NeuroConText: Contrastive Text-to-Brain Mapping for Neuroscientific Literature

Raphaël Meudec[‡], Fateme Ghayem[‡*], Jérôme Dockès, Demian Wassermann, and Bertrand Thirion

Université Paris-Saclay, Inria, CEA, Palaiseau, 91120, France
{raphael.meudec, fatemeh.ghayyem, jerome.dockes, demian.wassermann, bertrand.thirion}@inria.fr
[‡]Equally contributed, [*]Corresponding author

Abstract. Neuroscience faces challenges in reliability due to limited statistical power, reproducibility issues, and inconsistent terminology. To address these challenges, we introduce NeuroConText, the first brain meta-analysis model that uses a contrastive approach to enhance the association between text data and brain activation coordinates reported in 20K neuroscientific articles from PubMed Central. NeuroConText integrates the capabilities of recent large language models (LLMs) rather than traditional bag-of-words methods, to better capture the text semantic, and improve the association with brain activation. It is adapted to processing neuroscientific text regardless of length and generalizes well across various textual content—titles, abstracts, and full-body. Our experiments show NeuroConText significantly outperforms state-of-the-art methods with a threefold increase in linking text to brain activations in terms of recall@10. NeuroConText also allows decoding brain images from latent text representations, successfully maintaining the quality of brain image reconstruction compared to the state-of-the-art.

Keywords: Brain meta-analysis · Text-brain association · Contrastive representation learning · Large language models (LLM).

1 Introduction

Hundreds of neuroscience articles are published each year, highlighting the accumulating knowledge within this domain. However, these studies often face limitations, such as terminology inconsistency, small sample size, limited statistical power, and reproducibility issues that undermine the reliability of articles' findings [2, 3]. To tackle these challenges, brain meta-analysis aggregates results from multiple studies to gain statistical power and enhance reliability [7, 10, 15].

Several efforts for brain meta-analysis have significantly advanced neuroscience. **BrainMap** [13] is a database that provides structured information about articles by curating reported coordinates and annotating them based on a taxonomy strategy. However, as neuroimaging literature grows rapidly, it becomes difficult to manually curate databases to cover new publications. **NeuroSynth** [24] automates the meta-analysis of neuroimaging studies by identifying relevant articles through keywords in their abstracts and extracting brain

activation coordinates to create brain statistical maps. This process allows for the efficient analysis of the whole literature. However, it relies on fixed keywords, which limits its ability to capture complex research topics and detailed insights. **NeuroQuery** [6] addresses these challenges by using a multivariate model trained on full-text publications to predict brain locations for neuroscience queries. It incorporates feature selection and adaptive regularization methods to manage sparse, high-dimensional inputs effectively. However, it remains limited by a bag-of-words model and a regression-based method that struggles to yield high association scores between text and brain images. **Text2Brain** [17] links text and brain imaging data by leveraging semantic insights provided by large language models (LLM), but it is not effective in processing long input text, resulting in weaker associations between text and brain images.

While meta-analysis methods mainly rely on regression-based techniques, contrastive learning, a transformation-based approach, has been shown to effectively bridge text and image by establishing a shared latent space between the two modalities [23, 9, 20]. In particular, CLIP [20] is a contrastive approach for matching image-caption pairs. In CLIP, the contrastive learning framework involves training a model to identify correct (image, text) pairings from a batch of possible combinations. This is achieved by maximizing the cosine similarity between the embeddings of the actual pairs while minimizing the similarity of incorrect pairs, using a symmetric cross-entropy loss over these similarity scores. CLIP demonstrates how training a text and image encoder within a contrastive framework enables remarkable performance in associating text and image.

In this paper, inspired by the contrastive approach in CLIP [20], we introduce *NeuroConText*, the first contrastive-based approach for brain meta-analysis that enhances the association between neuroscientific texts and brain activation maps by using a shared latent space between text and image modalities. NeuroConText also leverages advanced LLMs like GPT-Neo-1.3B [1, 8] and Mistral-7B [11] to bring semantics into consideration, unlike bag-of-words methods. It can process text queries of any length, effectively handling diverse textual contents (title, abstract or full-body text) from neuroscientific articles. By using the Dictionary of Functional Modes (DiFuMo) [4] atlas, we address the high dimensionality challenge of fMRI images. Our experiments demonstrate that NeuroConText not only enhances the association scores by up to threefold compared to the baselines NeuroQuery and Text2Brain, but also maintains the brain image reconstructions from text latent representations.

2 Methodology

2.1 Data Preparation

Downloading Articles: To run coordinate-based brain meta-analysis, we first download and prepare neuroscientific articles from PubMed Central (PMC) using Pubget. Pubget is an open-access tool that extracts text, metadata, stereotactic coordinates, and Term Frequency-Inverse Document Frequency (TF-IDF)

[21] from the articles. We expand the dataset used in NeuroQuery [6] from 14K to 20K articles by incorporating recent publications.

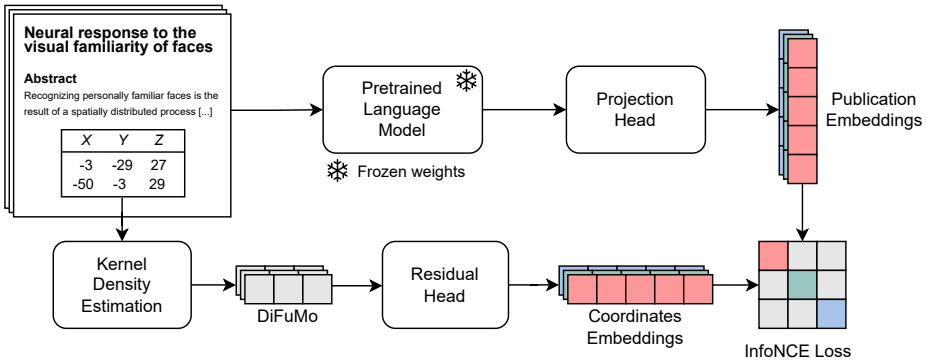
Text and Coordinate Feature Representation: To perform accurate meta-analysis on articles' text and coordinates, we first provide suitable representations of these two modalities. For text feature representation, we propose to use LLMs' embeddings. We extract the embeddings corresponding to the articles' data frames comparing four different models: SciBERT, GPT-Neo-125M, GPT-Neo-1.3B, and Mistral-7B. For long input queries, such as analyzing the full-body text of the articles, to overcome the token size limits of LLM, we divide the texts into chunks matching the LLM's token size limit. Subsequently, we generate embeddings for each chunk and represent the article's text by the average of these chunk embeddings. Our experiments with different aggregation strategies showed that averaging chunk embeddings provides the best results. This strategy for text data preparation addresses the limitations of the widely used TF-IDF [21] approach. While TF-IDF measures the significance of a term in an article compared to a set of articles, it remains a bag-of-words, and does not capture semantics. For the representation of stereotactic coordinates, we adopt the approach outlined in NeuroQuery to create Kernel Density Estimation (KDE) representations from activation peaks. To mitigate the risk of overfitting associated with high brain image dimensionality, we use the DiFuMo representation coefficients of the images, choosing dictionary sizes 256 and 512 [4, 22, 14].

2.2 NeuroConText: Our Proposed Framework for Contrastive Text-to-Brain Mapping

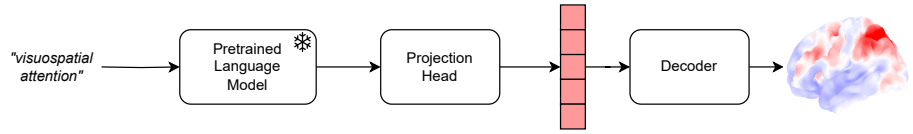
NeuroConText has two primary objectives: *i)* enhancing association between text content and brain activation coordinates reported in the neuroscientific articles, and *ii)* estimating brain activation maps from any given neuroscientific text. NeuroConText overall framework is shown in Fig.1. We detail it below.

A contrastive model for text and brain association (Fig.1-A): NeuroConText takes text embeddings and DiFuMo representation of the KDE as input features. Then, to bridge the gap between these two different modalities *i.e.* text and DiFuMo, we define a shared latent space through two distinct encoders. We consider a projection head to encode text embeddings, and a residual head to encode DiFuMo. We train these two networks simultaneously using the InfoNCE contrastive loss [18], as done in the CLIP paper [20]. This training allows the model to derive a shared latent space, capturing the intricate associations between the brain activation maps and the corresponding textual data.

Estimating brain activation map from an arbitrary text (Fig.1-B): We train a decoder to transform text latent representations, obtained from the output of the residual head, into brain images represented by DiFuMo coefficients. This training occurs on the same dataset used for the text and image encoders. Once the decoder is trained, in the inference phase, we can estimate brain images for any given neuroscientific query. This process involves extracting the query embeddings from the LLM used during data preparation, passing these



(A) Training of NeuroConText contrastive model on neuroscientific publications.



(B) Brain encoding of a query through the NeuroConText text latent representation.

Fig. 1: **NeuroConText:** (A) We train a contrastive model on a large corpus to retrieve a shared latent space between coordinates and text from neuroscientific articles. We leverage pre-trained LLMs to obtain an initial text embedding from the text and add a projection layer to align with embeddings of coordinates. Snowflakes denote models with frozen weights. (B) We train a decoder from text latent space to reproduce brain images from any query.

embeddings through the trained projection head to yield the text latent representation. Finally, we estimate the DiFuMo coefficients of the query by feeding the text latent representation into the trained decoder.

2.3 Related Work

Neuroquery is a brain meta-analysis tool trained on approximately 14K neuroimaging studies, which maps neuroscientific queries to brain images by extracting relevant terms through TF-IDF. It employs Gaussian Kernel Density Estimation on peak activation coordinates to generate spatial representations of brain activity. NeuroQuery uses Ridge regression to link TF-IDF features with brain activation density maps, thereby estimating activations from text.
Text2Brain overcomes the fixed size vocabulary from TF-IDF through SciBERT embeddings, which capture the semantic information of the literature. It leverages a convolutional neural network to produce 3D brain activation maps from the embeddings. To improve the model's generalizability across various cognitive concepts and tasks, Text2Brain uses data augmentation techniques, such as varying query types. There are several differences between our NeuroConText

method and Text2Brain, such as random text selection vs. long text processing, the choice of LLM (SciBERT vs. Mistral-7B), high-dimensional brain maps vs. DiFuMo coefficients, and regression via 3D CNN vs. introducing shared latent space via text and image encoders and the use of contrastive loss.

2.4 Evaluation Metrics Across Methodologies

The evaluation framework is divided into two categories: one for text-brain association tasks (Fig.1-A) and the other for mapping text to brain images (Fig.1-B).

The first task is the ability of the model to retrieve the right brain activation from a given input text. We evaluate the performance of NeuroConText on this retrieval task through recall@K and Mix&Match. For a given input text, the recall@K measures the presence of the true corresponding activation map in the top K retrieved maps. The Mix&Match metric [16] quantifies the frequency with which the estimated map more closely resembles the true map compared to a randomly selected map from the test set. We compute those metrics for the baselines by comparing the correlation of the predicted maps to the true target.

The second task is estimating brain activation from a given text. We leverage contrast descriptions from the Individual Brain Charting [19] library and compare the predicted activation to the average group through the Dice score [5]. For a given threshold t, the Dice score D_t measures how well activations cooccurs on the true target y and the predicted activation $\hat{y}$: $D_t = \frac{2(|\mathbb{1}_{\mathbf{y}>t}\mathbb{1}_{\hat{\mathbf{y}}>t}|)}{|\mathbb{1}_{\mathbf{y}>t}|+|\mathbb{1}_{\hat{\mathbf{y}}>t}|}$ where $\mathbb{1}_{\hat{\mathbf{y}}>t}$ indicates if a given voxel activation is above threshold t and $|.|$ indicates the norm, corresponding to a volume in image space.

3 Experimental Results

3.1 Model architecture

The input text is processed through the Mistral-7b, a projection head composed of a dense layer, GELU activation onto DiFuMo dimensional space, and two residual heads. Each residual head module comprises a fully connected layer, a GELU activation, a dropout with a rate of 0.5, and a normalization layer. The brain map input is processed through three residual heads composed of a dense layer and a GELU activation. We set the default batch size to 128, learning rate 5e-4, and weight decay 0.1, over 50 epochs. The output size is set to the DiFuMo dimension. The decoder consists of two residual heads with the abovementioned composition. The code and architecture details are publicly available at https://github.com/ghayem/NeuroConText.

3.2 Comparison with the baselines

In this section, we compare NeuroConText with NeuroQuery and Text2Brain. We use the Mistral-7B LLM, and we set the DiFuMo dictionary size to 512.

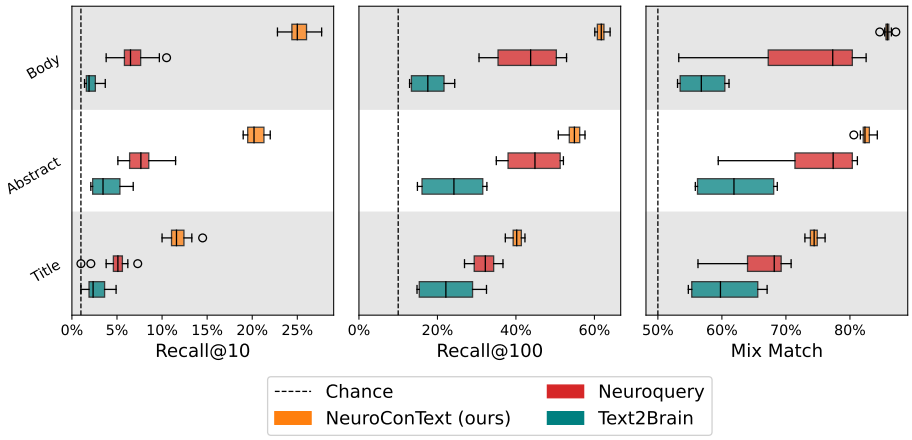


Fig. 2: Association of brain activation maps with article text: from a given query, either the title, abstract or body of a neuroscientific publication, we measure the ability of our model to associate the corresponding brain activation map. We report the recall@K, $K \in \{10, 100\}$, and mix&match score detailed in Section 2.4. The evaluation is performed on 15-fold cross-validation with 1K test set. NeuroConText performs better than NeuroQuery and Text2Brain. Our method also successfully generalizes to long full-body text, while Text2Brain and NeuroQuery fail. For more numerical details see Table **??** in the Appendix.

We split the dataset into 19K train and 1K test samples. Subsequently, we retrieve the text latent and DiFuMo latent of the test set through the trained projection head and residual head (Fig.1-A). Finally, we calculate association metrics, Recall@K for $K \in \{10, 100\}$, and Mix&Match, on the test data. For a fair comparison, we use the pre-trained models of NeuroQuery and Text2Brain to estimate brain maps from the text in the same test set and calculate the association metrics as detailed in section 2.4. To prevent leakage in evaluating the performance of the two baselines, we exclude articles used in the training of the NeuroQuery model from the test set. This experiment is performed with 15-folds cross-validation. The results, shown in Fig.2, depict the superiority of NeuroConText over the baselines in associating text with corresponding brain maps. Specifically, when comparing the recall@10 metric for body text, NeuroConText achieves up to 22.6%, significantly outperforming NeuroQuery, which only achieves 7%, and Text2Brain, which achieves 1.4%. Similar comparisons for other text sections and association metrics underscore the effectiveness of our contrastive method in linking text with brain maps, in contrast to the state-of-the-art methods, which fall short. Finally, we remark that expanding the training size from 10k to 19k enhanced NeuroContex's performance by 4.5% (see Fig.**??** in the Appendix).

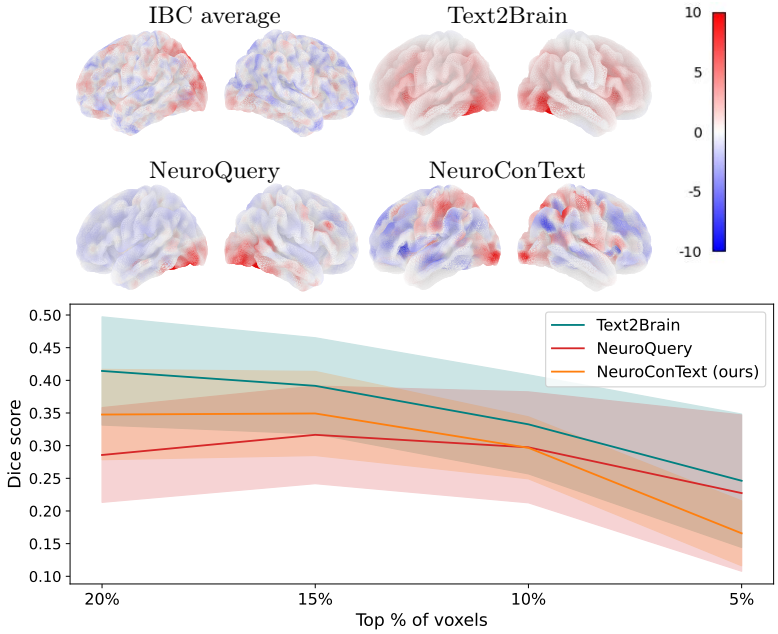


Fig. 3: (top) Thresholded projection on the surface of an IBC contrast definition: *read jabberwocky vs pseudo-word*. We project each image on the surface and threshold it, keeping 10% activation for each. Similarly, we plot the average IBC map for this contrast. The color bar shows the statistical value for IBC dataset (in z scale). (bottom) Encoding of IBC contrasts definition: We evaluate our encoding model on IBC contrast definitions, showing that encoding from the latent space performs similarly to other methods specifically trained for encoding.

3.3 Estimating brain maps from text latent space representation

We show the capability of NeuroConText to generate brain maps from text latent representations (see Fig.1-B). To assess the quality of these estimated brain maps, we encode the definition of each contrast from the Individual Brain Charting (IBC) [19] image dataset. We compare the encoded definition to the IBC average map with the Dice score detailed in section 2.4. As shown in Fig.3, the scores of brain maps reconstructed from the text latents closely resemble those derived from the baselines, indicating that the text latents of NeuroConText contain significant information about brain activations.

3.4 Ablation study

NeuroConText vs regression-based strategies We compare NeuroConText with regression-based methods to associate texts with brain maps. For NeuroConText, we follow the procedure in Section 3.2. As a regression-based approach,

Table 1: A comparison between proposed contrastive model NeuroConText and regression-based strategies in text-brain association.

Method \ Metric [%]	recall@10	recall@100	mix&match
NeuroConText (ours)	$\mathbf{22.6 \pm 1.4}$	$\mathbf{57.8 \pm 1.6}$	$\mathbf{84.2 \pm 0.9}$
RidgeCV (linear model)	14.9 ± 0.8	43.7 ± 1.2	76.5 ± 0.4
MLP (non-linear model)	11.4 ± 0.4	39.4 ± 1.2	74.1 ± 0.8

Table 2: Comparing different LLMs and DiFuMo size on NeuroConText performance. Mistral-7B paired with the DiFuMo 512 achieves the highest scores.

Setup \ Metric [%]		recall@10	recall@100	mix&match
DiFuMo 256	Mistral-7B	$\mathbf{19.8 \pm 0.9}$	$\mathbf{51.6 \pm 1.3}$	$\mathbf{81 \pm 0.6}$
	GPT-Neo-1.3B	18.1 ± 0.6	48.2 ± 1.3	79.4 ± 0.3
	GPT-Neo-125M	15.1 ± 0.6	42.3 ± 1.3	76.4 ± 0.3
	SciBERT	15.1 ± 0.6	42.8 ± 0.9	76.9 ± 0.3
DiFuMo 512	Mistral-7B	$\mathbf{22.6 \pm 1.4}$	$\mathbf{57.8 \pm 1.6}$	$\mathbf{84.2 \pm 0.9}$
	GPT-Neo-1.3B	21.5 ± 1.1	54.8 ± 1.1	82.7 ± 0.5
	GPT-Neo-125M	17.5 ± 1.1	48.2 ± 1.5	79.7 ± 0.7
	SciBERT	17.9 ± 0.8	50.3 ± 1.5	81 ± 0.8

we consider two models: RidgeCV (linear model) and a Multilayer perception (MLP) (non-linear model). We set the MLP architecture with three linear layers of 512 units each, incorporating layer normalization after the first two layers and a dropout (p=0.5) after the first layer. This model was trained over 50 epochs using the Adam optimizer [12] with a learning rate of $5e^{-4}$. We train these two models on the average Mistral-7B embeddings of the full-body chunks to estimate the DiFuMo coefficients. We evaluate the association performance of the trained models on the test data with the metrics detailed in 2.4. As detailed in Table 1, the regression-based methods fail to associate text with the brain effectively. Meanwhile, the regression-based method in this experiment significantly outperforms NeuroQuery and Text2Brain baselines, as shown in Fig.2. This comes from replacing the TF-IDF and SciBERT embedding features in the baselines with Mistral-7B embeddings and the proposed strategy for leveraging the full-body text.

Impact of large language model and DiFuMo size. We compare different LLMs and DiFuMo sizes on the performance of our proposed method. We replicated the experiment in section 3.2, changing the LLM to GPT-Neo-1.3B, GPT-Neo-125M, or SciBERT, alongside two DiFuMo sizes: 256 and 512. The findings are summarized in Table 2. We see that the larger language model with Mistral-7B significantly enhances performance, achieving a score more than 4% higher than the SciBERT. The results indicate that the DiFuMo with size 512 outperforms 256, improving scores by approximately 3%.

4 Conclusion

In this paper, we introduced NeuroConText, a new coordinate-based brain meta-analysis framework. NeuroConText performs text-brain association on neuroscientific articles by defining a shared latent space between text and coordinates modalities through a contrastive loss. By training a decoder, NeuroConText also estimates brain maps from text. NeuroConText provides a new framework to leverage articles' full-body texts by segmenting long texts and using recent advancements in LLM technology. We performed several experiments to compare the performance of NeuroConText with two baselines, NeuroQuery and Text2Brain. The results demonstrated NeuroConText's superior performance in associating texts to the brain, but show limited gain in predicting brain maps from text latent representations. The latent space could benefit from combining different loss objectives adapted to downstream applications. Our experiments also included various ablation studies. We evaluated the impact of full-body text, advanced LLMs, and DiFuMo dictionary size on text-brain association and mapping. In our ablation study, we also showed that NeuroConText enhances association scores and text-to-brain mapping compared to regression-based methods. As a future work, we will incorporate data augmentation into our framework, and investigate other contrastive techniques.

Acknowledgement This work is supported by the KARAIB AI chair (ANR-20-CHIA-0025-01), the ANR-22-PESN-0012 France 2030 program, and the HORIZON-INFRA-2022-SERV-B-01 EBRAINS 2.0 infrastructure project.

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Black, S., Leo, G., Wang, P., Leahy, C., Biderman, S.: Gpt-neo: Large scale autoregressive language modeling with mesh-tensorflow. Zenodo **1.0** (2021)
2. Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafò, M.R.: Power failure: why small sample size undermines the reliability of neuroscience. Nature reviews neuroscience **14**(5), 365–376 (2013)
3. Carp, J.: The secret lives of experiments: methods reporting in the fmri literature. Neuroimage **63**(1), 289–300 (2012)
4. Dadi, K., Varoquaux, G., Machlouzarides-Shalit, A., Gorgolewski, K.J., Wassermann, D., Thirion, B., Mensch, A.: Fine-grain atlases of functional modes for fmri analysis. NeuroImage **221**, 117126 (2020)
5. Dice, L.R.: Measures of the Amount of Ecologic Association Between Species. Ecology **26**(3), 297–302 (1945). https://doi.org/10.2307/1932409, https://www.jstor.org/stable/1932409
6. Dockès, J., Poldrack, R.A., Primet, R., Gözükan, H., Yarkoni, T., Suchanek, F., Thirion, B., Varoquaux, G.: Neuroquery, comprehensive meta-analysis of human brain mapping. Elife **9**, e53385 (2020)
7. Fox, P.T., Parsons, L.M., Lancaster, J.L.: Beyond the single study: function/location metanalysis in cognitive neuroimaging. Current opinion in neurobiology **8**(2), 178–187 (1998)
8. Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., Phang, J., He, H., Thite, A., Nabeshima, N., et al.: The pile: An 800gb dataset of diverse text for language modeling. arXiv preprint arXiv:2101.00027 (2020)
9. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9729–9738 (2020)
10. Horn, J.D.V., Grafton, S.T., Rockmore, D., Gazzaniga, M.S.: Sharing neuroimaging studies of human cognition. Nature neuroscience **7**(5), 473–481 (2004)
11. Jiang, A.Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D.S., Casas, D.d.l., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., et al.: Mistral 7b. arXiv preprint arXiv:2310.06825 (2023)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
13. Laird, A.R., Lancaster, J.J., Fox, P.T.: Brainmap: the social evolution of a human brain mapping database. Neuroinformatics **3**, 65–77 (2005)
14. Menuet, R., Meudec, R., Dockès, J., Varoquaux, G., Thirion, B.: Comprehensive decoding mental processes from web repositories of functional brain images. Sci. Rep. **12**(1), 7050 (Apr 2022)
15. Minzenberg, M.J., Laird, A.R., Thelen, S., Carter, C.S., Glahn, D.C.: Meta-analysis of 41 functional neuroimaging studies of executive function in schizophrenia. Archives of general psychiatry **66**(8), 811–822 (2009)
16. Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.M., Malave, V.L., Mason, R.A., Just, M.A.: Predicting human brain activity associated with the meanings of nouns. science **320**(5880), 1191–1195 (2008)
17. Ngo, G.H., Nguyen, M., Chen, N.F., Sabuncu, M.R.: Text2brain: Synthesis of brain activation maps from free-form text query. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VII 24. pp. 605–614. Springer (2021)

18. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
19. Pinho, A.L., Amadon, A., Ruest, T., Fabre, M., Dohmatob, E., Denghien, I., Ginisty, C., Becuwe-Desmidt, S., Roger, S., Laurier, L., Joly-Testault, V., Médiouni-Cloarec, G., Doublé, C., Martins, B., Pinel, P., Eger, E., Varoquaux, G., Pallier, C., Dehaene, S., Hertz-Pannier, L., Thirion, B.: Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping. Scientific Data **5**, 180105 (Jun 2018). https://doi.org/10.1038/sdata.2018.105
20. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PMLR (2021)
21. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Information processing & management **24**(5), 513–523 (1988)
22. Thomas, A.W., Ré, C., Poldrack, R.A.: Self-supervised learning of brain dynamics from broad neuroimaging data. arXiv preprint arXiv:2206.11417 (2022)
23. Tian, Y., Krishnan, D., Isola, P.: Contrastive multiview coding. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16. pp. 776–794. Springer (2020)
24. Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D.: Large-scale automated synthesis of human functional neuroimaging data. Nature methods **8**(8), 665–670 (2011)