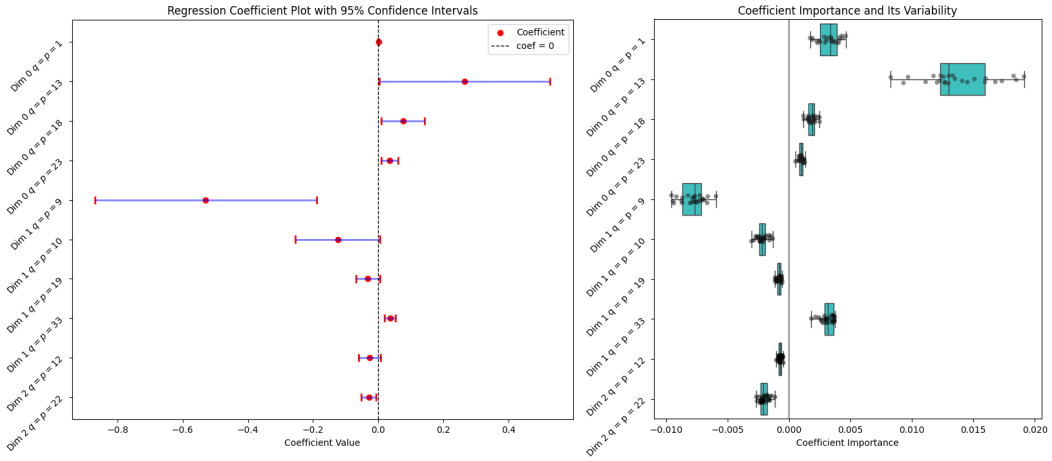

Figure 63: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: PTL and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 20: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: PTL and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$1.0268e{-2}$	$3.1473e{-3}$	$1.7389e{-2}$	$3.0379e{-3}$	$3.1600e0$	$3.5320e{-2}$
DIM 0 $q = p = 2$	$8.0419e{-2}$	$3.5483e{-2}$	$1.2535e{-1}$	$1.7892e{-2}$	$4.4886e0$	$3.6000e{-3}$
DIM 0 $q = p = 3$	$-3.5544e{-1}$	$-6.1130e{-1}$	$-9.9571e{-2}$	$9.5457e{-2}$	$3.6268e0$	$3.0440e{-2}$
DIM 0 $q = p = 9$	$-3.2986e0$	$-5.4383e0$	$-1.1588e0$	$8.9503e{-1}$	$3.2245e0$	$4.0360e{-2}$
DIM 0 $q = p = 18$	$7.0178e{-2}$	$2.0545e{-2}$	$1.1981e{-1}$	$1.2005e{-2}$	$5.4401e0$	$2.0560e{-2}$
DIM 0 $q = p = 23$	$2.3662e{-2}$	$7.3796e{-3}$	$3.9945e{-2}$	$3.6630e{-3}$	$5.5682e0$	$2.6040e{-2}$
DIM 0 $q = p = 32$	$-2.2148e{-2}$	$-3.6412e{-2}$	$-7.8849e{-3}$	$3.2943e{-3}$	$7.5597e0$	$1.6000e{-3}$
DIM 0 $q = p = 35$	$1.9227e{-2}$	$5.1714e{-4}$	$3.7937e{-2}$	$5.8812e{-3}$	$3.7248e0$	$4.5240e{-2}$
DIM 1 $q = p = 9$	$-4.6281e{-1}$	$-7.0142e{-1}$	$-2.2420e{-1}$	$6.0491e{-2}$	$7.0598e0$	$1.6400e{-3}$
DIM 1 $q = p = 10$	$-1.3839e{-1}$	$-2.2630e{-1}$	$-5.0483e{-2}$	$1.6225e{-2}$	$7.7840e0$	$7.5200e{-3}$
DIM 1 $q = p = 19$	$-3.0829e{-2}$	$-5.6030e{-2}$	$-5.6277e{-3}$	$7.2475e{-3}$	$4.3476e0$	$3.5880e{-2}$
DIM 1 $q = p = 35$	$3.5057e{-2}$	$1.8972e{-2}$	$5.1142e{-2}$	$6.1734e{-3}$	$4.1389e0$	$1.1440e{-2}$
DIM 2 $q = p = 6$	$-1.4861e1$	$-2.8733e1$	$-9.8855e{-1}$	$3.8387e0$	$4.5248e0$	$2.9120e{-2}$
DIM 2 $q = p = 18$	$2.3164e{-2}$	$7.8150e{-3}$	$3.8512e{-2}$	$3.5853e{-3}$	$6.0314e0$	$1.2400e{-2}$
DIM 2 $q = p = 19$	$-2.4316e{-2}$	$-3.8713e{-2}$	$-9.9189e{-3}$	$4.6657e{-3}$	$3.6749e0$	$4.6480e{-2}$
DIM 2 $q = p = 20$	$2.9925e{-2}$	$1.4701e{-2}$	$4.5150e{-2}$	$4.5436e{-3}$	$4.8935e0$	$1.1920e{-2}$
DIM 2 $q = p = 22$	$-2.2814e{-2}$	$-3.9557e{-2}$	$-6.0712e{-3}$	$3.8954e{-3}$	$5.9441e0$	$1.4160e{-2}$

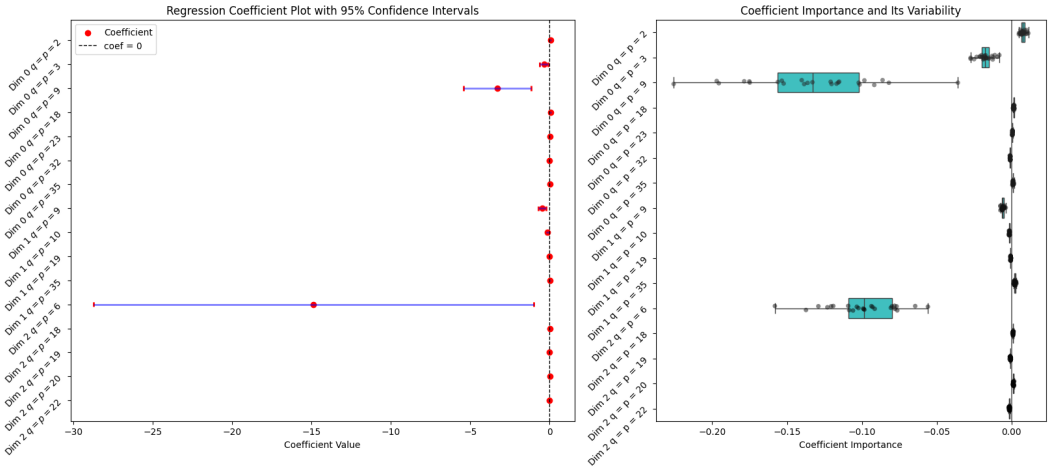

Figure 64: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: PTL and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 21: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: PTL and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$-5.2923e{-3}$	$-1.6630e{-2}$	$6.0458e{-3}$	$4.4800e{-3}$	$1.7909e0$	$2.5344e{-1}$
DIM 0 $q = p = 11$	$-5.8523e{-1}$	$-1.1147e0$	$-5.5797e{-2}$	$1.7857e{-1}$	$4.2272e0$	$1.6520e{-2}$
DIM 0 $q = p = 19$	$-1.1286e{-1}$	$-1.7756e{-1}$	$-4.8149e{-2}$	$1.4506e{-2}$	$6.2654e0$	$1.0680e{-2}$
DIM 0 $q = p = \infty$	$1.7962e{-1}$	$5.1970e{-2}$	$3.0728e{-1}$	$2.6219e{-2}$	$6.6191e0$	$1.3560e{-2}$
DIM 1 $q = p = 1$	$-1.7499e{-2}$	$-2.7279e{-2}$	$-7.7197e{-3}$	$3.3008e{-3}$	$5.3990e0$	$2.4400e{-3}$
DIM 1 $q = p = 2$	$7.4784e{-1}$	$2.5412e{-1}$	$1.2416e0$	$1.5426e{-1}$	$4.4264e0$	$2.0920e{-2}$
DIM 1 $q = p = 10$	$1.4712e{-1}$	$1.4032e{-2}$	$2.8020e{-1}$	$3.1971e{-2}$	$6.4546e0$	$6.2000e{-3}$
DIM 1 $q = p = 12$	$5.3914e{-2}$	$1.4478e{-2}$	$9.3351e{-2}$	$1.2628e{-2}$	$5.1767e0$	$5.6400e{-3}$
DIM 2 $q = p = 1$	$2.5956e{-1}$	$1.6803e{-1}$	$3.5110e{-1}$	$3.3323e{-2}$	$5.9471e0$	$3.6000e{-4}$
DIM 2 $q = p = 2$	$-9.3916e0$	$-1.3255e1$	$-5.5281e0$	$1.6880e0$	$3.9068e0$	$1.3040e{-2}$
DIM 2 $q = p = 14$	$-3.9924e{-2}$	$-6.8999e{-2}$	$-1.0850e{-2}$	$7.5473e{-3}$	$4.3309e0$	$4.9200e{-2}$
DIM 2 $q = p = 28$	$3.9050e{-2}$	$1.4120e{-2}$	$6.3979e{-2}$	$9.7509e{-3}$	$4.3271e0$	$1.5960e{-2}$

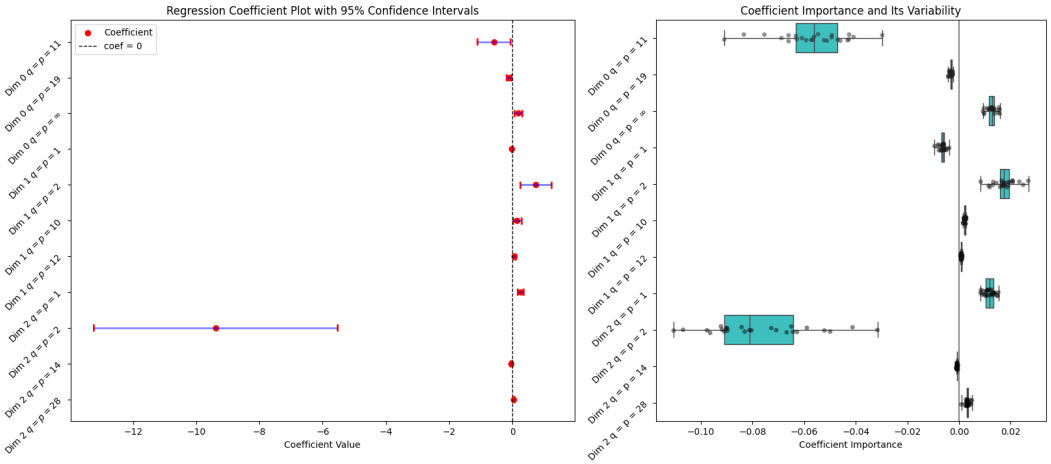

Figure 65: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: PTL and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 22: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: PTL and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$-2.0458e{-2}$	$-2.9290e{-2}$	$-1.1626e{-2}$	$4.6079e{-3}$	$1.3974e0$	$2.8840e{-1}$
DIM 0 $q = p = 19$	$-4.4577e{-2}$	$-8.7760e{-2}$	$-1.3936e{-3}$	$1.1679e{-2}$	$4.4383e0$	$3.8120e{-2}$
DIM 0 $q = p = 29$	$2.6851e{-2}$	$1.1790e{-2}$	$4.1912e{-2}$	$2.7699e{-3}$	$8.8938e0$	$2.2000e{-3}$
DIM 0 $q = p = \infty$	$1.5250e{-1}$	$6.4918e{-2}$	$2.4008e{-1}$	$1.8435e{-2}$	$8.5399e0$	$9.6000e{-4}$
DIM 1 $q = p = 1$	$-7.2707e{-3}$	$-1.4486e{-2}$	$-5.5603e{-5}$	$2.6363e{-3}$	$3.8352e0$	$1.9560e{-2}$
DIM 1 $q = p = 10$	$2.1890e{-1}$	$1.2080e{-1}$	$3.1700e{-1}$	$1.5871e{-2}$	$1.1342e1$	$4.4000e{-4}$
DIM 1 $q = p = 12$	$3.6330e{-2}$	$8.1226e{-3}$	$6.4537e{-2}$	$7.3995e{-3}$	$5.1413e0$	$1.8200e{-2}$
DIM 2 $q = p = 1$	$1.5831e{-1}$	$1.0277e{-1}$	$2.1384e{-1}$	$1.6827e{-2}$	$9.5096e0$	$0.0000e0$
DIM 2 $q = p = 2$	$-5.0261e0$	$-7.0774e0$	$-2.9747e0$	$6.3579e{-1}$	$7.5334e0$	$4.0000e{-5}$
DIM 2 $q = p = 14$	$-2.9303e{-2}$	$-5.1254e{-2}$	$-7.3525e{-3}$	$5.5759e{-3}$	$5.7891e0$	$7.7200e{-3}$
DIM 2 $q = p = 15$	$-2.9817e{-2}$	$-4.9779e{-2}$	$-9.8541e{-3}$	$4.9745e{-3}$	$5.6993e0$	$1.1400e{-2}$
DIM 2 $q = p = 16$	$-2.8393e{-2}$	$-4.6181e{-2}$	$-1.0604e{-2}$	$6.3687e{-3}$	$4.2164e0$	$1.1880e{-2}$
DIM 2 $q = p = 20$	$1.3214e{-2}$	$2.7608e{-3}$	$2.3667e{-2}$	$2.1354e{-3}$	$6.4423e0$	$1.5160e{-2}$
DIM 2 $q = p = 29$	$7.5011e{-2}$	$4.7473e{-2}$	$1.0255e{-1}$	$6.6481e{-3}$	$1.1495e1$	$0.0000e0$

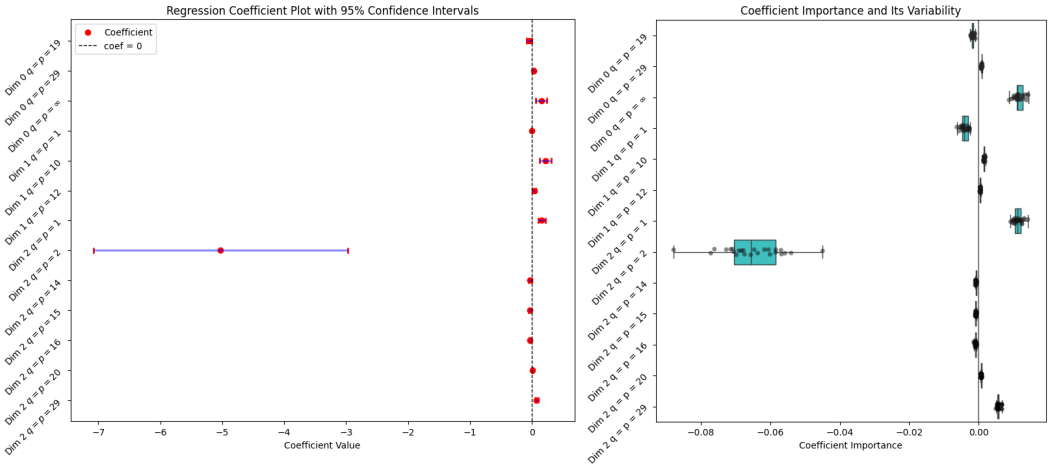

Figure 66: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: PTL and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 23: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: IFG and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$1.2152e{-2}$	$1.0494e{-2}$	$1.3810e{-2}$	$5.8618e{-4}$	$1.8943e1$	$0.0000e0$
DIM 0 $q = p = 1$	$3.4086e{-3}$	$2.8282e{-3}$	$3.9890e{-3}$	$1.8962e{-4}$	$1.8268e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-1.8577e{-1}$	$-2.1542e{-1}$	$-1.5612e{-1}$	$1.0606e{-2}$	$1.7234e1$	$0.0000e0$
DIM 0 $q = p = 3$	$5.1693e{-1}$	$4.2193e{-1}$	$6.1194e{-1}$	$3.3823e{-2}$	$1.4441e1$	$0.0000e0$
DIM 0 $q = p = 4$	$-3.4284e{-1}$	$-4.2229e{-1}$	$-2.6339e{-1}$	$2.6157e{-2}$	$1.1914e1$	$0.0000e0$
DIM 0 $q = p = \infty$	$3.6600e{-2}$	$1.9244e{-2}$	$5.3956e{-2}$	$4.9681e{-3}$	$6.0574e0$	$2.6400e{-3}$
DIM 1 $q = p = 1$	$8.0467e{-3}$	$2.9246e{-3}$	$1.3169e{-2}$	$1.4108e{-3}$	$6.2215e0$	$2.4800e{-3}$
DIM 1 $q = p = 2$	$-3.4178e{-1}$	$-5.4512e{-1}$	$-1.3844e{-1}$	$6.3094e{-2}$	$6.8980e0$	$1.2000e{-4}$
DIM 1 $q = p = 3$	$1.0227e0$	$3.1841e{-1}$	$1.7270e0$	$2.2583e{-1}$	$6.2329e0$	$6.8000e{-4}$
DIM 1 $q = p = 4$	$-7.0560e{-1}$	$-1.3062e0$	$-1.0494e{-1}$	$1.7930e{-1}$	$5.4286e0$	$5.9200e{-3}$

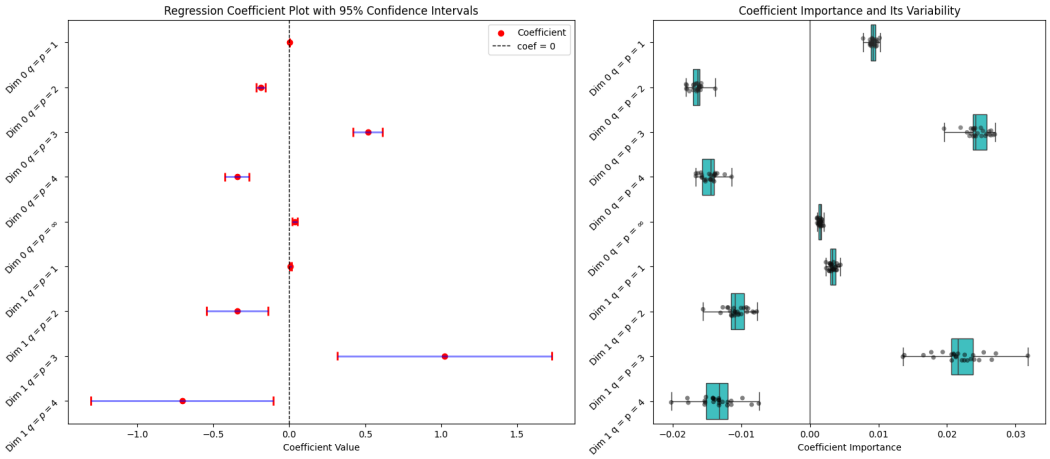

Figure 67: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: IFG and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 24: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: IFG and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$1.4132e{-2}$	$1.2836e{-2}$	$1.5428e{-2}$	$3.5541e{-4}$	$4.0213e1$	$0.0000e0$
DIM 0 $q = p = 1$	$3.5718e{-3}$	$3.1534e{-3}$	$3.9901e{-3}$	$1.3535e{-4}$	$2.7680e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-1.8855e{-1}$	$-2.1799e{-1}$	$-1.5912e{-1}$	$1.0323e{-2}$	$2.0283e1$	$0.0000e0$
DIM 0 $q = p = 3$	$5.9426e{-1}$	$4.3111e{-1}$	$7.5740e{-1}$	$4.8883e{-2}$	$1.3972e1$	$0.0000e0$
DIM 0 $q = p = 4$	$-6.2831e{-1}$	$-9.1999e{-1}$	$-3.3663e{-1}$	$8.0982e{-2}$	$9.2786e0$	$0.0000e0$
DIM 0 $q = p = 5$	$2.3541e{-1}$	$6.2997e{-2}$	$4.0782e{-1}$	$4.6127e{-2}$	$6.3883e0$	$2.5200e{-3}$
DIM 1 $q = p = 1$	$9.1191e{-3}$	$5.2884e{-3}$	$1.2950e{-2}$	$9.6548e{-4}$	$8.2422e0$	$1.2000e{-4}$
DIM 1 $q = p = 2$	$-4.5065e{-1}$	$-6.4678e{-1}$	$-2.5452e{-1}$	$4.8220e{-2}$	$8.0120e0$	$4.0000e{-4}$

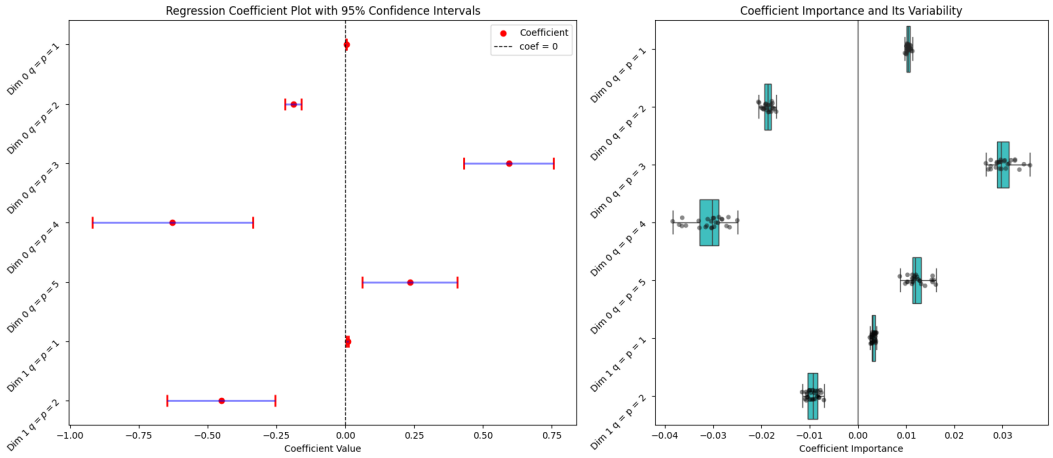

Figure 68: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: IFG and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 25: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: IFG and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$7.6982e{-3}$	$5.2758e{-4}$	$1.4869e{-2}$	$2.0497e{-3}$	$3.9723e0$	$4.7920e{-2}$
DIM 0 $q = p = 1$	$2.4968e{-3}$	$1.7600e{-3}$	$3.2336e{-3}$	$2.4391e{-4}$	$1.0014e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-1.3040e{-1}$	$-1.7178e{-1}$	$-8.9025e{-2}$	$1.9467e{-2}$	$5.5568e0$	$0.0000e0$
DIM 0 $q = p = 11$	$-1.6410e{-1}$	$-3.0071e{-1}$	$-2.7502e{-2}$	$2.7423e{-2}$	$5.6992e0$	$3.6080e{-2}$
DIM 1 $q = p = 1$	$1.1696e{-2}$	$5.6391e{-3}$	$1.7753e{-2}$	$2.2258e{-3}$	$5.5314e0$	$6.4000e{-4}$
DIM 1 $q = p = 2$	$-3.5225e{-1}$	$-6.5589e{-1}$	$-4.8604e{-2}$	$1.1324e{-1}$	$3.7184e0$	$2.0640e{-2}$
DIM 1 $q = p = 3$	$1.7138e0$	$2.2874e{-1}$	$3.1988e0$	$5.6054e{-1}$	$3.7436e0$	$1.5360e{-2}$
DIM 1 $q = p = 4$	$-4.9525e0$	$-8.6806e0$	$-1.2243e0$	$1.2788e0$	$4.1077e0$	$1.6600e{-2}$
DIM 1 $q = p = 11$	$-5.7324e{-2}$	$-9.7330e{-2}$	$-1.7318e{-2}$	$7.0890e{-3}$	$6.9779e0$	$2.2200e{-2}$
DIM 1 $q = p = 17$	$2.2443e{-2}$	$2.4178e{-4}$	$4.4645e{-2}$	$5.8184e{-3}$	$5.1042e0$	$2.1840e{-2}$
DIM 2 $q = p = 19$	$-1.5601e{-2}$	$-2.8013e{-2}$	$-3.1898e{-3}$	$3.3055e{-3}$	$5.6017e0$	$8.7600e{-3}$

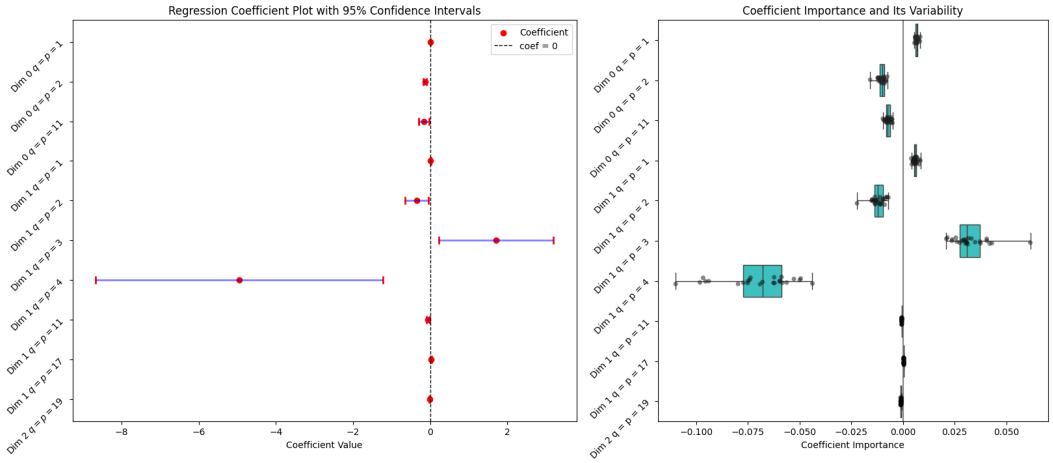

Figure 69: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: IFG and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 26: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: IFG and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$1.0846e{-2}$	$5.5497e{-3}$	$1.6142e{-2}$	$1.9059e{-3}$	$3.9423e0$	$2.5320e{-2}$
DIM 0 $q = p = 1$	$2.9811e{-3}$	$2.4451e{-3}$	$3.5171e{-3}$	$1.8159e{-4}$	$1.6916e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-1.4613e{-1}$	$-1.8164e{-1}$	$-1.1062e{-1}$	$1.8017e{-2}$	$8.5486e0$	$0.0000e0$
DIM 0 $q = p = 3$	$3.9672e{-1}$	$1.5808e{-1}$	$6.3536e{-1}$	$1.4258e{-1}$	$2.8312e0$	$3.8480e{-2}$
DIM 1 $q = p = 1$	$1.5140e{-2}$	$1.0645e{-2}$	$1.9635e{-2}$	$1.1797e{-3}$	$1.0358e1$	$0.0000e0$
DIM 1 $q = p = 2$	$-5.5080e{-1}$	$-7.9807e{-1}$	$-3.0353e{-1}$	$7.3269e{-2}$	$5.9266e0$	$3.4400e{-3}$
DIM 1 $q = p = 3$	$1.9133e0$	$6.3249e{-1}$	$3.1941e0$	$3.7926e{-1}$	$4.7817e0$	$1.8600e{-2}$
DIM 1 $q = p = 9$	$4.5235e{-1}$	$1.7541e{-1}$	$7.2928e{-1}$	$3.8963e{-2}$	$1.0325e1$	$6.4400e{-3}$
DIM 1 $q = p = 11$	$-3.2621e{-2}$	$-6.3679e{-2}$	$-1.5635e{-3}$	$8.0543e{-3}$	$5.2390e0$	$1.8800e{-2}$
DIM 1 $q = p = 13$	$-1.5051e{-2}$	$-3.1761e{-2}$	$1.6596e{-3}$	$4.7727e{-3}$	$4.0961e0$	$4.6600e{-2}$
DIM 1 $q = p = 19$	$1.4403e{-2}$	$2.4424e{-3}$	$2.6364e{-2}$	$2.9250e{-3}$	$5.4581e0$	$1.8040e{-2}$
DIM 2 $q = p = 19$	$-2.5808e{-2}$	$-3.7367e{-2}$	$-1.4250e{-2}$	$3.2017e{-3}$	$6.4056e0$	$2.3200e{-3}$

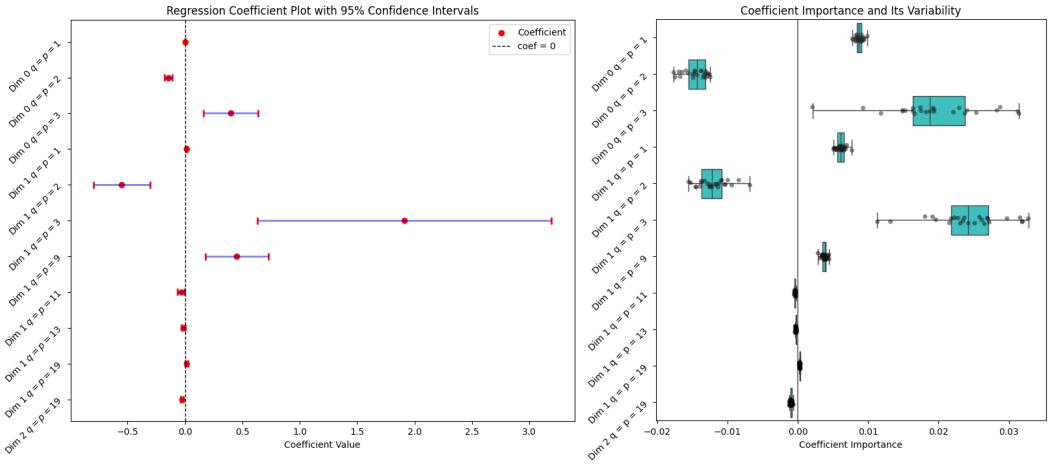

Figure 70: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: IFG and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 27: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: MFG and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$1.3996e{-3}$	$-1.7717e{-5}$	$2.8168e{-3}$	$3.2696e{-4}$	$4.8814e0$	$4.2120e{-2}$
DIM 0 $q = p = \infty$	$-2.6933e{-2}$	$-4.0844e{-2}$	$-1.3022e{-2}$	$3.2240e{-3}$	$6.7520e0$	$4.6800e{-3}$
DIM 1 $q = p = 1$	$1.0460e{-2}$	$6.3590e{-3}$	$1.4562e{-2}$	$1.1523e{-3}$	$8.4706e0$	$4.0000e{-5}$
DIM 1 $q = p = 2$	$-3.5318e{-1}$	$-5.5446e{-1}$	$-1.5191e{-1}$	$6.1924e{-2}$	$5.2441e0$	$9.3600e{-3}$
DIM 2 $q = p = 1$	$8.2001e{-2}$	$5.0725e{-2}$	$1.1328e{-1}$	$1.0095e{-2}$	$8.2064e0$	$0.0000e0$
DIM 2 $q = p = 2$	$-2.2388e0$	$-3.1626e0$	$-1.3149e0$	$3.4067e{-1}$	$6.9419e0$	$4.0000e{-5}$
DIM 2 $q = p = 3$	$5.7072e0$	$2.5076e0$	$8.9068e0$	$1.1902e0$	$5.1960e0$	$3.1600e{-3}$
DIM 2 $q = p = 4$	$-3.9269e0$	$-6.6403e0$	$-1.2135e0$	$9.9198e{-1}$	$4.2785e0$	$1.7200e{-2}$

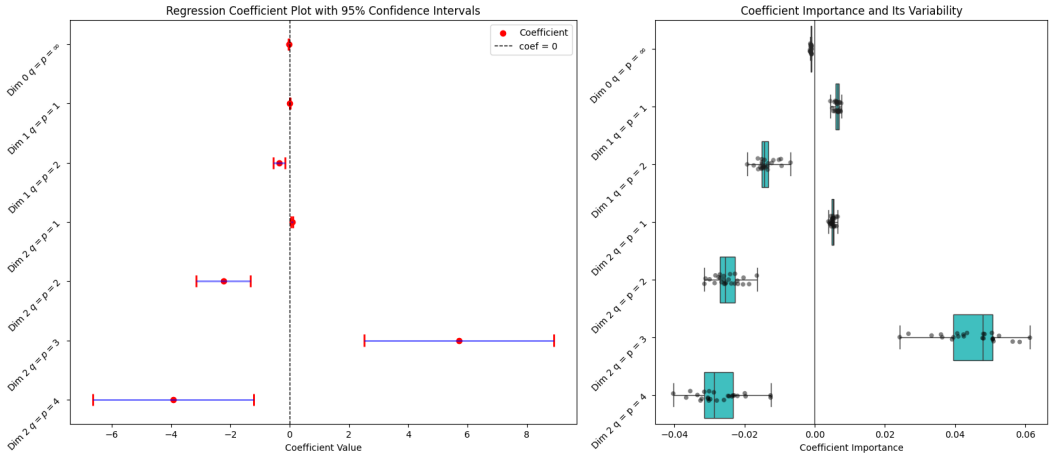

Figure 71: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: MFG and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 28: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: MFG and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$-4.7110e{-3}$	$-8.5453e{-3}$	$-8.7671e{-4}$	$1.2991e{-3}$	$3.4611e0$	$5.7360e{-2}$
DIM 0 $q = p = 1$	$5.7421e{-4}$	$1.6202e{-4}$	$9.8641e{-4}$	$1.2700e{-4}$	$5.2930e0$	$9.2800e{-3}$
DIM 0 $q = p = 25$	$-2.9137e{-2}$	$-5.2704e{-2}$	$-5.5700e{-3}$	$8.2875e{-3}$	$3.6270e0$	$3.2360e{-2}$
DIM 0 $q = p = 27$	$3.1774e{-2}$	$2.9214e{-3}$	$6.0626e{-2}$	$6.5506e{-3}$	$4.9152e0$	$4.5000e{-2}$
DIM 1 $q = p = 1$	$6.1024e{-3}$	$2.7949e{-3}$	$9.4098e{-3}$	$8.7892e{-4}$	$7.1884e0$	$8.4000e{-4}$
DIM 1 $q = p = 2$	$-2.7618e{-1}$	$-4.5964e{-1}$	$-9.2723e{-2}$	$5.1548e{-2}$	$5.9465e0$	$3.5200e{-3}$
DIM 1 $q = p = 9$	$3.1466e{-1}$	$5.9412e{-2}$	$5.6991e{-1}$	$5.2077e{-2}$	$5.3063e0$	$4.9400e{-2}$
DIM 1 $q = p = 14$	$-2.4714e{-2}$	$-3.6881e{-2}$	$-1.2547e{-2}$	$4.0324e{-3}$	$4.2467e0$	$1.9320e{-2}$
DIM 1 $q = p = 27$	$2.8023e{-2}$	$4.7186e{-3}$	$5.1327e{-2}$	$8.8894e{-3}$	$3.2704e0$	$4.7680e{-2}$
DIM 2 $q = p = 8$	$-7.9962e{-1}$	$-1.3330e0$	$-2.6619e{-1}$	$9.0912e{-2}$	$7.4300e0$	$1.8280e{-2}$
DIM 2 $q = p = 16$	$6.8196e{-3}$	$7.1987e{-4}$	$1.2919e{-2}$	$1.6549e{-3}$	$4.3284e0$	$4.2920e{-2}$

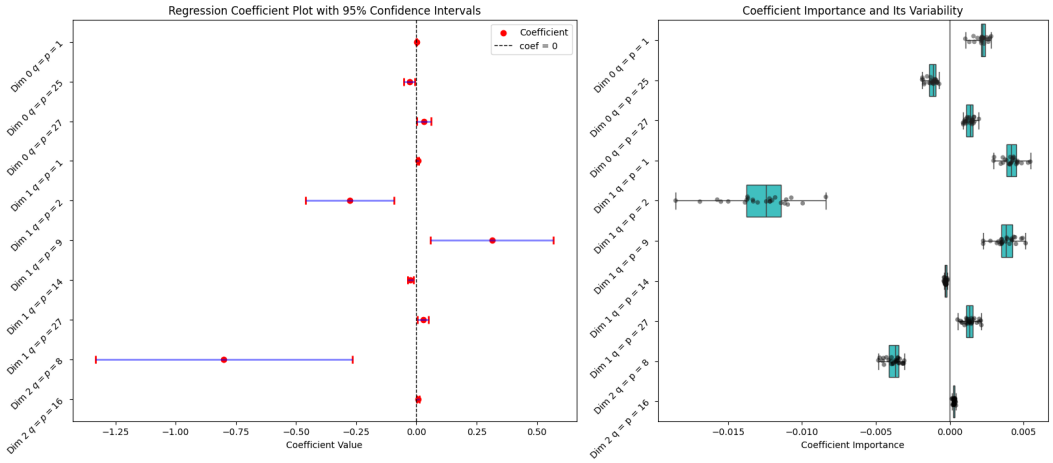

Figure 72: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: MFG and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 29: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: MFG and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$-5.1684e{-3}$	$-1.0312e{-2}$	$-2.4684e{-5}$	$9.4711e{-4}$	$4.6363e0$	$1.1100e{-1}$
DIM 0 $q = p = 1$	$5.2862e{-4}$	$-8.4094e{-5}$	$1.1413e{-3}$	$1.8096e{-4}$	$4.5573e0$	$2.1240e{-2}$
DIM 1 $q = p = 1$	$9.6453e{-3}$	$4.9470e{-3}$	$1.4344e{-2}$	$1.1250e{-3}$	$7.1302e0$	$2.4000e{-3}$

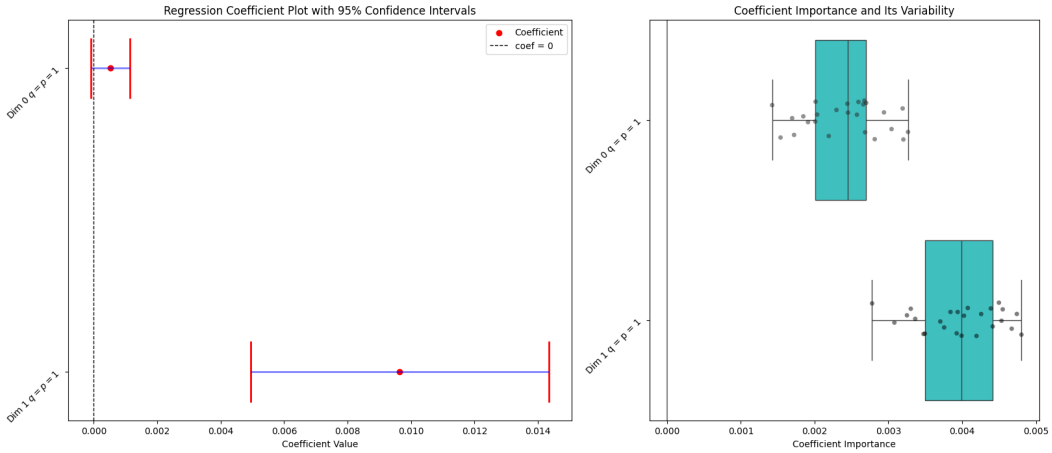

Figure 73: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: MFG and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 30: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: MFG and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$-6.1642e{-3}$	$-9.9033e{-3}$	$-2.4250e{-3}$	$1.1508e{-3}$	$4.8619e0$	$9.4400e{-3}$
DIM 0 $q = p = 1$	$1.5097e{-3}$	$1.0501e{-3}$	$1.9693e{-3}$	$1.1655e{-4}$	$1.0806e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-1.2065e{-1}$	$-1.5796e{-1}$	$-8.3346e{-2}$	$1.2607e{-2}$	$7.2642e0$	$0.0000e0$
DIM 0 $q = p = 3$	$6.5482e{-1}$	$4.1103e{-1}$	$8.9862e{-1}$	$8.2447e{-2}$	$5.8368e0$	$5.2000e{-4}$
DIM 0 $q = p = 4$	$-1.3070e0$	$-2.0279e0$	$-5.8613e{-1}$	$2.3516e{-1}$	$3.7931e0$	$3.6440e{-2}$
DIM 0 $q = p = 16$	$3.6950e{-2}$	$1.0030e{-2}$	$6.3870e{-2}$	$6.5351e{-3}$	$5.1724e0$	$2.7240e{-2}$
DIM 1 $q = p = 1$	$6.5595e{-3}$	$3.0950e{-3}$	$1.0024e{-2}$	$9.5399e{-4}$	$7.6541e0$	$2.8000e{-4}$
DIM 1 $q = p = 2$	$-1.8015e{-1}$	$-3.6976e{-1}$	$9.4571e{-3}$	$6.1966e{-2}$	$3.6304e0$	$4.8040e{-2}$
DIM 1 $q = p = 17$	$1.6031e{-2}$	$1.5339e{-3}$	$3.0529e{-2}$	$3.4386e{-3}$	$5.0093e0$	$3.4280e{-2}$
DIM 2 $q = p = \infty$	$-5.8378e{-1}$	$-1.3998e0$	$2.3229e{-1}$	$2.0274e{-1}$	$4.5419e0$	$4.4680e{-2}$

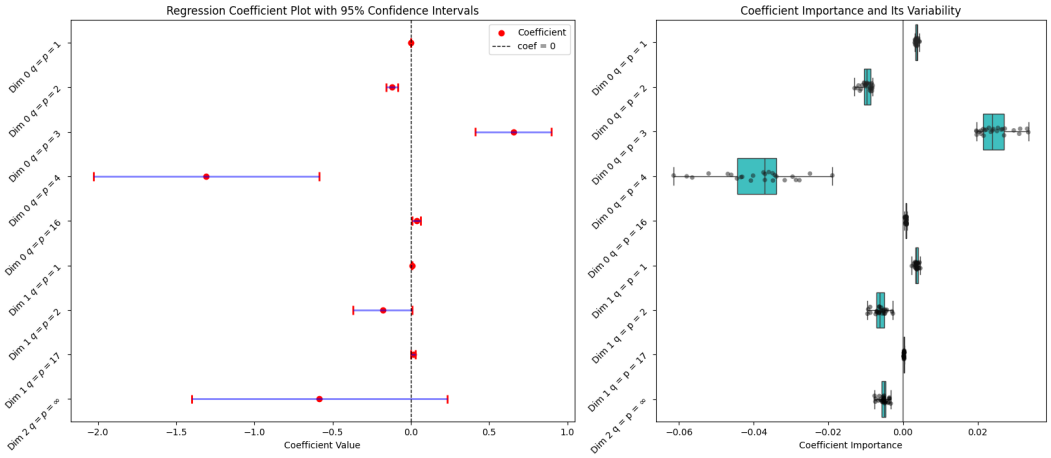

Figure 74: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: MFG and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 31: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: IFGorb and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$4.0872e{-3}$	$3.4022e{-3}$	$4.7722e{-3}$	$2.6092e{-4}$	$1.4420e1$	$0.0000e0$
DIM 0 $q = p = 1$	$1.3223e{-3}$	$1.1003e{-3}$	$1.5443e{-3}$	$1.0719e{-4}$	$1.2588e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-8.1721e{-2}$	$-9.4914e{-2}$	$-6.8527e{-2}$	$6.6871e{-3}$	$1.1971e1$	$0.0000e0$
DIM 0 $q = p = 3$	$2.4845e{-1}$	$2.0322e{-1}$	$2.9368e{-1}$	$2.2562e{-2}$	$1.0374e1$	$0.0000e0$
DIM 0 $q = p = 4$	$-1.7140e{-1}$	$-2.0797e{-1}$	$-1.3484e{-1}$	$1.7459e{-2}$	$8.8373e0$	$0.0000e0$
DIM 0 $q = p = \infty$	$1.2560e{-2}$	$6.2300e{-3}$	$1.8889e{-2}$	$2.0932e{-3}$	$3.9695e0$	$2.6040e{-2}$
DIM 1 $q = p = 1$	$5.6828e{-3}$	$3.7792e{-3}$	$7.5864e{-3}$	$6.4386e{-4}$	$8.0676e0$	$0.0000e0$
DIM 1 $q = p = 2$	$-2.3237e{-1}$	$-3.1542e{-1}$	$-1.4931e{-1}$	$2.6543e{-2}$	$8.2521e0$	$0.0000e0$
DIM 1 $q = p = 3$	$5.8867e{-1}$	$3.0317e{-1}$	$8.7416e{-1}$	$9.0539e{-2}$	$6.3086e0$	$4.0000e{-4}$
DIM 1 $q = p = 4$	$-3.2586e{-1}$	$-5.6097e{-1}$	$-9.0763e{-2}$	$7.0794e{-2}$	$4.6249e0$	$1.5760e{-2}$

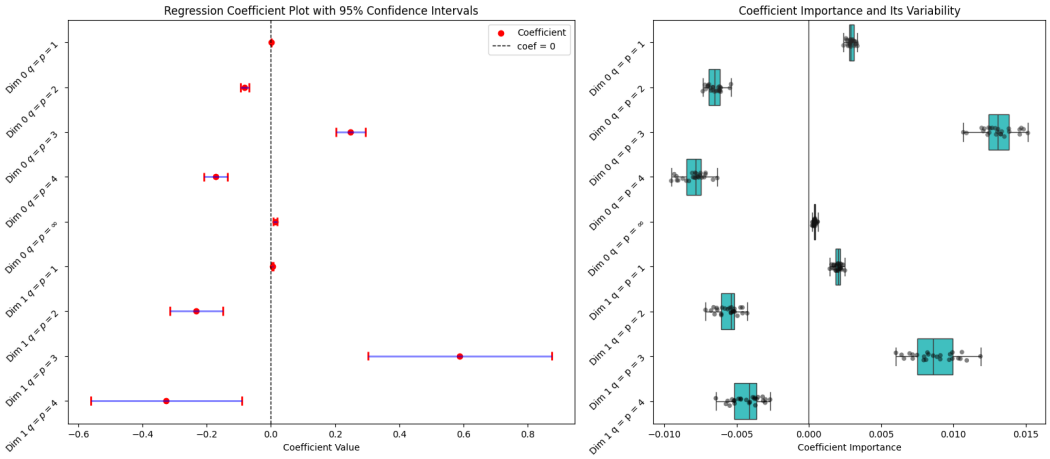

Figure 75: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: IFGorb and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 32: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: IFGorb and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$4.8182e{-3}$	$4.2913e{-3}$	$5.3451e{-3}$	$2.1064e{-4}$	$2.2569e1$	$0.0000e0$
DIM 0 $q = p = 1$	$1.5360e{-3}$	$1.3878e{-3}$	$1.6842e{-3}$	$4.0320e{-5}$	$3.6555e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-9.3796e{-2}$	$-1.0299e{-1}$	$-8.4602e{-2}$	$3.5408e{-3}$	$2.6042e1$	$0.0000e0$
DIM 0 $q = p = 3$	$2.7084e{-1}$	$2.3770e{-1}$	$3.0397e{-1}$	$1.5124e{-2}$	$1.8135e1$	$0.0000e0$
DIM 0 $q = p = 4$	$-1.7918e{-1}$	$-2.0667e{-1}$	$-1.5169e{-1}$	$1.3121e{-2}$	$1.4126e1$	$0.0000e0$
DIM 0 $q = p = \infty$	$1.0916e{-2}$	$6.3192e{-3}$	$1.5514e{-2}$	$1.5222e{-3}$	$7.7552e0$	$0.0000e0$
DIM 1 $q = p = 1$	$4.7381e{-3}$	$3.3740e{-3}$	$6.1021e{-3}$	$4.7664e{-4}$	$1.1084e1$	$0.0000e0$
DIM 1 $q = p = 2$	$-2.1427e{-1}$	$-2.7891e{-1}$	$-1.4964e{-1}$	$2.5617e{-2}$	$8.5284e0$	$0.0000e0$
DIM 1 $q = p = 3$	$6.1890e{-1}$	$3.8935e{-1}$	$8.4846e{-1}$	$9.3479e{-2}$	$6.2885e0$	$4.0000e{-5}$
DIM 1 $q = p = 4$	$-4.0466e{-1}$	$-5.9288e{-1}$	$-2.1644e{-1}$	$7.1281e{-2}$	$5.1461e0$	$1.6800e{-3}$
DIM 2 $q = p = \infty$	$-2.3388e{-1}$	$-3.7978e{-1}$	$-8.7981e{-2}$	$3.8588e{-2}$	$5.6401e0$	$9.5600e{-3}$

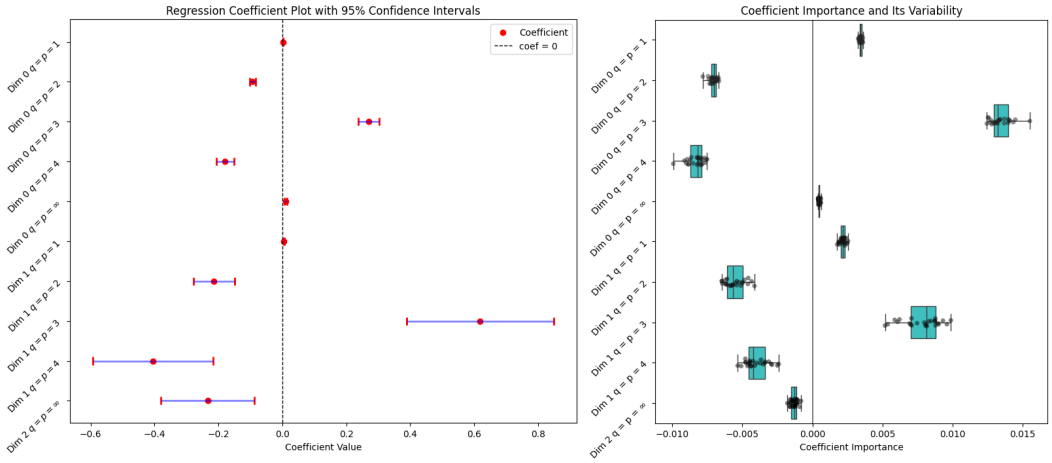

Figure 76: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: IFGorb and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 33: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: IFGorb and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\|t\|$	p
(INTERCEPT)	$4.1390e{-3}$	$3.4394e{-3}$	$4.8387e{-3}$	$2.2546e{-4}$	$1.7859e1$	$0.0000e0$
DIM 0 $q = p = 1$	$1.4564e{-3}$	$1.2480e{-3}$	$1.6647e{-3}$	$7.8326e{-5}$	$1.8145e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-7.5929e{-2}$	$-8.8622e{-2}$	$-6.3237e{-2}$	$5.2309e{-3}$	$1.3765e1$	$0.0000e0$
DIM 0 $q = p = 3$	$1.9063e{-1}$	$1.4891e{-1}$	$2.3236e{-1}$	$1.6993e{-2}$	$1.0269e1$	$0.0000e0$
DIM 0 $q = p = 4$	$-1.0865e{-1}$	$-1.4263e{-1}$	$-7.4677e{-2}$	$1.3197e{-2}$	$7.0897e0$	$0.0000e0$
DIM 1 $q = p = 1$	$2.2672e{-3}$	$2.7942e{-4}$	$4.2551e{-3}$	$5.0943e{-4}$	$4.4725e0$	$4.3880e{-2}$
DIM 2 $q = p = 1$	$2.5250e{-2}$	$1.0881e{-2}$	$3.9619e{-2}$	$4.4580e{-3}$	$5.7732e0$	$3.4800e{-3}$
DIM 2 $q = p = 2$	$-5.6038e{-1}$	$-9.4559e{-1}$	$-1.7516e{-1}$	$1.2595e{-1}$	$4.4972e0$	$1.7400e{-2}$

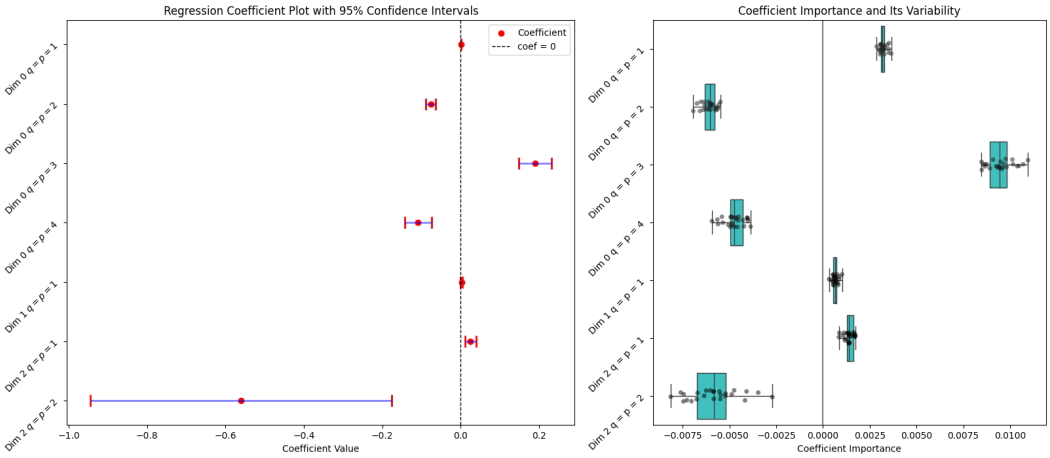

Figure 77: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: IFGorb and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 34: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: IFGorb and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$4.8847e{-3}$	$4.3702e{-3}$	$5.3992e{-3}$	$1.4309e{-4}$	$3.3710e1$	$0.0000e0$
DIM 0 $q = p = 1$	$1.4334e{-3}$	$1.2884e{-3}$	$1.5784e{-3}$	$5.8138e{-5}$	$2.4453e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-7.1814e{-2}$	$-8.1013e{-2}$	$-6.2615e{-2}$	$4.0488e{-3}$	$1.8175e1$	$0.0000e0$
DIM 0 $q = p = 3$	$1.7307e{-1}$	$1.4182e{-1}$	$2.0432e{-1}$	$1.3839e{-2}$	$1.2966e1$	$0.0000e0$
DIM 0 $q = p = 4$	$-9.4650e{-2}$	$-1.2056e{-1}$	$-6.8738e{-2}$	$1.0656e{-2}$	$9.2866e0$	$0.0000e0$
DIM 1 $q = p = 1$	$2.7981e{-3}$	$1.3820e{-3}$	$4.2143e{-3}$	$4.5536e{-4}$	$6.4843e0$	$4.8000e{-4}$
DIM 1 $q = p = 2$	$-1.5449e{-1}$	$-2.2333e{-1}$	$-8.5637e{-2}$	$2.7692e{-2}$	$4.8937e0$	$2.4000e{-3}$
DIM 1 $q = p = 3$	$4.6717e{-1}$	$2.2176e{-1}$	$7.1258e{-1}$	$1.0415e{-1}$	$3.6902e0$	$2.5640e{-2}$

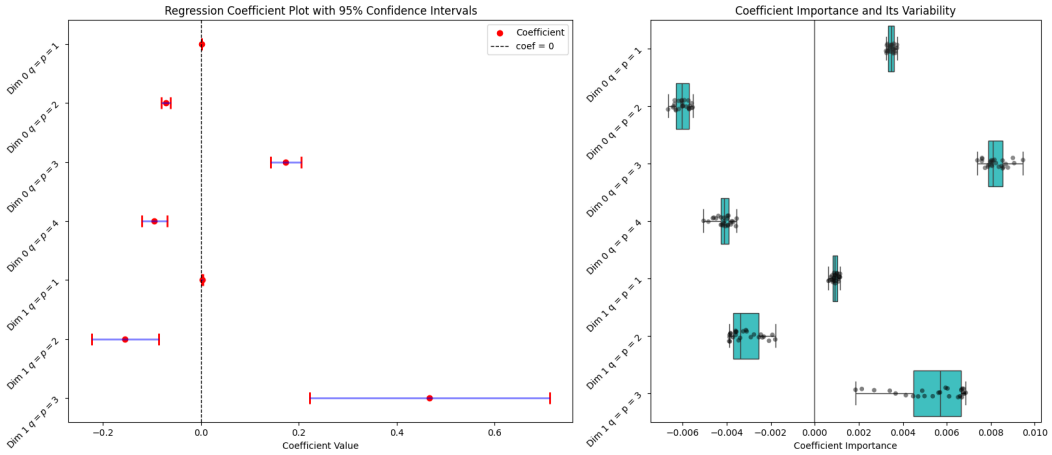

Figure 78: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: IFGorb and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 35: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: PCC and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$2.8611e{-3}$	$2.1406e{-3}$	$3.5817e{-3}$	$2.1186e{-4}$	$1.4518e1$	$0.0000e0$
DIM 0 $q = p = 1$	$9.0675e{-4}$	$7.0720e{-4}$	$1.1063e{-3}$	$6.4641e{-5}$	$1.2677e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-6.5188e{-2}$	$-7.7791e{-2}$	$-5.2586e{-2}$	$4.8102e{-3}$	$1.2467e1$	$0.0000e0$
DIM 0 $q = p = 3$	$1.9371e{-1}$	$1.4809e{-1}$	$2.3934e{-1}$	$1.6057e{-2}$	$1.1113e1$	$0.0000e0$
DIM 0 $q = p = 4$	$-1.2030e{-1}$	$-1.5858e{-1}$	$-8.2019e{-2}$	$1.2365e{-2}$	$9.2920e0$	$0.0000e0$
DIM 1 $q = p = 1$	$4.7096e{-3}$	$2.7731e{-3}$	$6.6460e{-3}$	$5.4751e{-4}$	$9.0836e0$	$0.0000e0$
DIM 1 $q = p = 2$	$-1.5204e{-1}$	$-2.2845e{-1}$	$-7.5625e{-2}$	$2.4904e{-2}$	$6.2067e0$	$6.8000e{-4}$
DIM 1 $q = p = 3$	$4.5030e{-1}$	$1.7614e{-1}$	$7.2447e{-1}$	$8.9790e{-2}$	$5.0927e0$	$5.4800e{-3}$
DIM 1 $q = p = 4$	$-3.2848e{-1}$	$-5.6898e{-1}$	$-8.7975e{-2}$	$7.6992e{-2}$	$4.2219e0$	$2.2800e{-2}$
DIM 2 $q = p = 1$	$1.4710e{-2}$	$1.6354e{-4}$	$2.9256e{-2}$	$4.7872e{-3}$	$3.6213e0$	$4.3920e{-2}$

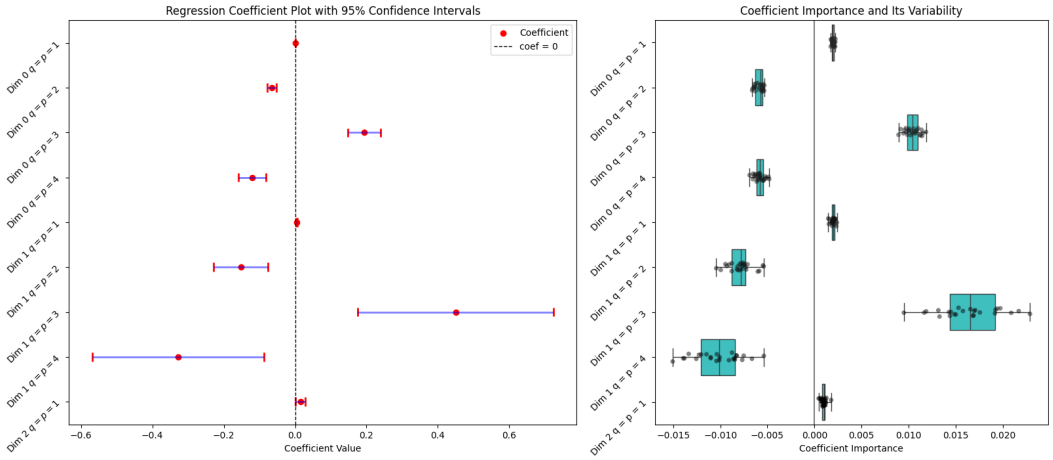

Figure 79: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: PCC and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 36: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: PCC and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$3.7603e{-3}$	$3.1880e{-3}$	$4.3325e{-3}$	$1.8512e{-4}$	$1.9837e1$	$0.0000e0$
DIM 0 $q = p = 1$	$8.0481e{-4}$	$6.6375e{-4}$	$9.4586e{-4}$	$4.2085e{-5}$	$2.0047e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-4.8646e{-2}$	$-5.7792e{-2}$	$-3.9500e{-2}$	$2.4564e{-3}$	$2.0622e1$	$0.0000e0$
DIM 0 $q = p = 3$	$1.3175e{-1}$	$9.6681e{-2}$	$1.6683e{-1}$	$8.7956e{-3}$	$1.6049e1$	$0.0000e0$
DIM 0 $q = p = 4$	$-8.1977e{-2}$	$-1.1222e{-1}$	$-5.1737e{-2}$	$7.0822e{-3}$	$1.2585e1$	$0.0000e0$
DIM 1 $q = p = 1$	$5.7958e{-3}$	$4.4040e{-3}$	$7.1875e{-3}$	$4.2205e{-4}$	$1.2288e1$	$0.0000e0$
DIM 1 $q = p = 2$	$-2.3960e{-1}$	$-2.9739e{-1}$	$-1.8181e{-1}$	$1.7591e{-2}$	$1.1567e1$	$0.0000e0$
DIM 1 $q = p = 3$	$7.0155e{-1}$	$4.8696e{-1}$	$9.1615e{-1}$	$6.3730e{-2}$	$8.8358e0$	$0.0000e0$
DIM 1 $q = p = 4$	$-4.6938e{-1}$	$-6.5634e{-1}$	$-2.8241e{-1}$	$5.3688e{-2}$	$6.8522e0$	$7.2000e{-4}$

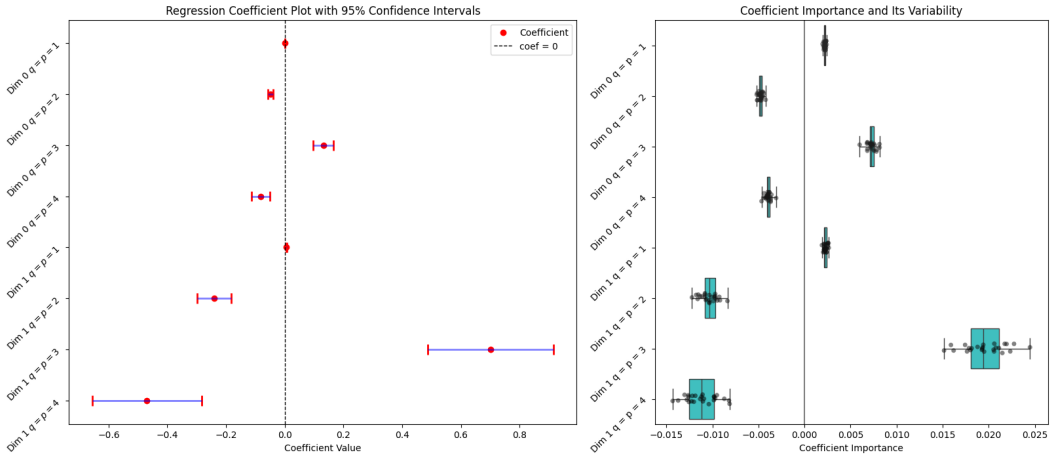

Figure 80: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: PCC and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 37: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: PCC and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\|t\|$	p
(INTERCEPT)	$1.7410e{-3}$	$1.1174e{-3}$	$2.3645e{-3}$	$1.7364e{-4}$	$1.0852e1$	$0.0000e0$
DIM 0 $q = p = 1$	$4.4869e{-4}$	$2.0435e{-4}$	$6.9302e{-4}$	$6.0081e{-5}$	$7.5765e0$	$8.8000e{-4}$
DIM 0 $q = p = 2$	$-2.5439e{-2}$	$-3.6694e{-2}$	$-1.4184e{-2}$	$2.6778e{-3}$	$1.0059e1$	$0.0000e0$
DIM 0 $q = p = 3$	$7.2124e{-2}$	$3.6451e{-2}$	$1.0780e{-1}$	$8.5342e{-3}$	$9.5325e0$	$0.0000e0$
DIM 0 $q = p = 4$	$-3.9465e{-2}$	$-6.8156e{-2}$	$-1.0774e{-2}$	$6.5266e{-3}$	$7.4546e0$	$2.4400e{-3}$
DIM 1 $q = p = 1$	$5.8645e{-3}$	$3.7476e{-3}$	$7.9814e{-3}$	$5.7891e{-4}$	$1.0361e1$	$0.0000e0$
DIM 1 $q = p = 2$	$-1.8806e{-1}$	$-2.8471e{-1}$	$-9.1417e{-2}$	$2.5273e{-2}$	$6.9312e0$	$1.3600e{-3}$
DIM 1 $q = p = 3$	$5.1224e{-1}$	$1.6175e{-1}$	$8.6273e{-1}$	$9.3985e{-2}$	$4.3896e0$	$4.0080e{-2}$

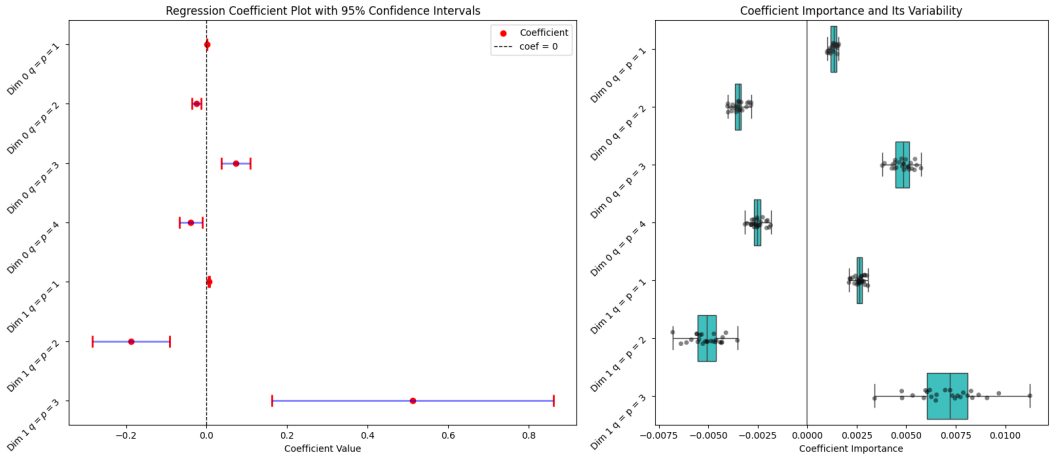

Figure 81: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: PCC and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 38: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: PCC and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$2.9357e{-3}$	$2.4457e{-3}$	$3.4258e{-3}$	$9.3103e{-5}$	$2.9678e1$	$0.0000e0$
DIM 0 $q = p = 1$	$7.9441e{-4}$	$5.9643e{-4}$	$9.9239e{-4}$	$5.0970e{-5}$	$1.5234e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-3.9411e{-2}$	$-5.0555e{-2}$	$-2.8267e{-2}$	$3.0967e{-3}$	$1.1528e1$	$0.0000e0$
DIM 0 $q = p = 3$	$1.4536e{-1}$	$8.6666e{-2}$	$2.0406e{-1}$	$1.5243e{-2}$	$7.7238e0$	$3.2000e{-4}$
DIM 0 $q = p = 4$	$-2.6746e{-1}$	$-4.1629e{-1}$	$-1.1862e{-1}$	$3.4332e{-2}$	$5.9503e0$	$1.3400e{-2}$
DIM 0 $q = p = 5$	$3.6745e{-1}$	$1.6384e{-1}$	$5.7106e{-1}$	$4.4056e{-2}$	$6.7737e0$	$7.4400e{-3}$
DIM 0 $q = p = 6$	$-2.1043e{-1}$	$-3.2185e{-1}$	$-9.9004e{-2}$	$2.3175e{-2}$	$7.8935e0$	$2.4800e{-3}$
DIM 0 $q = p = \infty$	$1.0477e{-2}$	$-3.4399e{-4}$	$2.1299e{-2}$	$2.2325e{-3}$	$5.2951e0$	$4.6320e{-2}$
DIM 1 $q = p = 1$	$5.0501e{-3}$	$3.3402e{-3}$	$6.7601e{-3}$	$5.6453e{-4}$	$1.0015e1$	$0.0000e0$
DIM 1 $q = p = 2$	$-2.0141e{-1}$	$-3.0629e{-1}$	$-9.6519e{-2}$	$4.0548e{-2}$	$6.6267e0$	$0.0000e0$
DIM 1 $q = p = 3$	$5.6797e{-1}$	$-1.1902e{-1}$	$1.2550e0$	$2.5080e{-1}$	$3.9494e0$	$1.9440e{-2}$

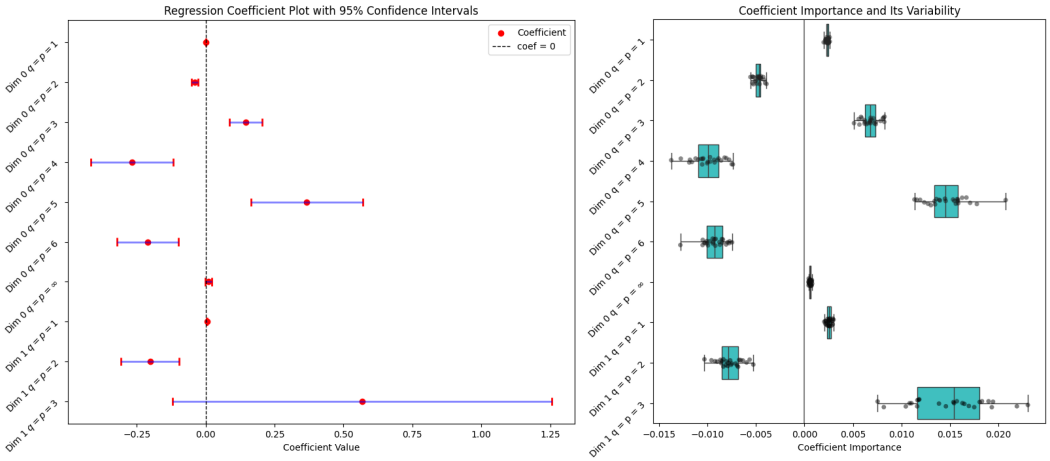

Figure 82: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: PCC and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 39: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: dmPFC and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$1.8114e{-3}$	$4.4714e{-4}$	$3.1757e{-3}$	$3.5342e{-4}$	$5.2203e0$	$2.0280e{-2}$
DIM 0 $q = p = 1$	$5.6317e{-4}$	$3.5552e{-4}$	$7.7083e{-4}$	$5.1113e{-5}$	$1.1222e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-3.3020e{-2}$	$-4.5732e{-2}$	$-2.0308e{-2}$	$3.5458e{-3}$	$9.0300e0$	$0.0000e0$
DIM 0 $q = p = 3$	$8.8753e{-2}$	$2.0050e{-2}$	$1.5746e{-1}$	$2.3931e{-2}$	$3.7293e0$	$3.4800e{-2}$
DIM 0 $q = p = 11$	$6.7070e{-2}$	$2.5333e{-2}$	$1.0881e{-1}$	$1.2946e{-2}$	$5.1823e0$	$7.5200e{-3}$
DIM 1 $q = p = 1$	$2.1597e{-3}$	$7.1246e{-4}$	$3.6069e{-3}$	$2.9764e{-4}$	$6.4576e0$	$1.4760e{-2}$
DIM 1 $q = p = 2$	$-9.8263e{-2}$	$-1.6514e{-1}$	$-3.1383e{-2}$	$1.3596e{-2}$	$5.4787e0$	$4.2560e{-2}$
DIM 1 $q = p = 14$	$-1.1407e{-2}$	$-1.7035e{-2}$	$-5.7794e{-3}$	$1.5985e{-3}$	$6.5236e0$	$1.2400e{-3}$
DIM 1 $q = p = \infty$	$-8.3111e{-2}$	$-1.4977e{-1}$	$-1.6452e{-2}$	$1.5753e{-2}$	$4.7382e0$	$4.5160e{-2}$
DIM 2 $q = p = 16$	$3.0546e{-3}$	$6.0739e{-4}$	$5.5017e{-3}$	$6.9384e{-4}$	$4.0559e0$	$4.7080e{-2}$
DIM 2 $q = p = 17$	$-2.5786e{-3}$	$-4.3843e{-3}$	$-7.7285e{-4}$	$4.2269e{-4}$	$7.1505e0$	$2.2800e{-3}$

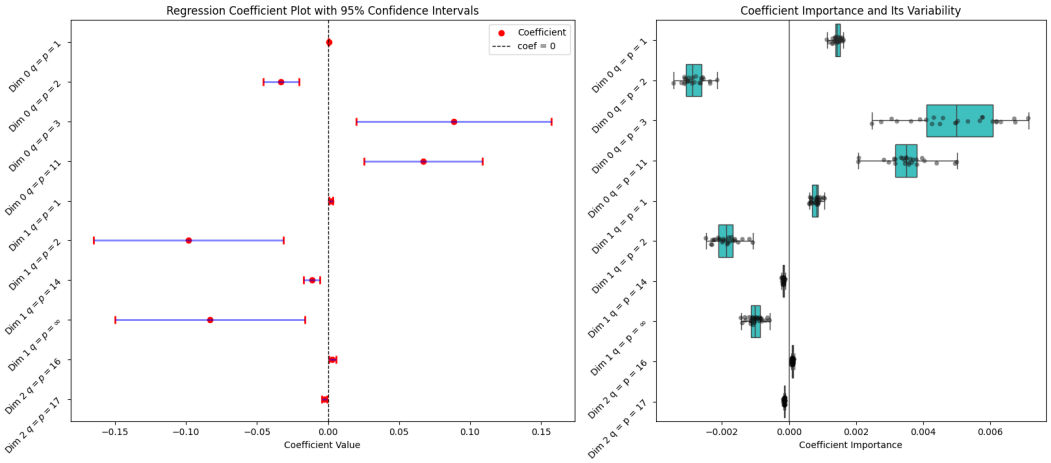

Figure 83: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: dmPFC and the hemisphere: L(eft) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 40: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: dmPFC and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$3.1380e{-4}$	$-8.5901e{-4}$	$1.4866e{-3}$	$3.1654e{-4}$	$1.3174e0$	$5.2556e{-1}$
DIM 0 $q = p = 1$	$5.7039e{-4}$	$4.2105e{-4}$	$7.1974e{-4}$	$5.4346e{-5}$	$1.1831e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-2.9279e{-2}$	$-3.9455e{-2}$	$-1.9103e{-2}$	$5.5756e{-3}$	$6.4241e0$	$0.0000e0$
DIM 0 $q = p = 3$	$1.0095e{-1}$	$3.9324e{-2}$	$1.6258e{-1}$	$4.3705e{-2}$	$3.0269e0$	$8.4000e{-4}$
DIM 0 $q = p = 11$	$6.7116e{-2}$	$3.3675e{-2}$	$1.0056e{-1}$	$1.0482e{-2}$	$4.5881e0$	$1.4600e{-2}$
DIM 1 $q = p = 1$	$1.4933e{-3}$	$4.5474e{-4}$	$2.5318e{-3}$	$2.9036e{-4}$	$5.8492e0$	$4.1200e{-3}$
DIM 1 $q = p = 2$	$-5.6171e{-2}$	$-1.0607e{-1}$	$-6.2675e{-3}$	$1.3891e{-2}$	$4.8249e0$	$2.0200e{-2}$
DIM 1 $q = p = 22$	$-6.8932e{-3}$	$-1.0936e{-2}$	$-2.8502e{-3}$	$7.2532e{-4}$	$9.1046e0$	$2.3200e{-3}$
DIM 1 $q = p = 23$	$4.5706e{-3}$	$1.2615e{-3}$	$7.8797e{-3}$	$8.2250e{-4}$	$4.4876e0$	$4.8360e{-2}$
DIM 1 $q = p = \infty$	$-4.9737e{-2}$	$-9.6498e{-2}$	$-2.9763e{-3}$	$1.0266e{-2}$	$6.1257e0$	$1.5040e{-2}$

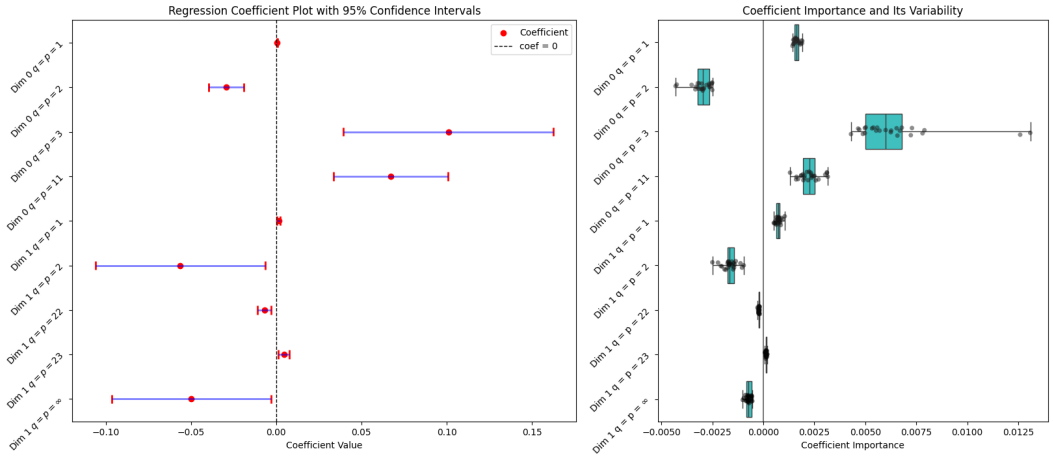

Figure 84: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: dmPFC and the hemisphere: L(eft) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 41: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: dmPFC and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$4.8202e{-4}$	$-7.8001e{-4}$	$1.7440e{-3}$	$2.9272e{-4}$	$1.9039e0$	$4.3344e{-1}$
DIM 0 $q = p = 1$	$4.0236e{-4}$	$2.1314e{-4}$	$5.9157e{-4}$	$7.8672e{-5}$	$5.4965e0$	$8.0000e{-5}$
DIM 0 $q = p = 2$	$-3.3138e{-2}$	$-4.5556e{-2}$	$-2.0720e{-2}$	$6.0554e{-3}$	$5.5138e0$	$0.0000e0$
DIM 0 $q = p = 3$	$1.3683e{-1}$	$6.3979e{-2}$	$2.0968e{-1}$	$3.5150e{-2}$	$4.1794e0$	$1.0400e{-3}$
DIM 1 $q = p = 1$	$3.8741e{-3}$	$2.3882e{-3}$	$5.3600e{-3}$	$4.1502e{-4}$	$8.7766e0$	$0.0000e0$
DIM 1 $q = p = 2$	$-1.1151e{-1}$	$-1.8843e{-1}$	$-3.4600e{-2}$	$1.9431e{-2}$	$5.1319e0$	$2.2000e{-2}$
DIM 1 $q = p = 14$	$-1.0395e{-2}$	$-1.6193e{-2}$	$-4.5961e{-3}$	$1.6721e{-3}$	$7.1680e0$	$2.4000e{-4}$
DIM 1 $q = p = 15$	$7.3163e{-3}$	$1.3269e{-3}$	$1.3306e{-2}$	$1.5463e{-3}$	$4.8113e0$	$3.2680e{-2}$
DIM 1 $q = p = 16$	$6.5807e{-3}$	$1.2895e{-3}$	$1.1872e{-2}$	$1.0768e{-3}$	$6.0820e0$	$2.4000e{-2}$

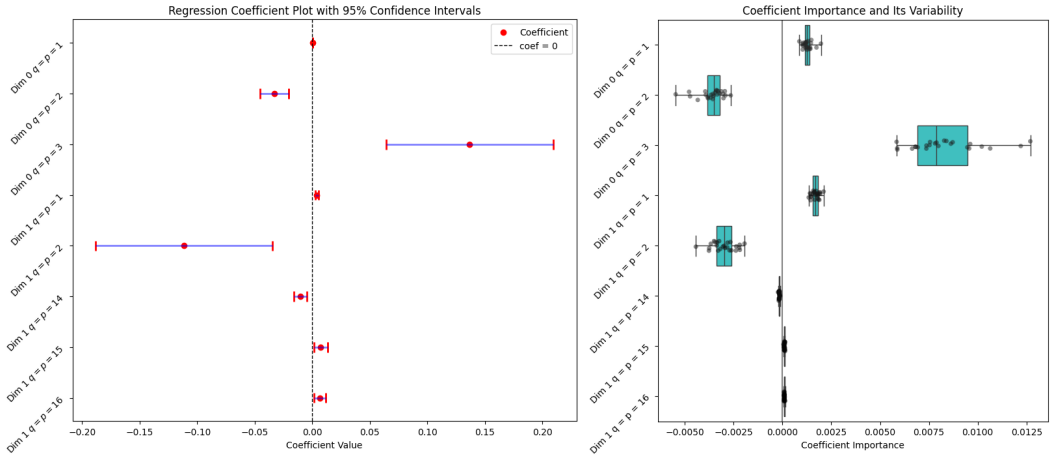

Figure 85: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: dmPFC and the hemisphere: R(ight) where the training data contains "Only Trained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

Table 42: This table summarizes the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: dmPFC and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs.

FEATURE	WEIGHT	95% CI LOWER	95% CI UPPER	SE	$\lvert t \rvert$	p
(INTERCEPT)	$1.9088e{-4}$	$-9.4506e{-4}$	$1.3268e{-3}$	$2.2561e{-4}$	$9.9811e{-1}$	$7.0760e{-1}$
DIM 0 $q = p = 1$	$6.6875e{-4}$	$5.2661e{-4}$	$8.1088e{-4}$	$5.6867e{-5}$	$1.0986e1$	$0.0000e0$
DIM 0 $q = p = 2$	$-4.0338e{-2}$	$-5.0386e{-2}$	$-3.0290e{-2}$	$3.3882e{-3}$	$1.2002e1$	$0.0000e0$
DIM 0 $q = p = 3$	$1.7443e{-1}$	$1.1521e{-1}$	$2.3,366e{-1}$	$1.8315e{-2}$	$9.8858e0$	$0.0000e0$
DIM 0 $q = p = 4$	$-3.2980e{-1}$	$-5.0822e{-1}$	$-1.5138e{-1}$	$5.6767e{-2}$	$5.9246e0$	$1.1600e{-3}$
DIM 1 $q = p = 1$	$2.5034e{-3}$	$1.3587e{-3}$	$3.6481e{-3}$	$4.2438e{-4}$	$6.8876e0$	$0.0000e0$
DIM 1 $q = p = 2$	$-9.7239e{-2}$	$-1.5799e{-1}$	$-3.6484e{-2}$	$2.1368e{-2}$	$4.2936e0$	$1.1640e{-2}$
DIM 1 $q = p = 13$	$-9.8151e{-3}$	$-1.4318e{-2}$	$-5.3121e{-3}$	$1.5374e{-3}$	$4.0777e0$	$2.7280e{-2}$
DIM 1 $q = p = 14$	$-6.9471e{-3}$	$-1.1090e{-2}$	$-2.8043e{-3}$	$9.1087e{-4}$	$9.0745e0$	$2.0000e{-4}$
DIM 1 $q = p = 22$	$3.4595e{-3}$	$1.0818e{-3}$	$5.8371e{-3}$	$4.9270e{-4}$	$6.2763e0$	$1.7400e{-2}$
DIM 2 $q = p = 9$	$-1.6268e{-2}$	$-3.2134e{-2}$	$-4.0227e{-4}$	$3.8915e{-3}$	$4.5707e0$	$4.7760e{-2}$

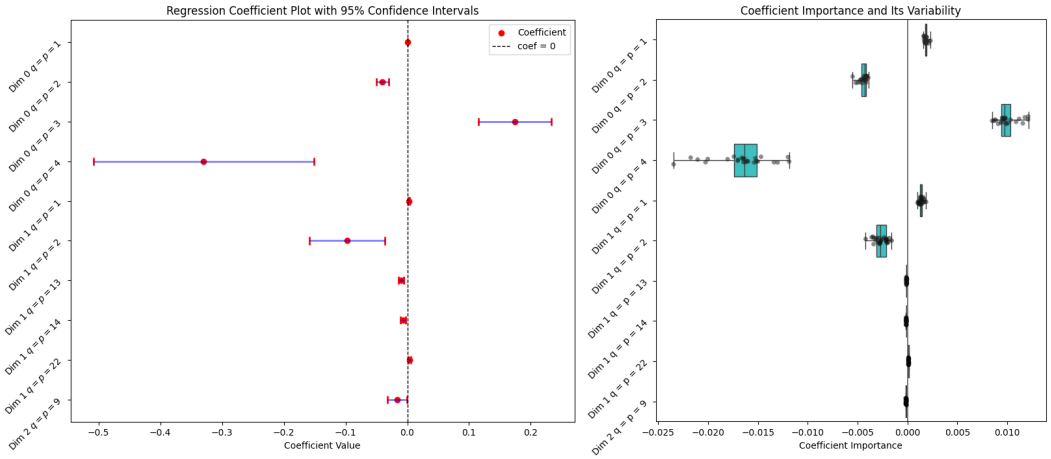

Figure 86: This figure illustrates the filtered-in features generated from the Second Pass (Appendix A.4.2) for the ROI: dmPFC and the hemisphere: R(ight) where the training data contains both "Only Trained" and "Trained plus Untrained" LLMs. On the left sub-figure, each feature's Weight is presented along with both the Lower and Upper Bounds of the 95% Confidence Intervals, while the right sub-figure illustrates its Coefficient Importance and Variability.

K ILLUSTRATIONS FOR BRAINSCORES IN DISCUSSIONS AND LIMITATIONS

This section presents various representations of brainscores corresponding to Section 4.

Additionally, we augment our illustrations with statistical analyses by examining correlations between the computed brainscores and the number of parameters [23] in each LLM (Section 4.2), the proportion of LLMs where trained ones outperform their untrained counterparts w.r.t their brainscores (Section 4.3), and whether the original LLM achieves a higher brainscore compared to its quantized version (Section 4.4), as outlined in Section 4.

[23]Refer to Table 3 in Appendix E.

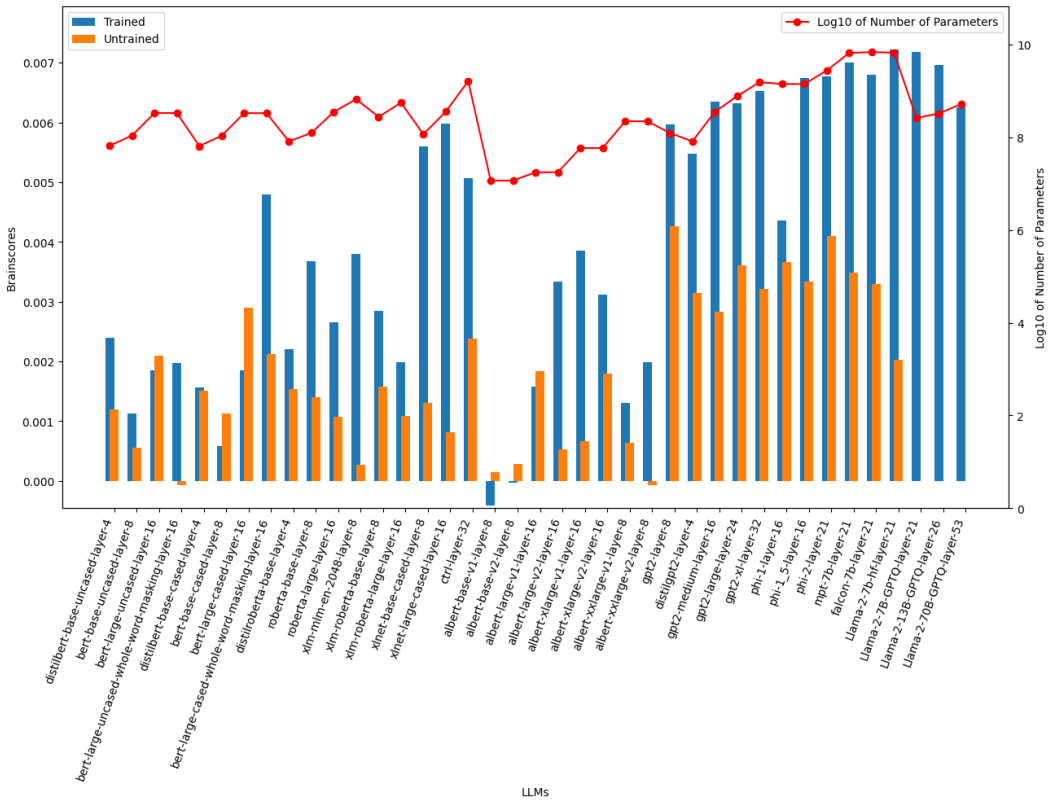

Figure 87: This figure illustrates the averaged brainscores computed for the intermediate-to-deep layer ($l = \frac{2}{3}n_{\text{layers}}$) according to Caucheteux et al. (2023) across all 39 LLMs, along with their untrained counterparts (excluding the last 3 LLMs). The brainscores are evaluated against the ROI: PCC and the (L)eft hemisphere, within the context of the task: "It's Not the Fall That Gets You". In this specific instance, 83.33% of the LLMs demonstrate that trained models outperform their untrained counterparts concerning their brainscores. Additionally, the Llama-2-7b-hf (trained) exhibits a higher brainscore compared to its quantized version, Llama-2-7B-GPTQ (trained). The figure also displays the logarithmic (with base 10) transformation of each LLM's number of parameters (refer to Table 3 in Appendix E).

Table 43: This table encapsulates the correlations, for the instance illustrated in Figure 87, across three metrics: Pearson r, Spearman ρ, and Kendall τ, supplemented with corresponding p values, obtained from correlating the computed brainscores with the number of parameters in each LLM (trained only). Each correlation is assessed within four distinct groups, determined by whether each LLM's number of parameters are represented in their raw value or transformed logarithmically (with base 10), and whether data points on quantized LLMs are incorporated. The decision to exclude quantized LLMs stems from their substantial reduction in the number of parameters, while maintaining comparable brainscores with their full-resolution configurations.

GROUP	PEARSON r	SPEARMAN ρ	KENDALL τ
CORR ALL TRAINED	50.87%	63.62%	47.46%
p	$9.44e{-4}$	$1.35e{-5}$	$2.27e{-5}$
CORR ALL TRAINED LOG	66.67%	63.62%	47.46%
p	$3.59e{-6}$	$1.35e{-5}$	$2.27e{-5}$
CORR NO QUANTIZED	58.37%	67.80%	51.56%
p	$1.87e{-4}$	$5.53e{-6}$	$1.07e{-5}$
CORR NO QUANTIZED LOG	69.92%	67.80%	51.56%
p	$2.10e{-6}$	$5.53e{-6}$	$1.07e{-5}$

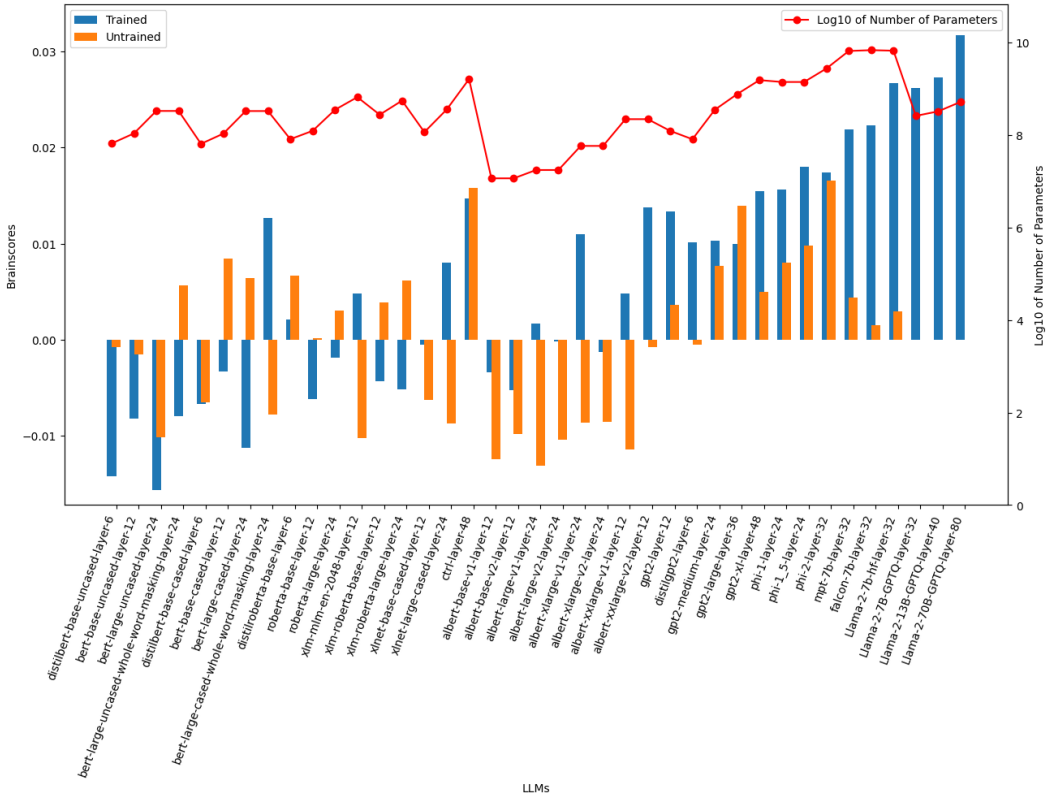

Figure 88: This figure illustrates the averaged brainscores computed for the last layer across all 39 LLMs, along with their untrained counterparts (excluding the last 3 LLMs). The brainscores are evaluated against the ROI: Whole Brain Mask and the (L)eft hemisphere, within the context of the task: "Pie Man". In this specific instance, 61.11% of the LLMs demonstrate that trained models outperform their untrained counterparts concerning their brainscores. Additionally, the Llama-2-7b-hf (trained) exhibits a higher brainscore compared to its quantized version, Llama-2-7B-GPTQ (trained). The figure also displays the logarithmic (with base 10) transformation of each LLM's number of parameters (refer to Table 3 in Appendix E).

Table 44: This table encapsulates the correlations, for the instance illustrated in Figure 88, across three metrics: Pearson r, Spearman ρ, and Kendall τ, supplemented with corresponding p values, obtained from correlating the computed brainscores with the number of parameters in each LLM (trained only). Each correlation is assessed within four distinct groups, determined by whether each LLM's number of parameters are represented in their raw value or transformed logarithmically (with base 10), and whether data points on quantized LLMs are incorporated. The decision to exclude quantized LLMs stems from their substantial reduction in the number of parameters, while maintaining comparable brainscores with their full-resolution configurations.

GROUP	PEARSON r	SPEARMAN ρ	KENDALL τ
CORR ALL TRAINED	51.33%	53.10%	37.70%
p	$8.33e{-4}$	$5.06e{-4}$	$7.66e{-4}$
CORR ALL TRAINED LOG	57.30%	53.10%	37.70%
p	$1.38e{-4}$	$5.06e{-4}$	$7.66e{-4}$
CORR NO QUANTIZED	65.98%	57.61%	41.66%
p	$1.20e{-5}$	$2.36e{-4}$	$3.75e{-4}$
CORR NO QUANTIZED LOG	63.93%	57.61%	41.66%
p	$2.70e{-5}$	$2.36e{-4}$	$3.75e{-4}$

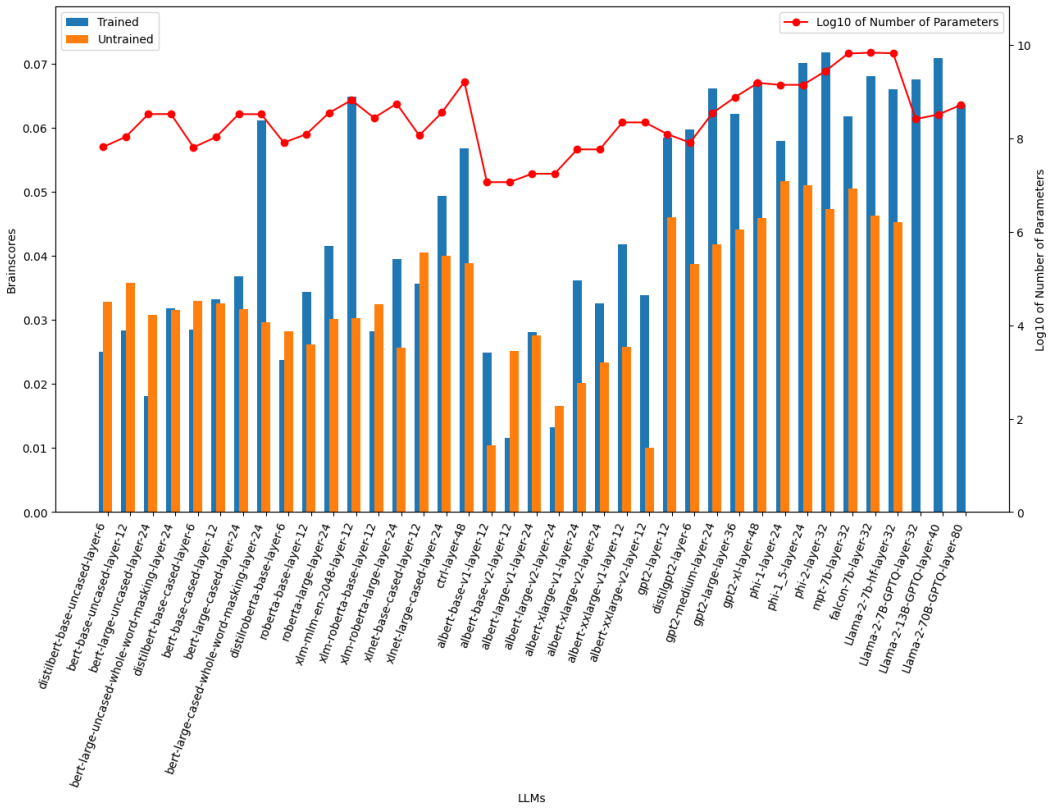

Figure 89: This figure illustrates the averaged brainscores computed for the last layer across all 39 LLMs, along with their untrained counterparts (excluding the last 3 LLMs). The brainscores are evaluated against the ROI: Whole Brain Mask and the (L)eft hemisphere, within the context of the task: "It's Not the Fall That Gets You". In this specific instance, 75% of the LLMs demonstrate that trained models outperform their untrained counterparts concerning their brainscores. But the Llama-2-7b-hf (trained) fails to exhibit a higher brainscore compared to its quantized version, Llama-2-7B-GPTQ (trained). The figure also displays the logarithmic (with base 10) transformation of each LLM's number of parameters (refer to Table 3 in Appendix E).

Table 45: This table encapsulates the correlations, for the instance illustrated in Figure 89, across three metrics: Pearson r, Spearman ρ, and Kendall τ, supplemented with corresponding p values, obtained from correlating the computed brainscores with the number of parameters in each LLM (trained only). Each correlation is assessed within four distinct groups, determined by whether each LLM's number of parameters are represented in their raw value or transformed logarithmically (with base 10), and whether data points on quantized LLMs are incorporated. The decision to exclude quantized LLMs stems from their substantial reduction in the number of parameters, while maintaining comparable brainscores with their full-resolution configurations.

GROUP	PEARSON r	SPEARMAN ρ	KENDALL τ
CORR ALL TRAINED	48.09%	72.99%	54.51%
p	$1.94e{-3}$	$1.34e{-7}$	$1.14e{-6}$
CORR ALL TRAINED LOG	74.67%	72.99%	54.51%
p	$4.79e{-8}$	$1.34e{-7}$	$1.14e{-6}$
CORR NO QUANTIZED	55.05%	78.17%	59.54%
p	$5.04e{-4}$	$1.82e{-8}$	$3.70e{-7}$
CORR NO QUANTIZED LOG	78.19%	78.17%	59.54%
p	$1.80e{-8}$	$1.82e{-8}$	$3.70e{-7}$

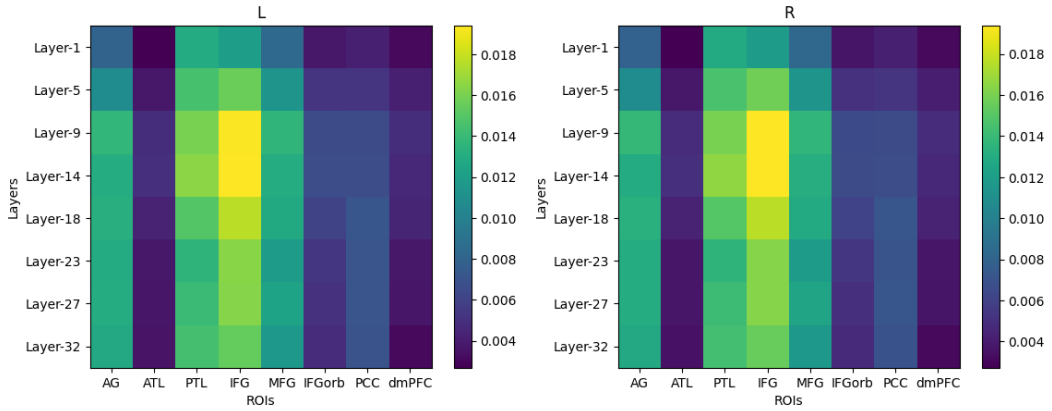

Figure 90: This figure depicts the averaged brainscores across the uniformly sampled eight layers, spanning from the first to the last layer (inclusive at both ends), of the LLM: Llama-2-7b-hf. These scores are computed concerning all eight language-related brain ROIs, within the context of the task: "It's Not the Fall That Gets You". The heatmap on the left represents the outcomes for the left hemisphere, while the one on the right illustrates those for the right hemisphere.

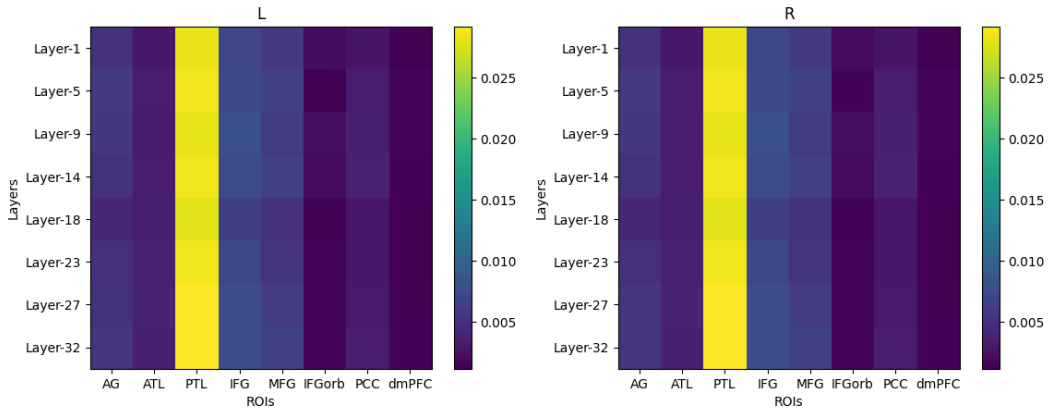

Figure 91: This figure depicts the averaged brainscores across the uniformly sampled eight layers, spanning from the first to the last layer (inclusive at both ends), of the LLM: Llama-2-7b-hf. These scores are computed concerning all eight language-related brain ROIs, within the context of the task: ”Shapes”. The heatmap on the left represents the outcomes for the left hemisphere, while the one on the right illustrates those for the right hemisphere.

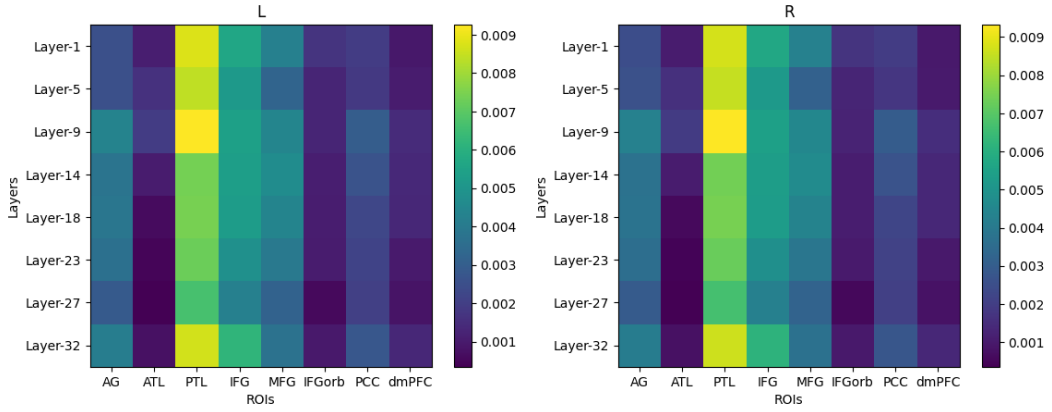

Figure 92: This figure depicts the averaged brainscores across the uniformly sampled eight layers, spanning from the first to the last layer (inclusive at both ends), of the LLM: Llama-2-7b-hf. These scores are computed concerning all eight language-related brain ROIs, within the context of the task: ”Pie Man”. The heatmap on the left represents the outcomes for the left hemisphere, while the one on the right illustrates those for the right hemisphere.