NiCLIP: Neuroimaging contrastive language-image pretraining model for predicting text from brain activation images

Julio A. Peraza[1], James D. Kent[2], Thomas E. Nichols[3], Jean-Baptiste Poline[4], Alejandro de la Vega[2], Angela R. Laird[1]

[1] Department of Physics, Florida International University, Miami, FL, USA
[2] Department of Psychology, University of Texas at Austin, Austin, TX, USA
[3] Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford, Oxford, UK
[4] Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada

Abstract

Predicting tasks or cognitive domains based on brain activation maps has remained an open question within the neuroscience community for many years. Meta-analytic functional decoding methods aim to tackle this issue by providing a quantitative estimation of behavioral profiles associated with specific brain regions. Existing methods face intrinsic challenges in neuroimaging meta-analysis, particularly in consolidating textual information from publications, as they rely on limited metrics that do not capture the semantic context of the text. The combination of large language models (LLMs) with advanced deep contrastive learning models (e.g., CLIP) for aligning text with images has revolutionized neuroimaging meta-analysis, potentially offering solutions to functional decoding challenges. In this work, we present NiCLIP, a contrastive language-image pretrained model that predicts cognitive tasks, concepts, and domains from brain activation patterns. We leveraged over 23,000 neuroscientific articles to train a CLIP model for text-to-brain association. We demonstrated that fine-tuned LLMs (e.g., BrainGPT models) outperform their base LLM counterparts. Our detailed evaluation of NiCLIP predictions revealed that performance is optimized when using full-text articles instead of abstracts, as well as a curated cognitive ontology with precise task-concept-domain mappings. Our results indicated that NiCLIP accurately predicts cognitive tasks from group-level activation maps provided by the Human Connectome Project across multiple domains (e.g., emotion, language, motor) and precisely characterizes the functional roles of specific brain regions, including the amygdala, hippocampus, and temporoparietal junction. However, NiCLIP showed limitations with noisy subject-level activation maps. NiCLIP represents a significant advancement in quantitative functional decoding for neuroimaging, offering researchers a powerful tool for hypothesis generation and scientific discovery.

Keywords: fMRI, meta-analysis, large language model, CLIP, functional decoding, neural network

Corresponding Author: **Dr. Angela R. Laird, Professor of Physics, Florida International University, Miami, FL, USA | Email: alaird@fiu.edu**

1. Introduction

One of the many open questions in neuroscience is whether cognitive processes can be inferred from the activation of a specific brain region (Poldrack, 2006). This task is commonly known in the literature by several different names (Salo et al., 2023), for instance "informal reverse inference" (Poldrack, 2011), "functional characterization analysis" (Bzdok et al., 2013a; Cieslik et al., 2013; Rottschy et al., 2013), "open-ended decoding" (Rubin et al., 2017), or "functional decoding" (Amft et al., 2015; Bzdok et al., 2013b; Nickl-Jockschat et al., 2015). Although the terminology lacks standardization, here we will refer to such methods as functional decoding. A way of functional decoding is employed by neuroscientists to explain instances of activation in a particular brain area by drawing qualitative associations from prior studies that observed activation in the same location. Such informal inferences are limited and susceptible to investigator biases. The release of coordinate-based meta-analytic (CBMA) databases, such as BrainMap (Fox et al., 2005; Laird et al., 2011, 2009, 2005) and Neurosynth (Yarkoni et al., 2011), along with the Cognitive Atlas ontology (Poldrack et al., 2011), has provided us with robust data to facilitate quantitative meta-analytic functional decoding in neuroimaging.

The implementation of meta-analytic functional decoding methods, specifically CBMA correlation decoders, on popular websites (e.g., Neurosynth (Yarkoni et al., 2011) and NeuroVault (Gorgolewski et al., 2015)) and meta-analysis software (e.g., NiMARE (Salo et al., 2023)) has boosted their popularity within the neuroimaging community. Functional decoding has equipped researchers with tools to quantitatively characterize brain activation patterns (Salo et al., 2023), addressing the complex task of traditional qualitative inference (Rubin et al., 2017). However, such methods depend on a straightforward computation of spatial similarities between input maps and a collection of meta-analytic maps derived from large-scale CBMA. This approach presents several limitations. For instance, correlation decoders are not based on a formal model, fail to identify latent structures associated with specific cognitive processes, and lack sensitivity to unseen brain patterns (Rubin et al., 2017). Moreover, correlation scores merely indicate the confidence of a decoder in its predictions, neglecting the accuracy since the labels assigned to meta-analytic maps for training the decoder are not validated. The GC-LDA model (Rubin et al., 2016, 2017) has addressed some of these limitations by learning joint probability from latent variables in the model, allowing predictions on unseen brain patterns. Nevertheless, GC-LDA is significantly influenced by the data, context, and assumptions established before model training, such as spatial priors, the number of topics, and other model parameters. Consequently, varying initial conditions are likely to yield different results. Additionally, GC-LDA operates entirely unsupervised, suggesting that the model is not optimized for classification or prediction accuracy. Thus, other models refined to enhance prediction parameters should outperform GC-LDA (Rubin et al., 2017).

More advanced decoding approaches using a large set of maps from NeuroVault, trained to predict tasks and concepts from the Cognitive Atlas (Poldrack et al., 2011), have demonstrated enhanced performance compared to the CBMA-based Neurosynth and GC-LDA methods (Mensch et al., 2021, 2017; Menuet et al., 2022; Varoquaux et al., 2018)(Menuet et al., 2022). However, despite ongoing efforts by the community to encourage researchers to upload their statistical maps to the NeuroVault

repository, the majority of maps from neuroimaging studies remain unshared or undistributed, which continues to limit NeuroVault images to a small number of tasks and mental functions and few images per task or function. Moreover, the metadata associated with statistical maps in NeuroVault is often inadequately annotated and requires more consistent adherence to standard ontologies. For instance, it is quite common to encounter images with inaccurately labeled image modality, statistics map type, and cognitive atlas task (Peraza et al., 2025). Although the use of whole-brain statistical maps along with a vocabulary based on an ontology (e.g., Cognitive Atlas) for training decoding models is generally preferred, the available labeled data and NeuroVault maps linked to publications remain quite limited (Peraza et al., 2025). In contrast, CBMA databases currently house the most extensive collection of neuroscience papers with associated peak activation coordinates, covering a diverse range of fMRI tasks and domains. As a result, a CBMA-based functional decoding approach may enable the inference of hidden and more diverse cognitive domains thanks to their extensive coverage of the neuroscientific literature, potentially mitigating unpredictable bias introduced by including solely statistical maps from NeuroVault. Publications in existing CBMA databases, however, are not annotated with an ontology.

Despite recent advances in neuroimaging meta-analysis, accurately annotating and consistently synthesizing results from thousands of publications remains a significant challenge. While automatic test scraping has enabled reliable extraction of text and brain activation coordinates, the representation of text for meta-analytic methods, which is needed for functional decoding, has used basic natural language processing (NLP) techniques, such as word counts and term-frequency inverse document frequency (TF-IDF) (Dockès et al., 2020; Rubin et al., 2016; Salo et al., 2023; Yarkoni et al., 2011). These metrics are limited because they rely on a bag-of-words approach. TF-IDF provides only a sparse view of word representation, fails to capture semantics, and overlooks intrinsic relationships between different words within the same conceptual family. Additionally, the vocabulary representing the text is confined to words present in the neuroscience corpora, restricting the inclusion of external terms and, more importantly, preventing the use of phrases, complete sentences, and term's definitions. The recent emergence of large language models (LLMs) has revolutionized the field of NLP (Liu et al., 2023; Vaswani et al., 2017). LLMs have demonstrated immense capabilities across various tasks, including text and code generation, translation, and summarization (Gunasekar et al., 2023; Srivastava et al., 2023). In particular, BrainGPT, an LLM fine-tuned on neuroscience literature, has exhibited remarkable abilities by surpassing experts in predicting neuroscience findings (Luo et al., 2024). Notably, BrainGPT outperformed neuroscientists in predicting experimental outcomes from novel studies, demonstrating its vast capability to integrate and synthesize information from past neuroscience publications. LLMs may significantly improve neuroimaging meta-analysis by enhancing text encoding and the synthesis of results from various articles. The semantic representation provided by LLM embeddings can capture the relationships between words based on contextual usage, enhancing the domain-specific contextual meaning of words and overcoming the limitations of basic NLP metrics. LLM can naturally enable the use of external terms and their definition to annotate the text of a publication, for instance, concepts from an established ontology like Cognitive Atlas.

LLM has also played a significant role in advancing the link between semantic meaning and images. In the context of brain imaging, Text2Brain has introduced a novel approach to large-scale neuroimaging meta-analysis by synthesizing peak activation coordinates from open-ended text queries (Ngo et al., 2021). Text2Brain is a neural network that learns the relationships between images and text, leveraging LLMs to estimate the semantic representation of the publication's text and utilizing 3D convolutional neural networks (CNN) to reconstruct brain maps. This approach has addressed the main limitations of traditional large-scale meta-analysis methods on Neurosynth (Yarkoni et al., 2011) and NeuroQuery (Dockès et al., 2020) regarding out-of-sample terms and their semantic representation. However, Text2Brain is less effective when processing large input texts (e.g., the entire corpus of an article), which reduces the association between text and brain images (Meudec et al., 2024). A new framework for text-brain association, NeuroConText, was recently proposed; it represents the first contrastive-based model derived from neuroimaging meta-analysis aimed at improving the association between neuroscientific text and brain maps (Meudec et al., 2024). This approach was inspired by CLIP, a contrastive learning framework for matching image-text pairs developed by OpenAI (Radford et al., 2021). NeuroConText also relies on LLM to encode information from texts while employing a simple brain parcellation to convert brain maps into a reduced feature space. Experimental results indicated that NeuroConText significantly outperformed previous approaches like NeuroQuery and Text2Brain in text-brain association tasks, achieving up to three times better recall@10 scores (i.e., matches based on top 10 most probable predictions). This framework demonstrated improved flexibility in handling longer text. However, NeuroConText exhibited limited capabilities in predicting brain activation maps from open-ended text queries. For that specific task, Text2Brain remains the leading coordinate-based meta-analysis framework today. The shared latent space, provided by CLIP, linking text from articles to their corresponding brain maps, not only provides an encoding model for forward inference (i.e., text-to-brain) but might also offer a functional decoding model (i.e., brain-to-text) that could be used for formal quantitative meta-analytic functional decoding. Such a functional decoding framework might enable predictions using custom vocabulary from an established ontology by projecting concepts and their definitions to the CLIP's latent space. However, a CLIP-based meta-analytic functional decoding model has not yet been implemented and validated. Further, no formal attempt to leverage task, concept, and their definitions from a defined ontology has been investigated for meta-analytic functional decoding.

In the current work, we present NiCLIP, a neuroimaging contrastive language-image pretraining model designed for predicting text from brain images. NiCLIP is built upon the NeuroConText model's CLIP architecture while enhancing the text encoding with the assistance of BrainGPT, a pre-trained LLM fine-tuned on neuroscience papers. The overall objective of this work was to improve brain meta-analytic functional decoding using pre-trained CLIP models. To accomplish this, we leveraged 23,000 neuroscientific articles with full text and reported brain activation coordinates from PubMed Central. First, we evaluated the effect of using the fine-tuned LLM on the CLIP models by comparing the two currently available BrainGPT versions (i.e., BrainGPT-7B-v0.1 and BrainGPT-7B-v0.2) with their corresponding pre-trained base LLMs (i.e., Llama-2-7b-chat-hf and Mistral-7B-v0.1, respectively), using the abstracts and the full text of the neuroimaging articles. Second, we assessed the functional decoding performance to predict cognitive tasks and concepts from task-fMRI group-average maps derived from

the Human Connectome Project (HCP) data release (Barch et al., 2013; Smith et al., 2013; Uğurbil et al., 2013; Van Essen et al., 2013, 2012), using a vocabulary from the Cognitive Atlas database (Poldrack et al., 2011). We explored the extent of NiCLIP's capability in predicting subject-level activation maps. Third, NiCLIP was tested on six anatomical brain regions defined by meta-analytic-based parcellations. The baseline models for evaluation included the Neurosynth decoder or correlation decoder and the GC-LDA decoder. Finally, we examined the impact that the LLM, the model-trained section of the articles, and cognitive ontology on the final prediction of the functional decoder. Collectively, this work presents the first LLM-powered CLIP model for formal reverse inference of brain activation maps.

2. Results

2.1. Overview

Figure 1 provides an overview of our general framework for preparing data to train the text-to-brain model (Fig. 1A) and performing functional decoding of a brain activation map (Fig. 1B). To train the CLIP model, we required text and image data. First, we searched the PubMed Central open-source collections for fMRI articles. Second, we extracted the text and activation coordinates from the downloaded articles using Pubget (Dockès et al., 2024). Third, we identified text and brain activation embeddings from the text and coordinates for each article, respectively. For the text features, we used a pre-trained LLM. Meanwhile, the brain activation features were obtained by transforming the coordinates into a modeled activation map using MKDA (Wager et al., 2007) methods and reducing the image dimensionality with a continuous brain parcellation defined by the DiFuMo 512 regions atlas (Dadi et al., 2020). Finally, we entered the normalized text and brain embeddings into the CLIP model to learn a shared latent space for the text and image encoders. To make predictions based on brain activation, we needed a cognitive ontology consisting of fMRI tasks with their respective definitions, along with concepts that are evoked by the tasks and their high-level domains. We first obtained a joint embedding of the task name and definition using a pre-trained LLM. The input brain activation embedding was determined using the DiFuMo 512 atlas. Then, the cognitive ontology and input activation embeddings were passed through the text and image encoders from the pre-trained CLIP model. Next, we determined the cosine similarity between the text and image embeddings in the shared latent space. This similarity was used to calculate the posterior probability of a task given an activation given the shared latent space.

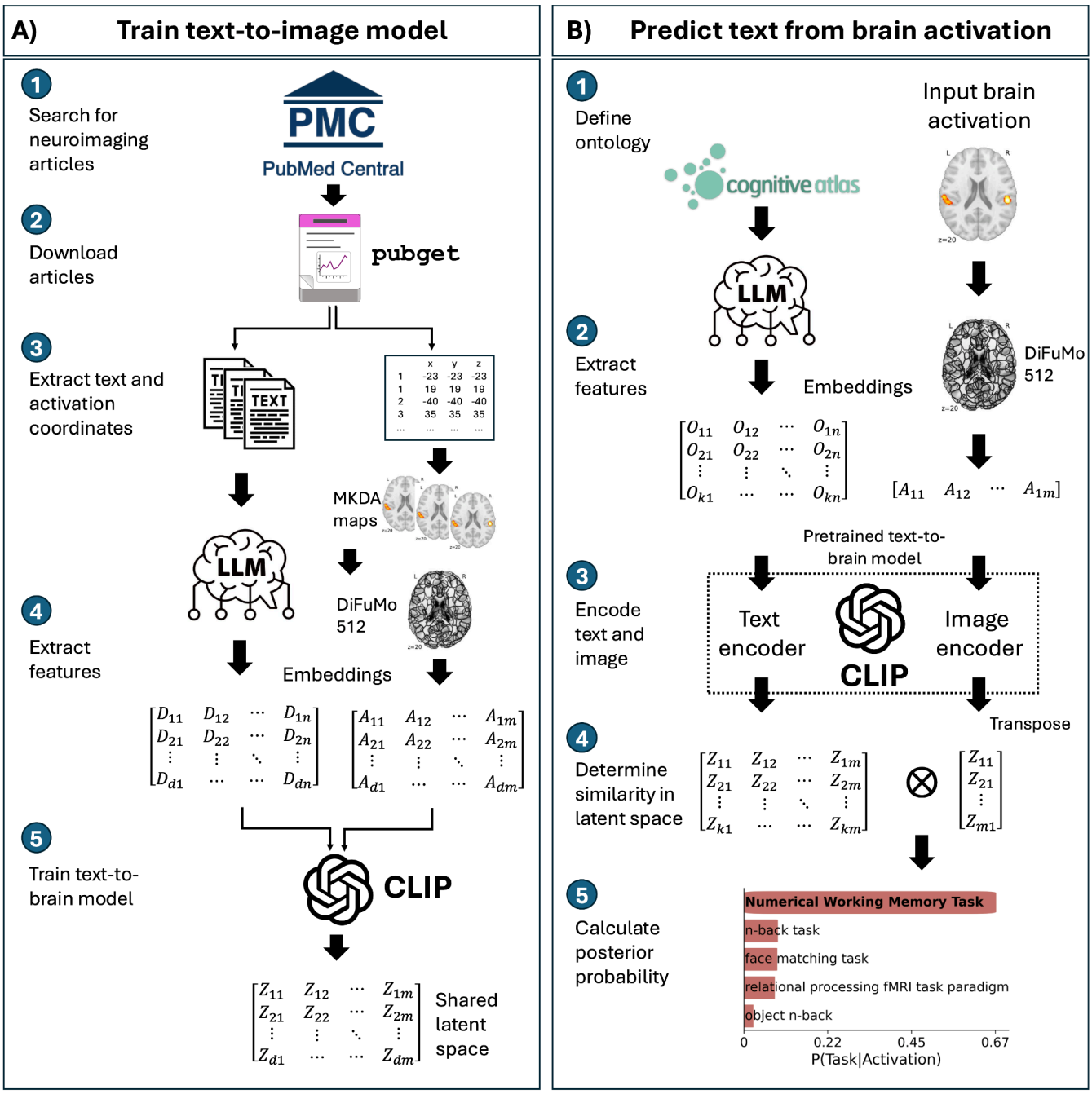

Figure 1. Overview of the framework for training the text-to-brain model and decoding brain activation maps. (A) The text-to-brain CLIP model was trained using text and brain activation coordinates sourced from a collection of fMRI articles downloaded from PubMed Central. Pubget was employed to download and preprocess the articles in a standardized format. Text embeddings were determined using a pre-trained LLM. In contrast, image embeddings were obtained by first calculating an MKDA-modeled activation brain map, and second applying a continuous brain parcellation defined by the DiFuMo 512 atlas. **(B)** The brain decoding model relies on a cognitive ontology to predict text from input brain activation. The embeddings of task names along with their definitions were extracted

using an LLM transformer, while image features were determined using the DiFuMo brain parcellation. The text and image embeddings were processed through the pre-trained text and image encoders from CLIP, yielding new embeddings in a shared latent space. The posterior probability representing the predicted task from the input activation was calculated based on the similarity between the text and image embeddings in the shared latent space.

2.2. CLIP model evaluation for text-to-brain association

We assessed the CLIP model's ability to match text to brain images using Recall@k and Mix&Match (see Methods sections for definitions). Four different large language models (i.e., BrainGPT-7B-v0.2, Mistral-7B-v0.1, BrainGPT-7B-v0.1, and Llama-2-7b-chat-hf) were evaluated for text features extraction across two document sections (i.e., abstract and body). Training the CLIP model with the body portion of the articles and embeddings extracted with BrainGPT-7B-v0.2 demonstrated the strongest performance across all metrics (Recall@10: 33.56±4.12, Recall@100: 71.95±4.18, Mix&Match: 90.04±1.66), closely followed by its foundational LLM, Mistral-7B-v0.1 (Recall@10: 33.36±4.05, Recall@100: 71.78±4.22, Mix&Match: 89.97±1.70). BrainGPT-7B-v0.1 and Llama-2-7b-chat-hf showed comparable but slightly lower performance, with Llama-2-7b-chat-hf marginally outperforming BrainGPT-7B-v0.1 on Recall@k and Mix&Match. When using only the abstract sections to train the CLIP model, all LLMs demonstrated lower performance than when training with the full body of the articles. BrainGPT-7B-v0.2 again achieved the highest scores (Recall@10: 24.01±1.84, Recall@100: 61.98±4.28, Mix&Match: 85.82±1.84), with BrainGPT-7B-v0.1 performing slightly better than Mistral and Llama models across all three metrics. Llama-2-7b-chat-hf consistently showed the lowest performance in the abstract section evaluation. These results indicate that BrainGPT-7B-v0.2 provided the most effective text-to-brain association among the tested LLMs, and the association performance was generally stronger with the full body of the articles compared to the abstracts.

Table 1
Text-to-brain association performance comparison of LLMs across article sections

CLIP model setting		Metrics		
Section	LLM	Recall@10	Recall@100	Mix&Match
Body	BrainGPT-7B-v0.2	**33.56 ± 4.12**	**71.95 ± 4.18**	**90.04 ± 1.66**
	Mistral-7B-v0.1	33.36 ± 4.05	71.78 ± 4.22	89.97 ± 1.70
	BrainGPT-7B-v0.1	32.30 ± 4.16	70.66 ± 4.81	89.65 ± 1.83
	Llama-2-7b-chat-hf	32.50 ± 4.08	71.08 ± 4.81	89.72 ± 1.84
Abstract	BrainGPT-7B-v0.2	24.01 ± 1.84	61.98 ± 4.28	85.82 ± 1.84
	Mistral-7B-v0.1	23.36 ± 2.12	61.37 ± 4.27	85.57 ± 1.88
	BrainGPT-7B-v0.1	23.51 ± 1.98	61.94 ± 4.09	85.73 ± 1.79
	Llama-2-7b-chat-hf	22.46 ± 2.07	60.69 ± 4.13	85.32 ± 1.87

Note: Values represent mean ± standard deviation across cross-validation sets. The best-performing value for each metric is highlighted in bold font.

2.3. NiCLIP model settings evaluation

Next, we used NiCLIP, a contrastive language-image pretraining model designed for predicting tasks, concepts, and domains from brain maps. Here, a "domain" refers to a high-level concept, as defined by the concept categories in Cognitive Atlas (https://www.cognitiveatlas.org/concepts/categories/all). NiCLIP relied on a cognitive ontology that comprises a collection of fMRI tasks with their definitions, along with the collection of concepts engaged in tasks and the major domains to which the concepts belong. The prediction of the task was based exclusively on the semantic content of the task name and its definition. In contrast, the predictions for concepts and domains were determined by the task-to-concept and concept-to-domain mappings established in the ontology.

We evaluated how different combinations of parameters influence the final prediction of NiCLIP (Table 2), across seven major domains of high interest in the neuroscience field, drawn from the HCP task fMRI dataset. These domains included tasks such as emotion, gambling, language, motor, relational, social, and working memory. We compared the results to two baseline decoding models (i.e., Neurosynth and GC-LDA decoder). Regarding the CLIP model settings, also presented in the first analysis in Table 1, we examined two sections of the articles (i.e., abstract and body) and four different pre-trained LLMs (i.e., BrainGPT-7B-v0.2, Mistral-7B-v0.1, BrainGPT-7B-v0.1, and Llama-2-7b-chat-hf). Next, for the vocabulary setting of the NiCLIP model, we utilized two different ontologies based on the Cognitive Atlas. Initially, we leveraged all tasks from the Cognitive Atlas database and their concept-to-task and concept-to-domain mappings. Then, we also tested a reduced version of the Cognitive Atlas, containing only 10% of the total tasks, with a more robust mapping from concept to task and the original mapping from concept to domain. Additionally, we employed two definitions for the embedding for NiCLIP, consisting of (1) embedding calculated using only the name of the task, and (2) embedding calculated using the name of the task along with its definition. The models with different parameter combinations were compared using Recall@4 for tasks and concepts, and Recall@2 for domains. In decoding, recall@k represents the likelihood that the model ranked a ground truth within the top k predictions. The analysis was performed on the seven group-level activation maps from the HCP dataset presented in Fig. 2.

The strongest performance across all categories was achieved by utilizing the complete text of the articles to train CLIP, BrainGPT-7B-v0.2 with the reduced Cognitive Atlas ontology, along with their definitions and task names (Task: 62.86%, Concept: 43.57%, Domain: 90.48% at Recall@2), significantly surpassing both baseline models with over a 40% improvement in recall scores. This configuration consistently yielded superior results across all tested LLMs, with BrainGPT-7B-v0.1 achieving the second-best performance (Task: 55%, Concept: 34.29%, Process: 78.57%). The reduced and enhanced versions of Cognitive Atlas consistently outperformed the original Cognitive Atlas ontology across all models, with performance further improved when using both the task name and definition rather than the name alone. Notably, some configurations with the full Cognitive Atlas ontology exhibited minimal or zero recall, particularly with Mistral-7B-v0.1. Baseline models (Neurosynth and GC-LDA) demonstrated limited effectiveness, with maximum Recall@4 scores of 20.71% for tasks, substantially lower than our NiCLIP model implementations. For abstract sections, we observed generally lower

performance compared to body sections, but with similar patterns. BrainGPT-7B-v0.1 with CogAt Reduced ontology and Name + Definition vocabulary achieved the highest concept recall (40.16%), while Llama-2-7b-chat-hf excelled in task predictions recall (44.29%) with the same configuration. The predictions of domains consistently showed higher recall rates than tasks and concepts across all models and configurations, with BrainGPT-7B-v0.2 also achieving the best performance (78.57%) when trained with the abstract section, utilizing the reduced ontology with both task name only and name plus definition.

Table 2
NiCLIP model settings evaluation and comparison against baseline models

CLIP Model settings		Vocabulary settings		Baseline Models		Our Model NiCLIP		
				Neurosynth	GC-LDA			
				Task	Task	Task	Concept	Domain
Section	LLM	Ontology	Embedding	Recall@4	Recall@4	Recall@4	Recall@4	Recall@2
Body	BrainGPT-7B-v0.2	CogAt. Reduced	Name + Def.	-	-	**62.86**	**43.57**	**90.48**
			Name	20.71	20.71	62.14	19.21	78.57
		CogAt	Name + Def.	-	-	28.57	13.97	57.14
			Name	02.86	17.14	10.71	7.62	76.19
	Mistral-7B-v0.1	CogAt. Reduced	Name + Def.	-	-	34.20	10.48	59.52
			Name	20.71	20.71	21.43	9.52	45.24
		CogAt	Name + Def.	-	-	14.29	2.86	40.48
			Name	02.86	17.14	0	0	38.10
	BrainGPT-7B-v0.1	CogAt. Reduced	Name + Def.	-	-	55.	34.29	78.57
			Name	20.71	20.71	41.43	26.43	85.71
		CogAt	Name + Def.	-	-	10.00	4.76	26.19
			Name	06.43	17.14	09.29	4.76	57.14
	Llama-2-7b-chat-hf	CogAt. Reduced	Name + Def.	-	-	49.29	26.27	47.62
			Name	20.71	20.71	13.57	15.16	28.57
		CogAt	Name + Def.	-	-	06.43	11.91	38.10
			Name	17.14	17.14	06.43	7.14	30.95
Abstract	BrainGPT-7B-v0.2	CogAt. Reduced	Name + Def.	-	-	36.43	17.14	78.57
			Name	20.71	20.71	33.57	25.24	78.57
		CogAt	Name + Def.	-	-	20.00	7.62	57.14
			Name	20.71	27.86	20.71	0	54.76
	Mistral-7B-v0.1	CogAt. Reduced	Name + Def.	-	-	37.14	20.00	59.52
			Name	20.71	27.86	20.00	15.95	42.86
		CogAt	Name + Def.	-	-	05.71	11.19	66.67
			Name	20.71	20.71	0	0	30.95
	BrainGPT-7B-v0.1	CogAt. Reduced	Name + Def.	-	-	37.86	40.16	64.29
			Name	20.71	20.71	37.14	28.33	71.43
		CogAt	Name + Def.	-	-	24.29	4.76	52.38
			Name	17.14	27.86	23.57	0	45.24
	Llama-2-7b-chat-hf	CogAt. Reduced	Name + Def.	-	-	44.29	15.95	66.67
			Name	20.71	20.71	26.43	21.67	47.62
		CogAt	Name + Def.	-	-	27.86	12.78	30.95
			Name	20.71	20.71	0	09.92	16.67

Note: The best-performing value is highlighted in bold

2.4. Evaluating NiCLIP for predicting text from group-level brain maps

Figure 2 presents the results of the top 5 predictions sorted by the magnitude of the posterior probabilities of task (P(T|A)), concept (P(C|A)), and domain (P(D|A)). For the following experiments, we leveraged the best-performing CLIP model from the previous analysis. Additional comparisons between the NiCLIP task predictions and the baseline models are provided in the supplementary information (Fig. S2).

First, we decoded the HCP emotion processing task, targeting the contrast "Face vs. Shape." This activation map was most strongly associated with a face identification task (27.4%), followed by functional localizer fMRI tasks (19.9%) and Pavlovian conditioning tasks (18.2%). Standard localizer and social cognition paradigms showed weaker associations (6.1% and 5.8%, respectively). For the concept prediction, we found the strongest associations with response selection (28.6%), visual object recognition (28.5%), and face identification/recognition (27.4% each). Consequently, domains such as Learning and Memory (41.6%) and Perception (36.7%) were strongest, followed by Action (33.3%). The Emotion domain itself was the fourth strongest prediction (15.7%).

Next, we decoded the HCP gambling task, which elicits activation of reward-related regions. Specifically, we selected the “Reward vs Baseline” contrast. NiCLIP predicted that this map was primarily associated with numerical working memory tasks (38.2%), followed by relational processing paradigms (22.1%) and spatial localizer paradigms (16.5%), with N-back tasks showing a moderate association (13.0%). As a result, we estimated the strongest association with concepts such as working memory (47.0%), numerical comparison (38.3%), and mathematical reasoning (38.2%). Response inhibition and relational comparison showed notable associations (22.1% each). This prediction yielded the highest association with domains of Executive Cognitive Control (65.0%) and Reasoning and Decision-Making (62.2%).

Following, we decoded the HCP language processing task, which consisted of blocks of story and math tasks. This task was designed to elicit activation of numerosity-related processing areas in the parietal cortex. The target contrast, “Story vs. Math,” was estimated to be more likely associated with social localizer paradigms (18.9%) and sentence/nonword language localizers (17.0%). Passive viewing (12.3%) and social cognition paradigms (12.0%) also showed moderate associations. The most probable concept associated with the target language maps was visual object recognition, with a probability of 27.4%, followed by reading (24.6%) and auditory sentence recognition (20.5%). Animacy detection and story comprehension showed identical associations (18.9% each). As anticipated, the most likely predicted domains were Perception (49.6%) and Language (44.7%), with Learning and Memory showing a lower probability (19.1%).

Subsequently, we decoded the contrast representing the average of all motor movement blocks against the baseline in the HCP task. As expected, NiCLIP predicted the strongest association with motor fMRI task paradigms (48.3%), followed by overt word repetition (27.5%). Other tasks showed substantially weaker associations, with a probability smaller than 6%. Equally expected, we discovered concepts such as movement, motor control, and motor learning, which all showed identical strong associations

(48.3% each), followed by word repetition and word generation (27.5% each). This demonstrates a clear and specific association between motor activations and motor-related concepts. Lastly, the distribution of the probability of concepts revealed Action as the most likely domain at 73.8%, followed by Language (56.6%) and Learning and Memory (48.7%).

In addition, for the HCP relational processing task, we used the "Relational vs. Baseline" contrast as our target. NiCLIP predicted the same relational processing fMRI task paradigms with a probability of 54.5% as the top associated task, followed by n-back tasks (14.9%) and spatial localizer paradigms (11.9%). The predicted concepts included response inhibition, relational comparison, and visual pattern recognition with a probability of 54.5% each, along with working memory (27.6%) and interference control (14.9%). Regarding the predicted domain, we found Executive Cognitive Control as our top prediction with a prominent 72.0%, followed by Action (22.8%) and Reasoning and Decision Making (20.4%).

We also decoded the HCP social cognition (TOM: theory of mind) task, targeting the contrast "TOM vs. Random." The strongest association for the target contrast was with Social localizer fMRI task paradigms (39.9%), followed by functional localizer fMRI tasks (29.1%) and social cognition paradigms (17.0%). In terms of the predicted concept, we found auditory sentence recognition had the most prominent association (58.0%), followed by animacy detection and story comprehension (39.9% each), and response selection (33.5%). The top predicted domains were Action (34.1%), Perception (22.7%), and Reasoning and Decision Making (21.4%). Interestingly, the Social Function domain itself was only the fourth strongest prediction (17.7%).

Finally, for the HCP working memory task, we decoded the contrast "2-Back vs. 0-Back." As expected, NiCLIP predicted that the Numerical Working Memory Tasks (67.2%) would have the highest probability, trailed by the n-back task (9.1%) and face-matching task (8.9%). For the concept predictions, we identified working memory with the strongest association (71.0%), followed by numerical comparison (67.3%) and mathematical reasoning (67.2%). Interference resolution and interference control showed weaker associations (9.1% each). As anticipated, the top predicted domains were Reasoning and Decision Making (89.5%) and Executive Cognitive Control (75.8%), with Learning and Memory having a considerably smaller probability (10.2%).

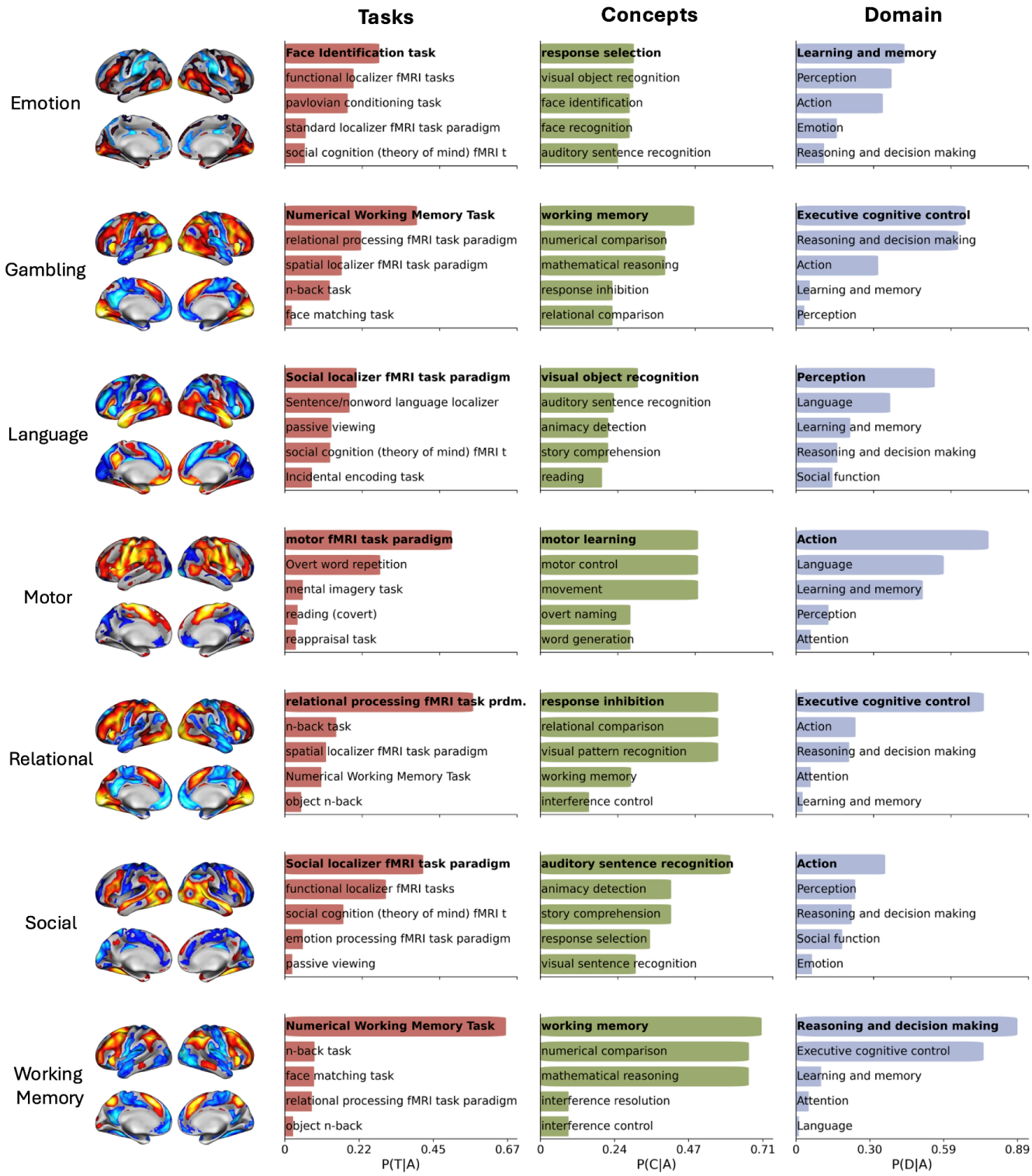

Figure 2. NiCLIP predicts tasks, concepts, and domains from brain activation patterns based on group-level maps. This analysis provides prediction probabilities across seven major cognitive

domains from the Human Connectome Project (HCP) task fMRI dataset. For each domain (Emotion, Gambling, Language, Motor, Relational, Social, and Working Memory), we present three types of predictions: the probability of a task given an activation pattern (P(T|A)), the probability of a concept given an activation (P(C|A)), and the probability of a domain given an activation (P(D|A)). Each prediction is illustrated with horizontal bars representing prediction strength, showcasing the top five predictions for each category.

2.5. Evaluating NiCLIP for predicting text from subject-level brain maps

Following the prediction of group-level activation maps, we explored whether NiCLIP can identify the underlying tasks, concepts, and domains in subject-level images. In practice, researchers typically report activation foci from group-level images of significant findings in scientific papers. Consequently, our model is capable of producing robust predictions for population-level average activation maps, as illustrated in Fig. 2. In contrast, subject-level images exhibit high variability and noise, and individual subject differences are known to result in distinct brain activation patterns within the same task and condition. We anticipated that these factors could potentially interfere with NiCLIP's predictions. For the next experiment, we also utilized the best-performing CLIP model along with a condensed version of the Cognitive Atlas as our cognitive ontology. The embeddings for the task names and their definitions were also determined using the same LLM employed for the CLIP model. We conducted a qualitative evaluation of the NiCLIP predictions on the seven major domains from the HCP tasks fMRI dataset. These domains included tasks along with their respective contrasts of interest, including emotion, gambling, language, motor, relational, social, and working memory. The subject-level images for each domain were sourced from the Individual Brain Charting (IBC) dataset (Pinho et al., 2018). We focused on one subject for our experiment. Figure 3 illustrates the results of the top 5 predictions sorted by the magnitude of the posterior probabilities (i.e., P(T|A), P(C|A), and P(D|A)). Additional comparisons between the NiCLIP task predictions and the baseline models are provided in the supplementary information (Fig. S3).

Overall, functional decoding on subject-level brain activation maps was not only less effective than the results on group-level images, but correct findings were rarely ranked among the top five predictions for most of the target brain images. For example, in emotion processing, the top predicted tasks were not related to emotion at all. The predictions were predominantly language-related, such as abstract/concrete judgment for bilinguals (51.2%), followed by contextual semantic priming and overt word repetition tasks with significantly lower probabilities of 8.8% and 5.6%, respectively. Consequently, the concepts were primarily language-related, with word comprehension showing the highest probability at 57.1%, followed closely by semantic categorization (53.1%) and semantic knowledge (53%). Such predictions indicated language as the most likely domain overall, with a notably high probability of 83.9%. Similarly, for gambling, none of the predicted tasks were oriented toward gambling. The top prediction was the object-one-back task (20.4%), followed by the spatial localizer task (14%). The leading predicted concepts revolved around working memory and other visual-related ideas, with no clear distinction among them. Four different domains were ranked very similarly. Continuing this trend, language and relational tasks showed no relevant task concepts within

the top five predictions. The most likely task corresponded to the aberrant behavior checklist - community, which is a behavioral task not necessarily aligned with language or relational processing. For this reason, the predicted concepts included autobiographical memory, self-monitoring, and monitoring, with executive cognitive control identified as the most likely domain. For motor, social, and working memory tasks, NiCLIP was able to predict the correct tasks within the top five predictions. For instance, the motor fMRI task paradigm was ranked first for motor with a probability of 32.9%. For working memory, we identified numerical working memory (10.5%) as the second most likely task. In social tasks, performance was somewhat reduced, and the social localizer task ranked fifth with a probability of 7.4%. However, the remainder of the top five predicted tasks largely comprised spurious findings that did not accurately reflect the target domain. Consequently, the majority of predicted concepts and domains for subject-level maps did not align with the expected results, unlike those found with group-level maps.

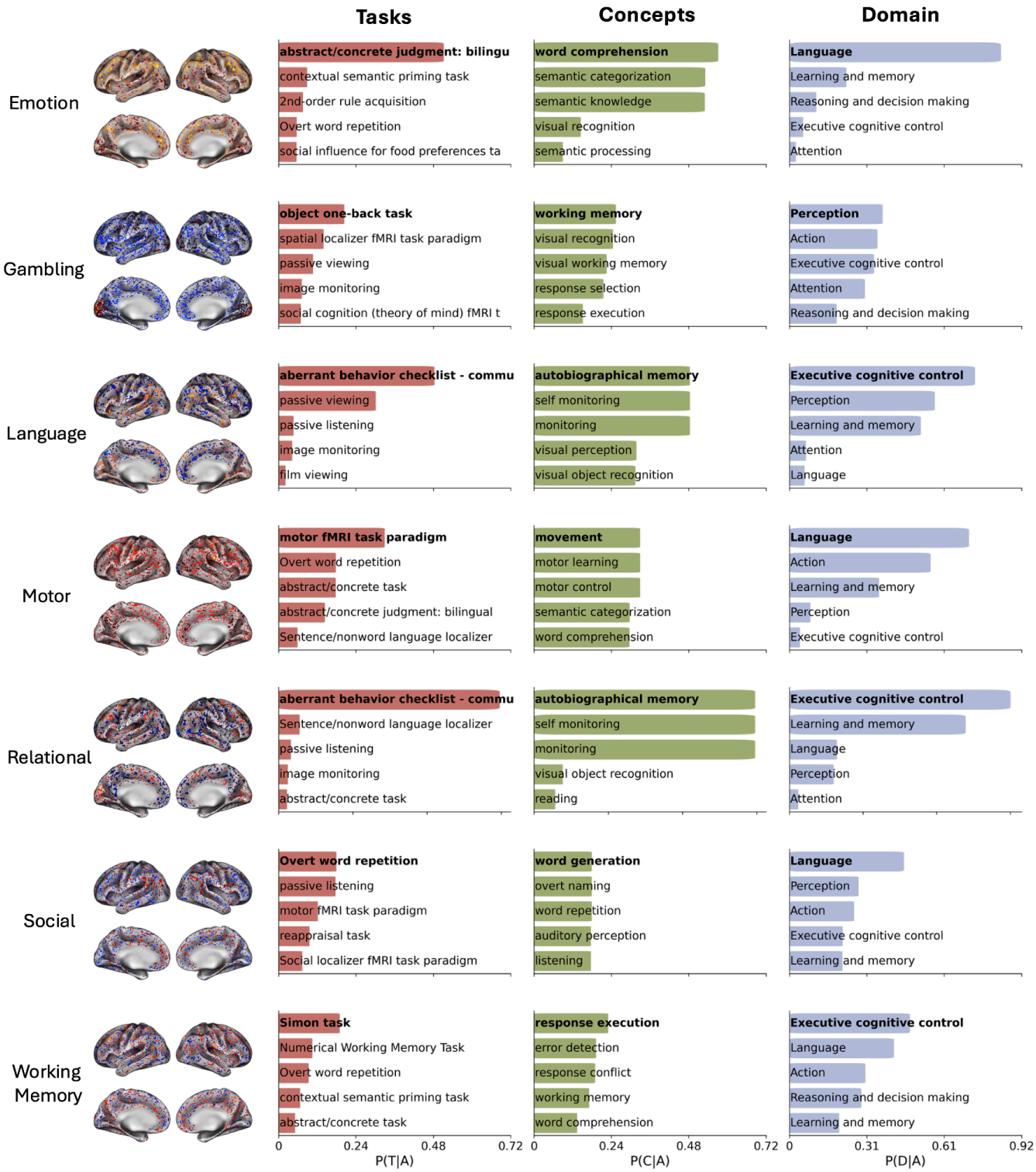

Figure 3. NiCLIP for predicting tasks, concepts, and domains from brain activation patterns of subject-level maps. Comprehensive analysis of prediction probabilities across seven major cognitive

domains from the IBC task fMRI dataset. For each domain (Emotion, Gambling, Language, Motor, Relational, Social, and Working Memory), we display three types of predictions: the probability of a task given an activation pattern (P(T|A)), the probability of a concept given an activation (P(C|A)), and the probability of a domain given an activation (P(D|A)). Each prediction is visualized with horizontal bars indicating prediction strength, with the top five predictions shown for each category.

2.6. Evaluating NiCLIP for predicting text from regions of interest maps

Next, we present a potential application of reverse inference methods for exploring novel hypotheses and discovering tasks, concepts, and domains associated with specific regions of interest (ROI). For this experiment, we also utilized the best-performing CLIP model, (i.e., BrainGPT-7B-v0.2), which was trained on the complete body of articles. The NiCLIP model relied on a reduced version of the Cognitive Atlas as its cognitive ontology. NiCLIP predictions on ROIs were tested on six highly popular anatomical brain regions, including the amygdala (Bzdok et al., 2013a), hippocampus (Plachti et al., 2019), insula (Chang et al., 2013), striatum (Liu et al., 2020), right temporoparietal junction (rTPJ) (Bzdok et al., 2013b), and ventromedial prefrontal cortex (vmPFC) (Chase et al., 2020). Figure 4 presents the results of predicting tasks, concepts, and domains across the six ROIs. Unlike the previous predictions on dense maps from Fig. 2, the top predicted task here resulted in a more distinct probability relative to the second-highest task. Additional comparisons between the NiCLIP task predictions and the baseline models are provided in the supplementary information (Fig. S3).

As anticipated, the highest predicted task for the amygdala ROI was the emotion processing task (67.7%), while other tasks showed a significantly weaker association (i.e., less than 5%). The top three predicted concepts were emotion, response selection, and feature comparison, each with nearly identical probability values of 69%, 67.9%, and 67.7%, respectively. Similarly, NiCLIP identified Action (70.6%), Emotion (69.6%), and Reasoning and Decision Making (68.7%) as the top three most likely domains linked to the amygdala. The hippocampus exhibited the strongest connections to incidental encoding (22.7%), emotion processing tasks (20.9%), and motor tasks (17.2%). The leading predicted concepts were predominantly memory-related (i.e., memory, memory trace, and memory acquisition), with similar probability values around 23%. Thus, the primary predicted domain was Learning and Memory, with a probability of nearly 60%, followed by Action and Reasoning and Decision Making (48.1%). As expected, the insula was associated with pain-related and audio-visual integration tasks, such as pain monitoring/discrimination (34.9%), reading (18.8%), and phasic pain (5.3%) tasks. The predicted concepts also aligned with the pain and visual families, including pain (38.9%), pain sensation (38.3%), pain habituation (38.3%), visual object recognition (22.1%), and reading (21.9%). In terms of domains, the insula was most likely associated with Emotion (76.9%), while other probable domains included Language (46.7%), Perception (45%), and Learning and Memory (22.9%). The striatum was most directly linked to the semantic task (52.1%), followed by social judgment (14%), and a pair of response inhibition tasks, namely stop signal and go/no-go tasks, with lower probabilities of 8.7% and 5.1%, respectively. The top predicted concepts for the striatum were language-related, specifically language, semantic categorization, and semantic information, each with identical probabilities of 52.2%. Consequently, the domains with the highest likelihood were Language (53.6%),

followed closely by Executive Control, Reasoning, and Decision Making (32.1% each). As foreseen, the rTPJ was singularly associated with social cognition (theory of mind) tasks, with a probability of 98.5%. These tasks corresponded to concepts such as social cognition, animacy perception, and animacy decision, each with equal probabilities (98.5%). Consequently, these three concepts were mapped to three distinct domains: Perception (98.5%), Reasoning and Decision Making, and Social Function (98.5%). Finally, the vmPFC region was predicted to be linked to a 2-stage decision task (45.7%), followed by a probabilistic classification task with a lower probability (11.2%). In terms of concepts, reward processing (47.2%), decision-making under uncertainty (45.9%), and reward valuation (45.7%) were predicted to be strongly associated with the target ROI. The most probable domains were Learning and Memory (63%), closely followed by Reasoning and Decision Making (50.4%).

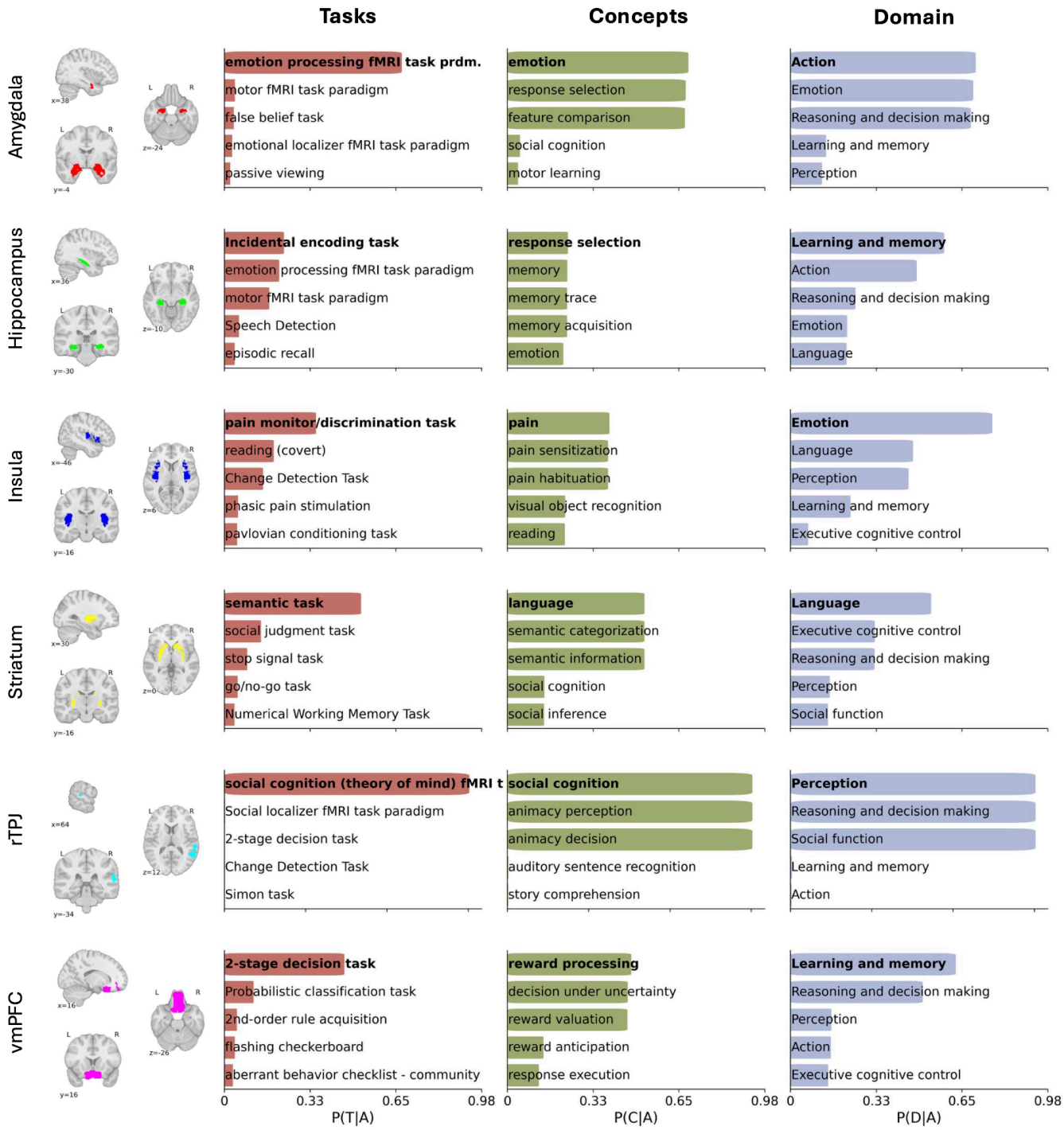

Figure 4. NiCLIP for predicting tasks, concepts, and domains from brain ROIs. We conducted a comprehensive analysis of prediction probabilities across six different ROIs. For each ROI (amygdala, hippocampus, insula, striatum, rTPJ, vmPFC), we display three types of predictions: the probability of a task given an activation pattern (P(T | A)), the probability of a concept given an activation (P(C | A)),

and the probability of a domain given an activation (P(D|A)). Each prediction is visualized with horizontal bars indicating prediction strength, with the top five predictions shown for each category.

3. Discussion

The topic of formal reverse inference on brain activation maps has remained an open question in the neuroscience community for many years. Although several attempts have been made to address this gap, mostly utilizing emerging meta-analytic approaches and databases, multiple limitations persist in meta-analytic methods for consolidating existing data for functional decoding. Additionally, the absence of a reliable cognitive ontology has hindered broader adoption and application. To tackle these current challenges, we present NiCLIP in this work, a contrastive language-image pretraining model designed to predict fMRI tasks, concepts, and domains from brain images. First, we compared four different LLM models and demonstrated that a fine-tuned LLM (e.g., BrainGPT-7B-v0.2) enhances text-to-brain association within the CLIP model architecture. Second, we evaluated the predictions of NiCLIP models on group-level activation maps from the HCP and found overall accurate predictions of tasks, concepts, and domains. However, when NiCLIP was applied to the same task on subject-level images, it produced inconsistent results, with accurate task predictions only for a few tasks, such as motor and working memory, while yielding incorrect predictions for others. Third, we investigated the capabilities of the NiCLIP model for decoding regions of interest. In this experiment, we successfully predicted tasks, concepts, and domains across six different meta-analytic ROIs tested. An extensive evaluation of NiCLIP revealed that it is crucial not only to utilize a fine-tuned LLM and train the CLIP model on the full text of the articles, but also to leverage a curated cognitive ontology with precise task-to-concept and concept-to-domain mappings.

Firstly, we demonstrated CLIP's unique ability to match article text with brain images, as shown in prior studies using the NeuroContext framework. In particular, we proved that a fine-tuned LLM (e.g., BrainGPT-7B-v0.2) improved text-to-brain association within the CLIP model architecture, compared to using a pre-trained, out-of-the-box LLM (e.g., Mistral-7B-v0.1). However, the Mistral-7B-v0.1 model outperformed BrainGPT-7B-v0.1, the fine-tuned version of Llama-2-7b-chat-hf, when utilizing the full text of the articles. Moreover, all tested LLMs with CLIP showed significantly lower performance when trained exclusively on the abstract sections of the articles compared to training with the full body text. Numerous labeled image datasets with task annotations exist and can be used to expand the training set of CLIP models, such as individual-subject images from large data consortia. However, employing the complete text of the publication, as opposed to a shorter text (e.g., the abstract section), yields over a 10% boost in recall@k scores and nearly a 5% improvement in mix&match scores. In summary, these results underscore the importance of utilizing the entire text of the publication for CLIP models, as the additional text provides more distinctive context for the associated brain maps. More importantly, we recommend employing domain-specific LLMs through fine-tuning, as they seem to produce better text-to-image associations than more generic models.

Secondly, we assessed the predictions of the NiCLIP model on group-level activation maps from the HCP database. We focused on seven major domains of high interest in the neuroscience field, including

emotion, gambling, language, motor, relational, social, and working memory. For emotion, we accurately predicted face identification and functional localization as the most likely tasks. In the case of gambling, we did not precisely find some expected tasks, such as the gambling fMRI or monetary incentive delay tasks. However, the top predictions included a numerical task, while the reasoning and decision-making domain was ranked second. Furthermore, the social and language localizer tasks exhibited the highest probability for the language HCP activation map. As anticipated, we identified the motor fMRI task paradigm as the top task associated with motor functions, while movement, motor concepts, and the action domain were rightly predicted with the highest probability. Similarly, for relational tasks, we discovered that the same relational processing fMRI task paradigm had the highest probability, followed by relational and response inhibition concepts, as well as the executive control domain. The social localizer fMRI task paradigm was also the top-ranked task for the social HCP images, with concepts such as auditory sentence recognition, animacy detection, and response comprehension. Lastly, for working memory, almost all top-predicted tasks were correctly linked to the target domain, including the numerical working memory task, followed by the n-back and object n-back tasks. Consequently, the concepts (i.e., working memory, numerical comparison, and mathematical reasoning) accurately reflected the expected results, ultimately yielding reasoning and decision making and executive cognitive control as the top two predicted domains. Taken together, in addition to identifying semantic similarities between tasks and their descriptions from our cognitive ontology, we were also able to observe the concept and domain distribution, providing additional context for our predictions. Overall, these findings validated NiCLIP as a precise and accurate decoding tool for group-level dense activation maps.

However, NiCLIP applied to the same tasks but on subject-level images yielded inconsistent results. We found only a few accurate task predictions for domains, such as motor and working memory. We predicted a motor fMRI task paradigm with the highest score for motor, with movement and motor concepts ranked the top. Yet, the language domain had a higher probability of prediction than action. Similarly, for working memory, the numerical working memory task showed the second-best prediction, while the working memory concept was ranked fourth. Although executive cognitive control had the top predicted domain, reasoning and decision making were predicted in fourth place. For the social HCP activation map, the social localizer fMRI task paradigm was ranked fifth, with no correctly predicted concepts in the top five. None of the correct task names and concepts associated with the remaining target HCP images (i.e., emotion, gambling, language, and relational) were ranked in the top five predictions. For instance, the aberrant behavioral checklist task was predicted with the highest probability for language and relational, while the most likely task for emotion was abstract/concrete judgment. Although discouraging, these results are not surprising. The activation pattern in subject-level images is known to be noisy and susceptible to inter-subject variability. In contrast, the images used for training CLIP consisted of modeled activation maps based on peak coordinates from population-level results. Therefore, for some cases, such as the HCP motor activation map, when peak activation in the somatosensory region is clearer, the model could perform relatively well. Still, in general, NiCLIP is set to underperform in noisy images, especially with activation in regions of low selectivity. Regardless, further evaluation with a more extensive sample of subjects is needed to truly assess the prediction capabilities of NiCLIP. Furthermore, we expect training the CLIP

model with subject-level images could potentially enhance NiCLIP's predictive robustness. By providing the model with sufficient subject-level training data, we can potentially utilize it for future applications on subject-level activation maps to study inter-individual differences. Overall, we currently recommend against using NiCLIP for predictions on individual subject images, but this can be improved in the future through dataset training expansion by including subject-level activation maps from relevant articles.

Thirdly, compared to other tools, NiCLIP not only accurately performs continuous decoding on dense activation maps but also makes predictions on sparse brain images, such as regions of interest. For instance, NiCLIP precisely linked the amygdala to emotion processing fMRI tasks, along with emotion concepts and domains. The hippocampus was predicted to be mostly associated with memory-related tasks, concepts, and domains. Moreover, the insula was associated with the emotion domain and pain-related tasks and concepts, and the striatum was strongly related to tasks, concepts, and domains relevant to language. Similarly, the rTPJ was exclusively connected to social cognition, while the vmPFC was predicted to be activated by decision-making tasks and related to reward and decision concepts. These promising capabilities open the door to potential new applications of reverse inference methods. In principle, researchers could explore novel hypotheses and discover the tasks, concepts, and domains associated with specific anatomical brain regions that have not been studied before. Others can leverage NiCLIP to describe functional domains of regions of interest that appear relevant to a specific population. Such predictions can also be used to contrast the functional associations linked to a region between healthy controls and clinical populations. In summary, the application of reverse inference does not have to be limited to brain activation maps alone. In principle, one could explore the functional domain of any region in the entire brain.

We evaluated how different combinations of parameters influence the final prediction of NiCLIP. In most instances, the decoding model trained on the complete text of the articles performed better than when trained only on the abstract section. Fine-tuned LLMs (i.e., BrainGPT-7B-v0.2 and BrainGPT-7B-v0.1) surpassed their base models (i.e., Mistral-7B-v0.1 and Llama-2-7b-chat-hf). The modified version of Cognitive Atlas also exceeded the original Cognitive Atlas ontology across all models while demonstrating improved performance when utilizing both the task name and definition rather than just the name. The findings illustrated that our NiCLIP model significantly outperforms traditional baseline approaches. The choice of base language model greatly affects performance, with BrainGPT-7B-v0.2 yielding the most promising results when training CLIP with the complete text of the articles. Overall, these results particularly emphasize the importance of cognitive ontology for reverse inference tasks. These parameters consistently resulted in the largest improvement in prediction recall when employing an enhanced version of the cognitive ontology. We believe that developing more standardized and accurate cognitive ontologies is crucial for fully unlocking LLM-powered functional decoders.

We note several limitations in this current work. The most significant weakness of our study pertains to the small size of the training sample. Typically, CLIP models are trained with hundreds of millions of image-text pairs (Radford et al., 2021). In contrast, our CLIP model was trained with just over 20,000 pairs, which currently is the largest sample of neuroscience papers with complete text information and reported activation coordinates. It is worth noting that (1) brain statistic images are much more constrained than naturalistic images, and (2) we employed a representation of the images based on immense quantities of training data, defined by the DiFuMO atlas. These two factors mitigate the smaller volume of training data relative to other CLIP applications, which requires multiple variations of the content of an image and advanced convolutional neural networks to extract image features. The sample size will increase in the future as more articles are published. However, the scientific literature is still growing at a pace that prevents us from achieving a sample size in the millions anytime soon (Bornmann and Mutz, 2015). In principle, we can integrate NeuroVault images into the training set (Gorgolewski et al., 2015). Combining modeled activation maps from peak coordinates with statistical maps has been successfully attempted before (Meudec et al., 2022), yielding improved performance for decoding tasks. We could also incorporate subject-level labeled images from major data consortia, such as ABIDE (Di Martino et al., 2017, 2014), ABCD (Casey et al., 2018), HCP (Barch et al., 2013; Smith et al., 2013; Uğurbil et al., 2013; Van Essen et al., 2013, 2012), ADNI (Jack Jr. et al., 2024), and ENIGMA (Thompson et al., 2014), among others. Collectively, these images could enhance our sample size to the hundreds of thousands. However, we must consider that our CLIP model appears to perform better when the text associated with the images is more extensive. Still, it remains unclear whether increasing the sample size, even with less detailed text information, can mitigate such limitations. More importantly, incorporating subject-level maps and other images from online repositories such as NeuroVault could potentially address another limitation: the lack of reliability of NiCLIP in making predictions on individual subject activation maps. Further analyses are required to evaluate these hypotheses.

Another significant constraint in the current study relates to the extent and accuracy of the cognitive ontology. Here, NiCLIP utilizes a cognitive ontology to define the fMRI tasks and their associations with concepts and domains. Despite the extensive collection of tasks and concepts, the Cognitive Atlas database remains limited. Moreover, as a community-driven ontology, these mappings reflect the opinions of individual researchers and are not necessarily factual (Poldrack et al., 2011). We demonstrated that a reduced representation of the Cognitive Atlas tasks with a more robust and comprehensive mapping of concepts outperforms the original Cognitive Atlas ontology. Following the trend of leveraging LLM for such challenging tasks, we believe that a data-driven ontology derived from the entire neuroscience literature has the potential to enhance existing cognitive ontologies, providing more accurate and standardized task names and definitions, along with improved task-to-concept mappings. Consequently, a data-driven ontology could naturally enhance NiCLIP predictions.

Lastly, the formulation of the reverse inference problem follows Bayes' theorem, where the posterior probability of a task name given an activation is directly proportional to its prior probability. Here, we have chosen to set the prior distribution on tasks based on their prevalence in the literature. However,

this assumption could be suboptimal for rare tasks, and a uniform prior could instead be used. Future work is needed to assess the actual influence and provide recommendations for selecting prior probabilities.

4. Conclusions

Taken together, this study presents NiCLIP, the first LLM-powered CLIP model for formal reverse inference of brain activation maps. We confirmed CLIP's unique ability to match article text to brain images. The results indicate that a fine-tuned LLM provides the most effective text-to-brain association, with generally stronger performance when using the full body of the articles. We validated NiCLIP as a precise decoding tool for group-level dense activation maps; however, our decoding model underperformed on individual subject images. Additionally, NiCLIP demonstrated notable performance with other image types, such as ROI maps. Lastly, we highlighted the importance of improving cognitive ontologies for reverse inference tasks. NiCLIP represents not only as a significant step forward relative to previous methods for describing existing activation maps but, more importantly, as an essential tool for testing novel hypotheses and advancing scientific discovery (Poldrack, 2006).

5. Methods

5.1. Datasets

5.1.1. PubMed neuroimaging papers

We conducted an extensive search for neuroimaging articles in PubMed Central using Pubget (https://neuroquery.github.io/pubget/pubget.html), an open-access Python tool for collecting data for biomedical text mining. We performed a query and retrieved articles that included keywords such as “fMRI” or “functional MRI” in the abstracts (e.g., "(fMRI[Abstract]) OR (functional MRI[Abstract])"). Pubget extracted text, metadata, and stereotactic coordinates from the articles in a standardized format. The data was then converted into a NiMARE database object to facilitate downstream tasks such as term-based and topic-based meta-analysis for the baseline models and to generate images from the extracted activation coordinates. As of February 2025, the PubMed search for neuroimaging articles yielded more than 30,000 papers. A total of 23,865 articles included at least one activation coordinate, a complete abstract, and full-text information.

The text embeddings $M \epsilon \Re^{23,865 \times 4,096}$ were derived using pre-trained LLMs. We compared four different LLM models: BrainGPT models (i.e., BrainGPT-7B-v0.1 and BrainGPT-7B-v0.2) (Luo et al., 2024), along with their corresponding foundational pre-trained LLM (i.e., Llama-2-7b-chat-hf and Mistral-7B-v0.1, respectively) (Jiang et al., 2023; Touvron et al., 2023). Due to the token size limitation of these LLMs, embeddings for long text were generated by averaging the embeddings from smaller chunks of the text. Article-specific brain images were generated from brain activation coordinates using a Multilevel Kernel Density Analysis (MKDA) (Wager et al., 2007). MKDA is a kernel-based method that convolves

each activation coordinate with a binary sphere of a set radius around the peak voxel. The coordinate-specific binary maps (i.e., one map corresponding to each coordinate from the set) were combined into a single modeled activation map by taking the maximum value for each voxel. In this work, we used a sphere with a radius of 10mm. Finally, brain image embeddings $A \epsilon \Re^{23,865 \times 512}$ were extracted from the modeled activation map using a continuous brain parcellation defined by the DiFuMo atlas with 512 regions (Dadi et al., 2020). Text and image embeddings were L2 normalized.

5.1.2. Cognitive Ontology

Cognitive Atlas (Poldrack et al., 2011) (https://www.cognitiveatlas.org/) is an online repository of cumulative knowledge from experienced researchers in psychology, cognitive science, and neuroscience. The repository currently offers two knowledge bases: 912 cognitive concepts and 851 tasks, complete with definitions and properties. The cognitive concepts establish relationships with other concepts and tasks, aiming to create a map between mental processes and brain functions. The task and concept metadata (i.e., names and definitions) were downloaded using the Cognitive Atlas API, while the task, concept, and domain relationships were obtained through NiMARE's 'download_cognitive_atlas' function. Here, a "domain" refers to a high-level concept, as define by the concept categories in Cognitive Atlas (https://www.cognitiveatlas.org/concepts/categories/all). Despite the extensive collection of tasks and concepts, the mapping (i.e., which tasks measure a concept and how concepts are grouped in domains) remains limited. Additionally, as a community-based ontology, these mappings reflect the opinions of individual researchers and may not always be factual. Therefore, we utilized a simplified representation of the Cognitive Atlas task to enhance the mapping of concepts to tasks defined in (Menuet et al., 2022).

5.1.3. Task fMRI datasets

The functional decoding model was assessed using the task-fMRI group-average maps from the HCP S1200 data release (Barch et al., 2013; Smith et al., 2013; Uğurbil et al., 2013; Van Essen et al., 2013, 2012). Specifically, we used the results in volume space reported in collection 457 in NeuroVault (https://neurovault.org/collections/457/). The HCP tasks target seven major domains that sample a diverse set of neural systems of high interest in the neuroscience field. These domains include (1) emotion processing, (2) category specific representations, (3) language processing (semantic and phonological processing), (4) visual, motion, somatosensory, and motor systems, (5) relational processing, (6) social cognition (theory of mind), (7) working memory/cognitive control systems. Additional details regarding these tasks and their available contrasts have been previously published (Barch et al., 2013). Next, the functional decoding model was evaluated using subject-level statistical images, as these data tend to have more variability and noise compared to group-level images. We used data from the Individual Brain Charting (IBC) dataset (Pinho et al., 2018), which contains the same tasks as those in the HCP.

5.1.4. Meta-analytic regions of interest

Next, the functional decoding model was evaluated on specific regions of interest (ROI). We used six highly popular anatomical brain regions, including the amygdala (Bzdok et al., 2013a), hippocampus (Plachti et al., 2019), insula (Chang et al., 2013), striatum (Liu et al., 2020), right temporoparietal junction (rTPJ) (Bzdok et al., 2013b), and ventromedial prefrontal cortex (vmPFC) (Chase et al., 2020). Here, we defined a single ROI for each brain region by combining all the seed regions. Fig. S2 presents the seed regions per ROI as defined by their corresponding meta-analytic-based parcellation analysis.

5.2. CLIP model for article and brain image association

The CLIP model architecture (Fig. 5A) adheres to the identical settings employed in the NeuroConText framework (Meudec et al., 2024). CLIP consists of both a text and an image encoder. The text encoder is characterized by a projection head and two residual heads, while the image encoder comprises three residual heads. The projection head takes the high-dimensional text embedding and projects it into a shared dimensional space to align with the image embedding dimension, which has already been reduced through brain parcellation. The projection head consists of a fully connected projection layer, followed by a GELU activation function, an identity fully connected layer, a dropout rate of 0.5 for regularization, and a normalization layer. Each residual head includes an identity fully connected layer and a GELU activation function, followed by a dropout rate of 0.5 and a normalization layer.

The model's hyperparameters for training were selected from the NeuroConText paper (Meudec et al., 2024), with the following configurations: batch size of 128, learning rate of 5e-4, weight decay of 0.1, and a total of 50 epochs. The contrastive loss function InfoNCE was employed for self-supervised learning during training, following the standard implementation of the CLIP model (Oord et al., 2019). Additionally, we incorporated an early stopping rule to prevent overfitting and avoid unnecessary computation during epochs without performance improvement. We monitored the validation loss across epochs while maintaining a patience window of 10 epochs to accommodate potential performance fluctuations. If the validation loss did not improve within the patience window, training was terminated, and the model weights corresponding to the lowest validation loss were saved. Finally, the model was trained and evaluated using distinct data loaders for training, validation, and testing, implementing a 23-fold cross-validation on the test set and a 22-fold nested cross-validation on the evaluation set. The fold size for the testing and evaluation was 1,000 samples, resulting in approximately 21,865 articles available for training.

5.3. NiCLIP: neuroimaging CLIP model for predicting tasks, concepts, and domains from brain images

The NiCLIP architecture (Fig. 5B) leverages the learned shared latent space from the CLIP model to make predictions on unseen data. Given a target activation map, the goal of the NiCLIP decoder is to predict the most likely task, concept, and domain associated with the activation pattern. NiCLIP relies on a cognitive ontology to define the fMRI tasks and their definitions, as well as their associations with

concepts and domains. Using Bayes' theorem, we can express the probability that a brain activation pattern (A_k) was produced by a certain fMRI task T_i as:

$$P\left(T_i|A_k\right) = \frac{P\left(A_k|T_i\right)P\left(T_i\right)}{\Sigma_j P\left(A_k|T_j\right)P\left(T_j\right)}$$

where $P\left(T_i|A_k\right)$ is the posterior probability, $P\left(A_k|T_i\right)$ is the likelihood of task T_i having an activation pattern A_k, and $P\left(T_i\right)$ is the prior probability of task T_i regardless of the activation pattern.

First, we use the pre-trained CLIP model to estimate the likelihood $P\left(A_k|T_i\right)$. To accomplish that, we use a vocabulary composed of all Cognitive Atlas tasks and their definitions. Initially, the vocabulary was encoded by the same LLM used to train CLIP, and then the embedding for each task was determined as a linear combination of the task name embedding and its definition:

$$Emb(T_i) = \alpha Emb\left(T_{i,name}\right) + (1 - \alpha)Emb\left(T_{i,description}\right)$$

where the parameter α represented the weight of the name embedding relative to its definition. Here, we considered $\alpha = 0.5$.

Next, the vocabulary embeddings were encoded by the text encoding layers from CLIP, projecting the vocabulary into the shared latent space with the images. For the target image, we extracted 512 features using the Difumo atlas (i.e., the same dimension used to train CLIP). We encoded the features with the image encoding layers, projecting the images into the same shared latent space of the vocabulary. Finally, we calculated the cosine similarity between the vocabulary embeddings and the input image embeddings and transformed the similarity into probability, resulting in the likelihood $P\left(A_k|T_i\right)$.

$$P\left(A_k|T\right) = softmax\left(Emb(T) \cdot Emb(A_k)\right)$$

The prior probability $P\left(T_i\right)$ was determined by the task representation in the neuroscience corpora used for training the CLIP model. The task representation was defined by the cosine similarity of all document task pairs DT. To obtain a measure of the global representation, we calculated the mean across all publications, resulting in a similarity score for each task with the corpora. Finally, the prior probability resulted in

$$P\left(T_i\right) = softmax(DT)$$

Now, since we have the posterior probabilities of the task given an activation P(T | A), we can use noisy-OR model to define the probability of a concept (C_j) given an activation A_k as:

$$P\left(C_j|A_k\right) = 1 - \prod_i\left(1 - P\left(T_i^j|A_k\right)\right)$$

where T_i^j represents tasks that measure the concept C_j. Similarly, we can obtain the probability of the domain D_j given an activation A_k as:

$$P\left(D_j|A_k\right) = 1 - \prod_i \left(1 - P\left(C_i^j|A_k\right)\right)$$

where C_i^j represents concepts categorized under the cognitive process D_j.

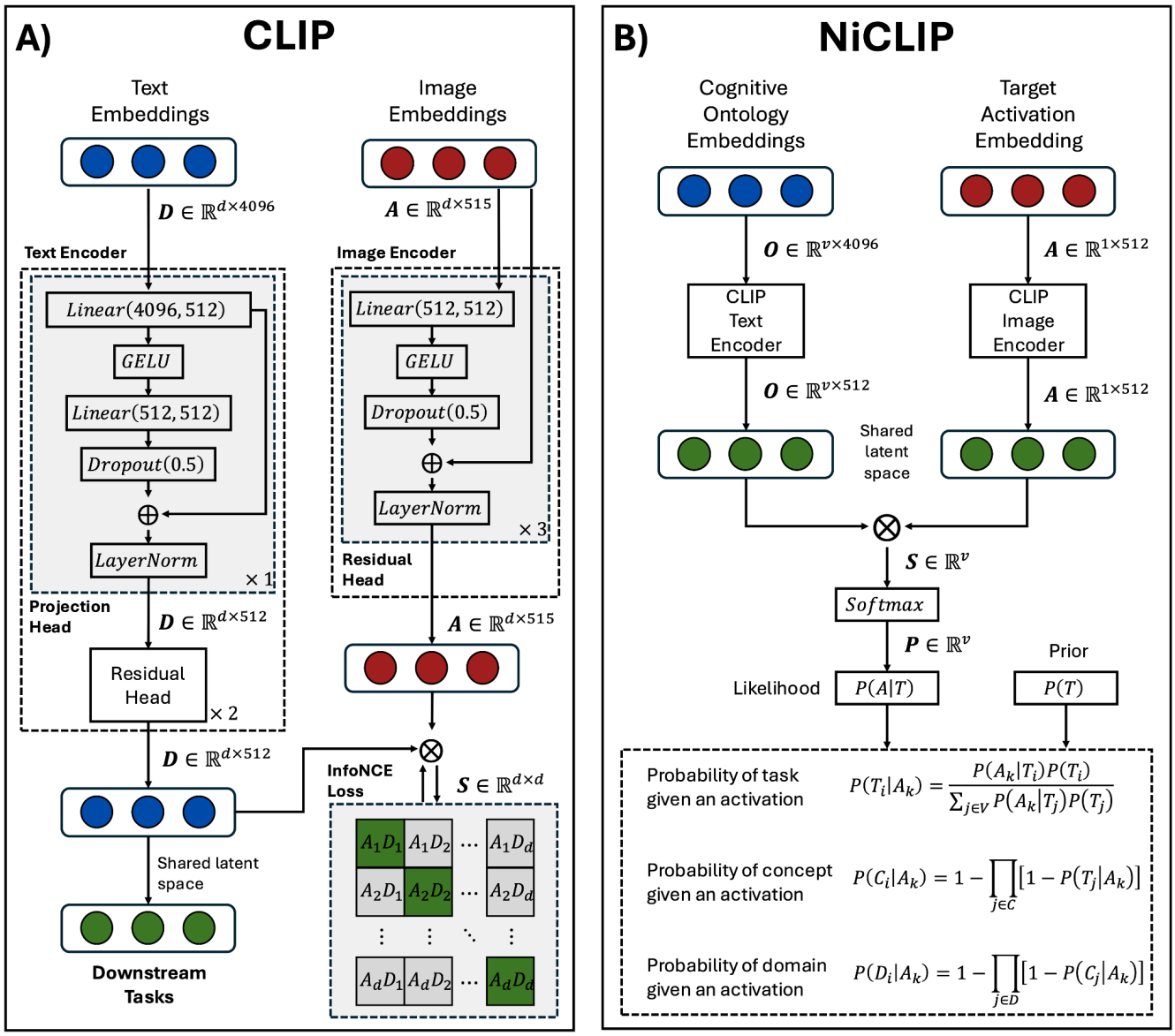

Figure 5. CLIP and NiCLIP model architecture. (A) The architecture of the CLIP model includes a text encoder and an image encoder that transform input embeddings into a shared latent space. The text encoder consists of a projection head and two residual heads, while the image encoder has three

residual heads. The projection head is defined by a linear projection layer, followed by a GELU activation function, a linear layer, a dropout layer, and culminating in a normalization layer. The residual head is made up of a linear identity layer, followed by a GELU activation and a dropout layer, concluding with a normalization layer. The output from the shared latent space is utilized for downstream tasks (e.g., functional decoding), and InfoNCE loss is applied in the latent space for self-supervised learning during training. **(B)** NiCLIP takes advantage of the task name embedding and the extracted features from a target activation map. These embeddings are encoded with the pre-trained CLIP text and image encoders. Cosine similarity is assessed in the shared latent space, and a softmax function converts them to a likelihood P(A | T). Using the prior probability P(T), we compute the posterior probability of a task given an activation P(T | A). Furthermore, the noisy-OR model is employed to determine the probability of a concept P(C | A) and domain P(D | A) given an activation.

5.4. Evaluation metrics

We assessed the CLIP model's ability to match brain images with their corresponding text using Recall@k and Mix&Match (Menuet et al., 2022; Meudec et al., 2024; Mitchell et al., 2008), which are commonly employed in contrastive learning models. Given an activation map, Recall@K quantifies the likelihood of finding its true corresponding text among the top-k ranked by similarity within the entire sample (Meudec et al., 2024). To determine the top-k maps, we first calculate the similarity matrix $S \epsilon \Re^{n \times n}$ between the image and text embeddings in the shared latent space. The identity matrix represents the true positive, as the diagonal element indicates the similarity between an image and its corresponding text. As specified in the InfoNCE loss, the model is trained so that the diagonal element of the similarity matrix approaches 1 while the nondiagonal elements decrease towards zero. The similarity matrix is sorted on a row-by-row basis, and we check whether the diagonal element (i, i) is among the top-k ranked elements. The final scores are computed by averaging across rows:

$$Recall@k = \frac{1}{n} \sum_{i} H(i)$$

where n represents the total number of samples (or rows), and H denotes the indicator function that yields one if the column j=i belongs to the top k sorted columns ($i \epsilon Top_{ki}$) and zero otherwise. Specifically, the best model was determined based on the highest test set Recall@10

Mix&Match measures the likelihood that a given brain map is more similar to its true corresponding text than the other text embeddings in the set (Meudec et al., 2024; Mitchell et al., 2008). In other words, this metric assesses whether the learned embeddings in the shared latent space representations are discriminative, indicating that an item is more similar to itself than to other items.

We also evaluated the sensitivity of the functional decoders, including baselines and NiCLIP, using Recall@K. In this context, the Recall@K value can be interpreted as the likelihood that a decoder ranks the true label among the top-k labels for a given brain activation map (Menuet et al., 2022). We utilized

the image metadata from the HCP and IBC datasets, which contains the corresponding Cognitive Atlas task as the ground truth.

Acknowledgments

Special thanks to the FIU Instructional & Research Computing Center (IRCC, http://ircc.fiu.edu) for providing the HPC and computing resources that contributed to the research results reported in this paper.

Ethical Statement

The Human Connectome Project provided the ethics and consent needed for the study and dissemination of HCP data. This secondary data analysis was approved by the Institutional Review Board of Florida International University.

Funding Statement

Funding for this project was provided by NIH R01-MH096906.

Data Accessibility

The functional neuroimaging data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; U54-MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. The results in volume space can be downloaded from collection 457 in NeuroVault (https://neurovault.org/collections/457/).

The Individual Brain Charting (IBC) dataset (Pinho et al., 2018) can be downloaded from collection 2138 in NeuroVault (https://neurovault.org/collections/2138/).

The meta-analytic parcellation images are publicly available for download at https://anima.fz-juelich.de/.

A simplified representation of the Cognitive Atlas task is available in GitHub https://github.com/Parietal-INRIA/fmri_decoding/blob/master/Data/labels/cogatlas_tasks_concepts_bertrand_mini.csv.

Code Availability

This project relied on multiple open-source Python packages, including: Jupyter (Kluyver et al., 2016), Matplotlib (Hunter, 2007), Neuromaps (Markello et al., 2022), NiBabel (Brett et al., 2020), Nilearn (Abraham et al., 2014), NiMARE (Salo et al., 2024, 2023), PyMARE (Yarkoni et al., 2024), NumPy (van der Walt et al., 2011), Pandas (McKinney, 2010), Scikit-learn (Pedregosa et al., 2011), SciPy (Virtanen et

al., 2020), Seaborn (Waskom, 2021), and SurfPlot (Gale et al., 2021),. We also used the HCP software Connectome Workbench (wb_command version 1.5.0, (Marcus et al., 2011)).

All code required to reproduce the analyses and figures in this paper is available on GitHub at https://github.com/NBCLab/brain-decoder. All data and resources that resulted from this paper are openly disseminated and made available on the Open Science Framework (OSF) at https://osf.io/dsj56/, including the links to the GitHub repository and figures.

Competing Interests

The authors declare no competing interests.

Author Contributions

ARL, JAP, AdlV, JBP, TEN, and JDK conceived and designed the project. JAP, AdlV, and JDK analyzed data. JAP and JDK contributed scripts and pipelines. JAP, TEN, and ARL wrote the paper, and all authors contributed to the revisions and approved the final version.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn. Front. Neuroinform. 8. https://doi.org/10.3389/fninf.2014.00014

Amft, M., Bzdok, D., Laird, A.R., Fox, P.T., Schilbach, L., Eickhoff, S.B., 2015. Definition and characterization of an extended social-affective default network. Brain Struct Funct 220, 1031–1049. https://doi.org/10.1007/s00429-013-0698-0

Barch, D.M., Burgess, G.C., Harms, M.P., Petersen, S.E., Schlaggar, B.L., Corbetta, M., Glasser, M.F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J.M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A.Z., Van Essen, D.C., 2013. Function in the human connectome: Task-fMRI and individual differences in behavior. NeuroImage, Mapping the Connectome 80, 169–189. https://doi.org/10.1016/j.neuroimage.2013.05.033

Bornmann, L., Mutz, R., 2015. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. Journal of the Association for Information Science and Technology 66, 2215–2222. https://doi.org/10.1002/asi.23329

Brett, M., Markiewicz, C.J., Hanke, M., Côté, M.-A., Cipollini, B., McCarthy, P., Jarecka, D., Cheng, C.P., Halchenko, Y.O., Cottaar, M., Larson, E., Ghosh, S., Wassermann, D., Gerhard, S., Lee, G.R., Wang, H.-T., Kastman, E., Kaczmarzyk, J., Guidotti, R., Duek, O., Daniel, J., Rokem, A., Madison, C., Moloney, B., Morency, F.C., Goncalves, M., Markello, R., Riddell, C., Burns, C., Millman, J., Gramfort, A., Leppäkangas, J., Sólon, A., van den Bosch, J.J.F., Vincent, R.D., Braun, H., Subramaniam, K., Gorgolewski, K.J., Raamana, P.R., Klug, J., Nichols, B.N., Baker, E.M., Hayashi, S., Pinsard, B., Haselgrove, C., Hymers, M., Esteban, O., Koudoro, S., Pérez-García, F., Oosterhof, N.N., Amirbekian, B., Nimmo-Smith, I., Nguyen, L., Reddigari, S., St-Jean, S., Panfilov, E., Garyfallidis, E., Varoquaux, G., Legarreta, J.H., Hahn, K.S., Hinds, O.P., Fauber, B., Poline, J.-B., Stutters, J., Jordan, K., Cieslak, M., Moreno, M.E., Haenel, V., Schwartz, Y., Baratz, Z., Darwin, B.C., Thirion, B., Gauthier, C., Papadopoulos Orfanos, D., Solovey, I., Gonzalez, I., Palasubramaniam, J., Lecher, J., Leinweber, K., Raktivan, K., Calábková, M., Fischer, P., Gervais, P., Gadde, S., Ballinger, T., Roos, T., Reddam, V.R., freec84, 2020. nipy/nibabel: 3.2.1. https://doi.org/10.5281/zenodo.4295521

Bzdok, D., Laird, A.R., Zilles, K., Fox, P.T., Eickhoff, S.B., 2013a. An investigation of the structural, connectional, and functional subspecialization in the human amygdala. Hum Brain Mapp 34, 3247–3266. https://doi.org/10.1002/hbm.22138

Bzdok, D., Langner, R., Schilbach, L., Jakobs, O., Roski, C., Caspers, S., Laird, A.R., Fox, P.T., Zilles, K., Eickhoff, S.B., 2013b. Characterization of the temporo-parietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding. Neuroimage 81, 381–392. https://doi.org/10.1016/j.neuroimage.2013.05.046

Casey, B.J., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., Soules, M.E., Teslovich, T., Dellarco, D.V., Garavan, H., Orr, C.A., Wager, T.D., Banich, M.T., Speer, N.K., Sutherland, M.T., Riedel, M.C., Dick, A.S., Bjork, J.M., Thomas, K.M., Chaarani, B., Mejia, M.H., Hagler, D.J., Daniela Cornejo, M., Sicat, C.S., Harms, M.P., Dosenbach, N.U.F., Rosenberg, M., Earl, E., Bartsch, H., Watts, R., Polimeni, J.R., Kuperman, J.M., Fair, D.A., Dale, A.M., 2018. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. Developmental Cognitive Neuroscience, The Adolescent Brain Cognitive Development (ABCD) Consortium: Rationale, Aims, and Assessment Strategy 32, 43–54. https://doi.org/10.1016/j.dcn.2018.03.001

Chang, L.J., Yarkoni, T., Khaw, M.W., Sanfey, A.G., 2013. Decoding the role of the insula in human cognition: functional parcellation and large-scale reverse inference. Cereb Cortex 23, 739–749. https://doi.org/10.1093/cercor/bhs065

Chase, H.W., Grace, A.A., Fox, P.T., Phillips, M.L., Eickhoff, S.B., 2020. Functional differentiation in the human ventromedial frontal lobe: A data‐driven parcellation. Hum Brain Mapp 41, 3266–3283. https://doi.org/10.1002/hbm.25014

Cieslik, E.C., Zilles, K., Caspers, S., Roski, C., Kellermann, T.S., Jakobs, O., Langner, R., Laird, A.R., Fox, P.T., Eickhoff, S.B., 2013. Is there "one" DLPFC in cognitive action control? Evidence for heterogeneity from co-activation-based parcellation. Cereb. Cortex 23, 2677–2689. https://doi.org/10.1093/cercor/bhs256

Dadi, K., Varoquaux, G., Machlouzarides-Shalit, A., Gorgolewski, K.J., Wassermann, D., Thirion, B., Mensch, A., 2020. Fine-grain atlases of functional modes for fMRI analysis. NeuroImage 221, 117126. https://doi.org/10.1016/j.neuroimage.2020.117126

Di Martino, A., O'Connor, D., Chen, B., Alaerts, K., Anderson, J.S., Assaf, M., Balsters, J.H., Baxter, L., Beggiato, A., Bernaerts, S., Blanken, L.M.E., Bookheimer, S.Y., Braden, B.B., Byrge, L., Castellanos, F.X., Dapretto, M., Delorme, R., Fair, D.A., Fishman, I., Fitzgerald, J., Gallagher, L., Keehn, R.J.J., Kennedy, D.P., Lainhart, J.E., Luna, B., Mostofsky, S.H., Müller, R.-A., Nebel, M.B., Nigg, J.T., O'Hearn, K., Solomon, M., Toro, R., Vaidya, C.J., Wenderoth, N., White, T., Craddock, R.C., Lord, C., Leventhal, B., Milham, M.P., 2017. Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. Sci Data 4, 170010. https://doi.org/10.1038/sdata.2017.10

Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., Deen, B., Delmonte, S., Dinstein, I., Ertl-Wagner, B., Fair, D.A., Gallagher, L., Kennedy, D.P., Keown, C.L., Keysers, C., Lainhart, J.E., Lord, C., Luna, B., Menon, V., Minshew, N.J., Monk, C.S., Mueller, S., Müller, R.-A., Nebel, M.B., Nigg, J.T., O'Hearn, K., Pelphrey, K.A., Peltier, S.J., Rudie, J.D., Sunaert, S., Thioux, M., Tyszka, J.M., Uddin, L.Q., Verhoeven, J.S., Wenderoth, N., Wiggins, J.L., Mostofsky, S.H., Milham, M.P., 2014. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiatry 19, 659–667. https://doi.org/10.1038/mp.2013.78

Dockès, J., Oudyk, K., Torabi, M., Vega, A.I. de la, Poline, J.-B., 2024. Mining the neuroimaging literature. eLife 13. https://doi.org/10.7554/eLife.94909.1

Dockès, J., Poldrack, R.A., Primet, R., Gözükan, H., Yarkoni, T., Suchanek, F., Thirion, B., Varoquaux, G., 2020. NeuroQuery, comprehensive meta-analysis of human brain mapping. eLife 9, e53385. https://doi.org/10.7554/eLife.53385

Fox, P.T., Laird, A.R., Fox, S.P., Fox, P.M., Uecker, A.M., Crank, M., Koenig, S.F., Lancaster, J.L., 2005. BrainMap taxonomy of experimental design: description and evaluation. Hum Brain Mapp 25, 185–198. https://doi.org/10.1002/hbm.20141

Gale, D.J., Vos de Wael., R., Benkarim, O., Bernhardt, B., 2021. Surfplot: Publication-ready brain surface figures. https://doi.org/10.5281/zenodo.5567926

Gorgolewski, K.J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S.S., Maumet, C., Sochat, V.V., Nichols, T.E., Poldrack, R.A., Poline, J.-B., Yarkoni, T., Margulies, D.S., 2015. NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. Frontiers in Neuroinformatics 9, 8. https://doi.org/10.3389/fninf.2015.00008

Gunasekar, S., Zhang, Y., Aneja, J., Mendes, C.C.T., Giorno, A.D., Gopi, S., Javaheripi, M., Kauffmann, P., Rosa, G. de, Saarikivi, O., Salim, A., Shah, S., Behl, H.S., Wang, X., Bubeck, S., Eldan, R., Kalai, A.T., Lee, Y.T., Li, Y., 2023. Textbooks Are All You Need.

https://doi.org/10.48550/arXiv.2306.11644
Hunter, J.D., 2007. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering 9, 90–95. https://doi.org/10.1109/MCSE.2007.55
Jack Jr., C.R., Arani, A., Borowski, B.J., Cash, D.M., Crawford, K., Das, S.R., DeCarli, C., Fletcher, E., Fox, N.C., Gunter, J.L., Ittyerah, R., Harvey, D.J., Jahanshad, N., Maillard, P., Malone, I.B., Nir, T.M., Reid, R.I., Reyes, D.A., Schwarz, C.G., Senjem, M.L., Thomas, D.L., Thompson, P.M., Tosun, D., Yushkevich, P.A., Ward, C.P., Weiner, M.W., Initiative, A.D.N., 2024. Overview of ADNI MRI. Alzheimer's & Dementia 20, 7350–7360. https://doi.org/10.1002/alz.14166
Jiang, A.Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D.S., Casas, D. de las, Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L.R., Lachaux, M.-A., Stock, P., Scao, T.L., Lavril, T., Wang, T., Lacroix, T., Sayed, W.E., 2023. Mistral 7B. https://doi.org/10.48550/arXiv.2310.06825
Kluyver, T., Ragan-Kelley, B., P&#233, Rez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., n, Abdalla, S., Willing, C., Team, J.D., 2016. Jupyter Notebooks – a publishing format for reproducible computational workflows. Positioning and Power in Academic Publishing: Players, Agents and Agendas 87–90. https://doi.org/10.3233/978-1-61499-649-1-87
Kober, H., Barrett, L.F., Joseph, J., Bliss-Moreau, E., Lindquist, K., Wager, T.D., 2008. Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis of neuroimaging studies. NeuroImage 42, 998–1031. https://doi.org/10.1016/j.neuroimage.2008.03.059
Laird, A.R., Eickhoff, S.B., Fox, P.M., Uecker, A.M., Ray, K.L., Saenz, J.J., McKay, D.R., Bzdok, D., Laird, R.W., Robinson, J.L., Turner, J.A., Turkeltaub, P.E., Lancaster, J.L., Fox, P.T., 2011. The BrainMap strategy for standardization, sharing, and meta-analysis of neuroimaging data. BMC Res Notes 4, 349. https://doi.org/10.1186/1756-0500-4-349
Laird, A.R., Eickhoff, S.B., Kurth, F., Fox, P.M., Uecker, A.M., Turner, J.A., Robinson, J.L., Lancaster, J.L., Fox, P.T., 2009. ALE Meta-Analysis Workflows Via the Brainmap Database: Progress Towards A Probabilistic Functional Brain Atlas. Front Neuroinform 3, 23. https://doi.org/10.3389/neuro.11.023.2009
Laird, A.R., Lancaster, J.L., Fox, P.T., 2005. BrainMap: the social evolution of a human brain mapping database. Neuroinformatics 3, 65–78. https://doi.org/10.1385/ni:3:1:065
Liu, X., Eickhoff, S.B., Hoffstaedter, F., Genon, S., Caspers, S., Reetz, K., Dogan, I., Eickhoff, C.R., Chen, J., Caspers, J., Reuter, N., Mathys, C., Aleman, A., Jardri, R., Riedl, V., Sommer, I.E., Patil, K.R., 2020. Joint Multi-modal Parcellation of the Human Striatum: Functions and Clinical Relevance. Neurosci Bull 36, 1123–1136. https://doi.org/10.1007/s12264-020-00543-1
Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., Ge, B., 2023. Summary of ChatGPT-Related research and perspective towards the future of large language models. Meta-Radiology 1, 100017. https://doi.org/10.1016/j.metrad.2023.100017
Luo, X., Rechardt, A., Sun, G., Nejad, K.K., Yáñez, F., Yilmaz, B., Lee, K., Cohen, A.O., Borghesani, V., Pashkov, A., Marinazzo, D., Nicholas, J., Salatiello, A., Sucholutsky, I., Minervini, P., Razavi, S., Rocca, R., Yusifov, E., Okalova, T., Gu, N., Ferianc, M., Khona, M., Patil, K.R., Lee, P.-S., Mata, R., Myers, N.E., Bizley, J.K., Musslick, S., Bilgin, I.P., Niso, G., Ales, J.M., Gaebler, M., Ratan Murty, N.A., Loued-Khenissi, L., Behler, A., Hall, C.M., Dafflon, J., Bao, S.D., Love, B.C., 2024. Large language models surpass human experts in predicting neuroscience results. Nat Hum Behav 1–11. https://doi.org/10.1038/s41562-024-02046-9
Marcus, D., Harwell, J., Olsen, T., Hodge, M., Glasser, M., Prior, F., Jenkinson, M., Laumann, T., Curtiss,

S., Van Essen, D., 2011. Informatics and Data Mining Tools and Strategies for the Human Connectome Project. Front. Neuroinform. 5. https://doi.org/10.3389/fninf.2011.00004

Markello, R.D., Hansen, J.Y., Liu, Z.-Q., Bazinet, V., Shafiei, G., Suárez, L.E., Blostein, N., Seidlitz, J., Baillet, S., Satterthwaite, T.D., Chakravarty, M.M., Raznahan, A., Misic, B., 2022. neuromaps: structural and functional interpretation of brain maps. https://doi.org/10.1101/2022.01.06.475081

McKinney, W., 2010. Data Structures for Statistical Computing in Python. Presented at the Python in Science Conference, Austin, Texas, pp. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Mensch, A., Mairal, J., Bzdok, D., Thirion, B., Varoquaux, G., 2017. Learning Neural Representations of Human Cognition across Many fMRI Studies, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.

Mensch, A., Mairal, J., Thirion, B., Varoquaux, G., 2021. Extracting representations of cognition across neuroimaging studies improves brain decoding. PLOS Computational Biology 17, e1008795. https://doi.org/10.1371/journal.pcbi.1008795

Menuet, R., Meudec, R., Dockès, J., Varoquaux, G., Thirion, B., 2022. Comprehensive decoding mental processes from Web repositories of functional brain images. Sci Rep 12, 7050. https://doi.org/10.1038/s41598-022-10710-1

Meudec, R., Dockès, J., Wassermann, D., Thirion, B., 2022. Peaks2Image: Reconstructing fMRI Statistical Maps from Peaks.

Meudec, R., Ghayem, F., Dockès, J., Wassermann, D., Thirion, B., 2024. NeuroConText: Contrastive Text-to-Brain Mapping for Neuroscientific Literature, in: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. Springer Nature Switzerland, Cham, pp. 325–335. https://doi.org/10.1007/978-3-031-72384-1_31

Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.-M., Malave, V.L., Mason, R.A., Just, M.A., 2008. Predicting Human Brain Activity Associated with the Meanings of Nouns. Science 320, 1191–1195. https://doi.org/10.1126/science.1152876

Ngo, G.H., Nguyen, M., Chen, N.F., Sabuncu, M.R., 2021. Text2Brain: Synthesis of Brain Activation Maps from Free-Form Text Query, in: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Springer International Publishing, Cham, pp. 605–614. https://doi.org/10.1007/978-3-030-87234-2_57

Nickl-Jockschat, T., Rottschy, C., Thommes, J., Schneider, F., Laird, A.R., Fox, P.T., Eickhoff, S.B., 2015. Neural networks related to dysfunctional face processing in autism spectrum disorder. Brain Struct Funct 220, 2355–2371. https://doi.org/10.1007/s00429-014-0791-z

Oord, A. van den, Li, Y., Vinyals, O., 2019. Representation Learning with Contrastive Predictive Coding. https://doi.org/10.48550/arXiv.1807.03748

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 6.

Peraza, J.A., Kent, J.D., Blair, R.W., Poline, J.-B., Nichols, T.E., Vega, A. de la, Laird, A.R., 2025. Advancing image-based meta-analysis for fMRI: A framework for leveraging NeuroVault data. https://doi.org/10.1101/2025.03.06.641922

Peraza, J.A., Salo, T., Riedel, M.C., Bottenhorn, K.L., Poline, J.-B., Dockès, J., Kent, J.D., Bartley, J.E., Flannery, J.S., Hill-Bowen, L.D., Lobo, R.P., Poudel, R., Ray, K.L., Robinson, J.L., Laird, R.W., Sutherland, M.T., de la Vega, A., Laird, A.R., 2024. Methods for decoding cortical gradients of

functional connectivity. Imaging Neuroscience 2, 1–32. https://doi.org/10.1162/imag_a_00081
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., 2018. Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping. Sci Data 5, 180105. https://doi.org/10.1038/sdata.2018.105
Plachti, A., Eickhoff, S.B., Hoffstaedter, F., Patil, K.R., Laird, A.R., Fox, P.T., Amunts, K., Genon, S., 2019. Multimodal Parcellations and Extensive Behavioral Profiling Tackling the Hippocampus Gradient. Cereb Cortex 29, 4595–4612. https://doi.org/10.1093/cercor/bhy336
Poldrack, R.A., 2011. Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. Neuron 72, 692–697. https://doi.org/10.1016/j.neuron.2011.11.001
Poldrack, R.A., 2006. Can cognitive processes be inferred from neuroimaging data? Trends in Cognitive Sciences 10, 59–63. https://doi.org/10.1016/j.tics.2005.12.004
Poldrack, R.A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., Parker, D.S., Sabb, F.W., Bilder, R.M., 2011. The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. Front Neuroinform 5, 17. https://doi.org/10.3389/fninf.2011.00017
Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I., 2021. Learning Transferable Visual Models From Natural Language Supervision. https://doi.org/10.48550/arXiv.2103.00020
Rottschy, C., Caspers, S., Roski, C., Reetz, K., Dogan, I., Schulz, J.B., Zilles, K., Laird, A.R., Fox, P.T., Eickhoff, S.B., 2013. Differentiated parietal connectivity of frontal regions for "what" and "where" memory. Brain Struct Funct 218, 1551–1567. https://doi.org/10.1007/s00429-012-0476-4
Rubin, T., Koyejo, O.O., Jones, M.N., Yarkoni, T., 2016. Generalized Correspondence-LDA Models (GC-LDA) for Identifying Functional Regions in the Brain, in: NIPS'16 Proceedings of the 30th International Conference on Neural Information Processing Systems. Presented at the 29th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, pp. 1118–1126.
Rubin, T.N., Koyejo, O., Gorgolewski, K.J., Jones, M.N., Poldrack, R.A., Yarkoni, T., 2017. Decoding brain activity using a large-scale probabilistic functional-anatomical atlas of human cognition. PLOS Computational Biology 13, e1005649. https://doi.org/10.1371/journal.pcbi.1005649
Salo, T., Yarkoni, T., Nichols, T.E., Poline, J.-B., Bilgel, M., Bottenhorn, K.L., Jarecka, D., Kent, J.D., Kimbler, A., Nielson, D.M., Oudyk, K.M., Peraza, J.A., Pérez, A., Reeders, P.C., Yanes, J.A., Laird, A.R., 2023. NiMARE: Neuroimaging Meta-Analysis Research Environment. Aperture Neuro 3, 1–32. https://doi.org/10.52294/001c.87681
Salo, T., Yarkoni, T., Nichols, T.E., Poline, J.-B., Kent, J.D., Gorgolewski, K.J., Glerean, E., Bottenhorn, K.L., Bilgel, M., Wright, J., Reeders, P., Kimbler, A., Nielson, D.N., Yanes, J.A., Pérez, A., Oudyk, K.M., Jarecka, D., Enge, A., Peraza, J.A., Eickhoff, S.B., De La Vega, A., Liu, Z.-Q., Laird, A.R., 2024. neurostuff/NiMARE: 0.4.1. https://doi.org/10.5281/zenodo.14183422
Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., Kelly, M., Laumann, T., Miller, K.L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A.Z., Vu, A.T., Woolrich, M.W., Xu, J., Yacoub, E., Uğurbil, K., Van Essen, D.C., Glasser, M.F., 2013. Resting-state fMRI in the Human Connectome Project. NeuroImage, Mapping the Connectome 80, 144–168. https://doi.org/10.1016/j.neuroimage.2013.05.039
Srivastava, A., Rastogi, A., Rao, A., Shoeb, A.A.M., Abid, A., Fisch, A., Brown, A.R., Santoro, A., Gupta,

Aditya, Garriga-Alonso, A., Kluska, A., Lewkowycz, A., Agarwal, A., Power, A., Ray, A., Warstadt, A., Kocurek, A.W., Safaya, A., Tazarv, A., Xiang, A., Parrish, A., Nie, A., Hussain, A., Askell, A., Dsouza, A., Slone, A., Rahane, A., Iyer, A.S., Andreassen, A., Madotto, A., Santilli, A., Stuhlmüller, A., Dai, A., La, A., Lampinen, A., Zou, A., Jiang, A., Chen, A., Vuong, A., Gupta, Animesh, Gottardi, A., Norelli, A., Venkatesh, A., Gholamidavoodi, A., Tabassum, A., Menezes, A., Kirubarajan, A., Mullokandov, A., Sabharwal, A., Herrick, A., Efrat, A., Erdem, A., Karakaş, A., Roberts, B.R., Loe, B.S., Zoph, B., Bojanowski, B., Özyurt, B., Hedayatnia, B., Neyshabur, B., Inden, B., Stein, B., Ekmekci, B., Lin, B.Y., Howald, B., Orinion, B., Diao, C., Dour, C., Stinson, C., Argueta, C., Ramírez, C.F., Singh, C., Rathkopf, C., Meng, C., Baral, C., Wu, C., Callison-Burch, C., Waites, C., Voigt, C., Manning, C.D., Potts, C., Ramirez, C., Rivera, C.E., Siro, C., Raffel, C., Ashcraft, C., Garbacea, C., Sileo, D., Garrette, D., Hendrycks, D., Kilman, D., Roth, D., Freeman, D., Khashabi, D., Levy, D., González, D.M., Perszyk, D., Hernandez, D., Chen, Danqi, Ippolito, D., Gilboa, D., Dohan, D., Drakard, D., Jurgens, D., Datta, D., Ganguli, D., Emelin, D., Kleyko, D., Yuret, D., Chen, Derek, Tam, D., Hupkes, D., Misra, D., Buzan, D., Mollo, D.C., Yang, D., Lee, D.-H., Schrader, D., Shutova, E., Cubuk, E.D., Segal, E., Hagerman, E., Barnes, E., Donoway, E., Pavlick, E., Rodola, E., Lam, E., Chu, E., Tang, E., Erdem, E., Chang, E., Chi, E.A., Dyer, E., Jerzak, E., Kim, E., Manyasi, E.E., Zheltonozhskii, E., Xia, F., Siar, F., Martínez-Plumed, F., Happé, F., Chollet, F., Rong, F., Mishra, G., Winata, G.I., Melo, G. de, Kruszewski, G., Parascandolo, G., Mariani, G., Wang, G., Jaimovitch-López, G., Betz, G., Gur-Ari, G., Galijasevic, H., Kim, H., Rashkin, H., Hajishirzi, H., Mehta, H., Bogar, H., Shevlin, H., Schütze, H., Yakura, H., Zhang, H., Wong, H.M., Ng, I., Noble, I., Jumelet, J., Geissinger, J., Kernion, J., Hilton, J., Lee, J., Fisac, J.F., Simon, J.B., Koppel, J., Zheng, J., Zou, J., Kocoń, J., Thompson, J., Wingfield, J., Kaplan, J., Radom, J., Sohl-Dickstein, J., Phang, J., Wei, J., Yosinski, J., Novikova, J., Bosscher, J., Marsh, J., Kim, J., Taal, J., Engel, J., Alabi, J., Xu, J., Song, J., Tang, J., Waweru, J., Burden, J., Miller, J., Balis, J.U., Batchelder, J., Berant, J., Frohberg, J., Rozen, J., Hernandez-Orallo, J., Boudeman, J., Guerr, J., Jones, J., Tenenbaum, J.B., Rule, J.S., Chua, J., Kanclerz, K., Livescu, K., Krauth, K., Gopalakrishnan, K., Ignatyeva, K., Markert, K., Dhole, K.D., Gimpel, K., Omondi, K., Mathewson, K., Chiafullo, K., Shkaruta, K., Shridhar, K., McDonell, K., Richardson, K., Reynolds, L., Gao, L., Zhang, L., Dugan, L., Qin, L., Contreras-Ochando, L., Morency, L.-P., Moschella, L., Lam, L., Noble, L., Schmidt, L., He, L., Colón, L.O., Metz, L., Şenel, L.K., Bosma, M., Sap, M., Hoeve, M. ter, Farooqi, M., Faruqui, M., Mazeika, M., Baturan, M., Marelli, M., Maru, M., Quintana, M.J.R., Tolkiehn, M., Giulianelli, M., Lewis, M., Potthast, M., Leavitt, M.L., Hagen, M., Schubert, M., Baitemirova, M.O., Arnaud, M., McElrath, M., Yee, M.A., Cohen, M., Gu, M., Ivanitskiy, M., Starritt, M., Strube, M., Swędrowski, M., Bevilacqua, M., Yasunaga, M., Kale, M., Cain, M., Xu, M., Suzgun, M., Walker, M., Tiwari, M., Bansal, M., Aminnaseri, M., Geva, M., Gheini, M., T, M.V., Peng, N., Chi, N.A., Lee, N., Krakover, N.G.-A., Cameron, N., Roberts, N., Doiron, N., Martinez, N., Nangia, N., Deckers, N., Muennighoff, N., Keskar, N.S., Iyer, N.S., Constant, N., Fiedel, N., Wen, N., Zhang, O., Agha, O., Elbaghdadi, O., Levy, O., Evans, O., Casares, P.A.M., Doshi, P., Fung, P., Liang, P.P., Vicol, P., Alipoormolabashi, P., Liao, P., Liang, P., Chang, P., Eckersley, P., Htut, P.M., Hwang, P., Miłkowski, P., Patil, P., Pezeshkpour, P., Oli, P., Mei, Q., Lyu, Q., Chen, Q., Banjade, R., Rudolph, R.E., Gabriel, R., Habacker, R., Risco, R., Millière, R., Garg, R., Barnes, R., Saurous, R.A., Arakawa, R., Raymaekers, R., Frank, R., Sikand, R., Novak, R., Sitelew, R., LeBras, R., Liu, R., Jacobs, R., Zhang, R., Salakhutdinov, R., Chi, R., Lee, R., Stovall, R., Teehan, R., Yang, R., Singh, Sahib, Mohammad, S.M., Anand, S., Dillavou, S., Shleifer, S., Wiseman, S., Gruetter, S., Bowman, S.R., Schoenholz, S.S., Han, S., Kwatra, S., Rous, S.A., Ghazarian, S., Ghosh, S., Casey, S., Bischoff, S., Gehrmann, S., Schuster, S., Sadeghi, S.,

Hamdan, S., Zhou, S., Srivastava, S., Shi, S., Singh, Shikhar, Asaadi, S., Gu, S.S., Pachchigar, S., Toshniwal, S., Upadhyay, S., Shyamolima, Debnath, Shakeri, S., Thormeyer, S., Melzi, S., Reddy, S., Makini, S.P., Lee, S.-H., Torene, S., Hatwar, S., Dehaene, S., Divic, S., Ermon, S., Biderman, S., Lin, S., Prasad, S., Piantadosi, S.T., Shieber, S.M., Misherghi, S., Kiritchenko, S., Mishra, S., Linzen, T., Schuster, T., Li, T., Yu, T., Ali, T., Hashimoto, T., Wu, T.-L., Desbordes, T., Rothschild, T., Phan, T., Wang, T., Nkinyili, T., Schick, T., Kornev, T., Tunduny, T., Gerstenberg, T., Chang, T., Neeraj, T., Khot, T., Shultz, T., Shaham, U., Misra, V., Demberg, V., Nyamai, V., Raunak, V., Ramasesh, V., Prabhu, V.U., Padmakumar, V., Srikumar, V., Fedus, W., Saunders, W., Zhang, W., Vossen, W., Ren, X., Tong, X., Zhao, X., Wu, X., Shen, X., Yaghoobzadeh, Y., Lakretz, Y., Song, Y., Bahri, Y., Choi, Y., Yang, Y., Hao, Y., Chen, Y., Belinkov, Y., Hou, Yu, Hou, Yufang, Bai, Y., Seid, Z., Zhao, Z., Wang, Zijian, Wang, Z.J., Wang, Zirui, Wu, Z., 2023. Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models. https://doi.org/10.48550/arXiv.2206.04615

Thompson, P.M., Stein, J.L., Medland, S.E., Hibar, D.P., Vasquez, A.A., Renteria, M.E., Toro, R., Jahanshad, N., Schumann, G., Franke, B., Wright, M.J., Martin, N.G., Agartz, I., Alda, M., Alhusaini, S., Almasy, L., Almeida, J., Alpert, K., Andreasen, N.C., Andreassen, O.A., Apostolova, L.G., Appel, K., Armstrong, N.J., Aribisala, B., Bastin, M.E., Bauer, M., Bearden, C.E., Bergmann, O., Binder, E.B., Blangero, J., Bockholt, H.J., Bøen, E., Bois, C., Boomsma, D.I., Booth, T., Bowman, I.J., Bralten, J., Brouwer, R.M., Brunner, H.G., Brohawn, D.G., Buckner, R.L., Buitelaar, J., Bulayeva, K., Bustillo, J.R., Calhoun, V.D., Cannon, D.M., Cantor, R.M., Carless, M.A., Caseras, X., Cavalleri, G.L., Chakravarty, M.M., Chang, K.D., Ching, C.R.K., Christoforou, A., Cichon, S., Clark, V.P., Conrod, P., Coppola, G., Crespo-Facorro, B., Curran, J.E., Czisch, M., Deary, I.J., de Geus, E.J.C., den Braber, A., Delvecchio, G., Depondt, C., de Haan, L., de Zubicaray, G.I., Dima, D., Dimitrova, R., Djurovic, S., Dong, H., Donohoe, G., Duggirala, R., Dyer, T.D., Ehrlich, S., Ekman, C.J., Elvsåshagen, T., Emsell, L., Erk, S., Espeseth, T., Fagerness, J., Fears, S., Fedko, I., Fernández, G., Fisher, S.E., Foroud, T., Fox, P.T., Francks, C., Frangou, S., Frey, E.M., Frodl, T., Frouin, V., Garavan, H., Giddaluru, S., Glahn, D.C., Godlewska, B., Goldstein, R.Z., Gollub, R.L., Grabe, H.J., Grimm, O., Gruber, O., Guadalupe, T., Gur, R.E., Gur, R.C., Göring, H.H.H., Hagenaars, S., Hajek, T., Hall, G.B., Hall, J., Hardy, J., Hartman, C.A., Hass, J., Hatton, S.N., Haukvik, U.K., Hegenscheid, K., Heinz, A., Hickie, I.B., Ho, B.-C., Hoehn, D., Hoekstra, P.J., Hollinshead, M., Holmes, A.J., Homuth, G., Hoogman, M., Hong, L.E., Hosten, N., Hottenga, J.-J., Hulshoff Pol, H.E., Hwang, K.S., Jack, C.R., Jenkinson, M., Johnston, C., Jönsson, E.G., Kahn, R.S., Kasperaviciute, D., Kelly, S., Kim, S., Kochunov, P., Koenders, L., Krämer, B., Kwok, J.B.J., Lagopoulos, J., Laje, G., Landen, M., Landman, B.A., Lauriello, J., Lawrie, S.M., Lee, P.H., Le Hellard, S., Lemaître, H., Leonardo, C.D., Li, C.-S., Liberg, B., Liewald, D.C., Liu, X., Lopez, L.M., Loth, E., Lourdusamy, A., Luciano, M., Macciardi, F., Machielsen, M.W.J., Macqueen, G.M., Malt, U.F., Mandl, R., Manoach, D.S., Martinot, J.-L., Matarin, M., Mather, K.A., Mattheisen, M., Mattingsdal, M., Meyer-Lindenberg, A., McDonald, C., McIntosh, A.M., McMahon, F.J., McMahon, K.L., Meisenzahl, E., Melle, I., Milaneschi, Y., Mohnke, S., Montgomery, G.W., Morris, D.W., Moses, E.K., Mueller, B.A., Muñoz Maniega, S., Mühleisen, T.W., Müller-Myhsok, B., Mwangi, B., Nauck, M., Nho, K., Nichols, T.E., Nilsson, L.-G., Nugent, A.C., Nyberg, L., Olvera, R.L., Oosterlaan, J., Ophoff, R.A., Pandolfo, M., Papalampropoulou-Tsiridou, M., Papmeyer, M., Paus, T., Pausova, Z., Pearlson, G.D., Penninx, B.W., Peterson, C.P., Pfennig, A., Phillips, M., Pike, G.B., Poline, J.-B., Potkin, S.G., Pütz, B., Ramasamy, A., Rasmussen, J., Rietschel, M., Rijpkema, M., Risacher, S.L., Roffman, J.L., Roiz-Santiañez, R., Romanczuk-Seiferth, N., Rose, E.J., Royle, N.A., Rujescu, D., Ryten, M.,

Sachdev, P.S., Salami, A., Satterthwaite, T.D., Savitz, J., Saykin, A.J., Scanlon, C., Schmaal, L., Schnack, H.G., Schork, A.J., Schulz, S.C., Schür, R., Seidman, L., Shen, L., Shoemaker, J.M., Simmons, A., Sisodiya, S.M., Smith, C., Smoller, J.W., Soares, J.C., Sponheim, S.R., Sprooten, E., Starr, J.M., Steen, V.M., Strakowski, S., Strike, L., Sussmann, J., Sämann, P.G., Teumer, A., Toga, A.W., Tordesillas-Gutierrez, D., Trabzuni, D., Trost, S., Turner, J., Van den Heuvel, M., van der Wee, N.J., van Eijk, K., van Erp, T.G.M., van Haren, N.E.M., van 't Ent, D., van Tol, M.-J., Valdés Hernández, M.C., Veltman, D.J., Versace, A., Völzke, H., Walker, R., Walter, H., Wang, L., Wardlaw, J.M., Weale, M.E., Weiner, M.W., Wen, W., Westlye, L.T., Whalley, H.C., Whelan, C.D., White, T., Winkler, A.M., Wittfeld, K., Woldehawariat, G., Wolf, C., Zilles, D., Zwiers, M.P., Thalamuthu, A., Schofield, P.R., Freimer, N.B., Lawrence, N.S., Drevets, W., Alzheimer's Disease Neuroimaging Initiative, EPIGEN Consortium, IMAGEN Consortium, Saguenay Youth Study (SYS) Group, 2014. The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. Brain Imaging Behav 8, 153–182. https://doi.org/10.1007/s11682-013-9269-5

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C.C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., Fuller, B., Gao, C., Goswami, V., Goyal, N., Hartshorn, A., Hosseini, S., Hou, R., Inan, H., Kardas, M., Kerkez, V., Khabsa, M., Kloumann, I., Korenev, A., Koura, P.S., Lachaux, M.-A., Lavril, T., Lee, J., Liskovich, D., Lu, Y., Mao, Y., Martinet, X., Mihaylov, T., Mishra, P., Molybog, I., Nie, Y., Poulton, A., Reizenstein, J., Rungta, R., Saladi, K., Schelten, A., Silva, R., Smith, E.M., Subramanian, R., Tan, X.E., Tang, B., Taylor, R., Williams, A., Kuan, J.X., Xu, P., Yan, Z., Zarov, I., Zhang, Y., Fan, A., Kambadur, M., Narang, S., Rodriguez, A., Stojnic, R., Edunov, S., Scialom, T., 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. https://doi.org/10.48550/arXiv.2307.09288

Uğurbil, K., Xu, J., Auerbach, E.J., Moeller, S., Vu, A.T., Duarte-Carvajalino, J.M., Lenglet, C., Wu, X., Schmitter, S., Van de Moortele, P.F., Strupp, J., Sapiro, G., De Martino, F., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D.A., Smith, S.M., Miller, K.L., Sotiropoulos, S.N., Jbabdi, S., Andersson, J.L.R., Behrens, T.E.J., Glasser, M.F., Van Essen, D.C., Yacoub, E., 2013. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. NeuroImage, Mapping the Connectome 80, 80–104. https://doi.org/10.1016/j.neuroimage.2013.05.012

van der Walt, S., Colbert, S.C., Varoquaux, G., 2011. The NumPy Array: A Structure for Efficient Numerical Computation. Computing in Science Engineering 13, 22–30. https://doi.org/10.1109/MCSE.2011.37

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., 2013. The WU-Minn Human Connectome Project: An overview. NeuroImage, Mapping the Connectome 80, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E.J., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., Della Penna, S., Feinberg, D., Glasser, M.F., Harel, N., Heath, A.C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S.E., Prior, F., Schlaggar, B.L., Smith, S.M., Snyder, A.Z., Xu, J., Yacoub, E., 2012. The Human Connectome Project: A data acquisition perspective. NeuroImage, Connectivity 62, 2222–2231. https://doi.org/10.1016/j.neuroimage.2012.02.018

Varoquaux, G., Schwartz, Y., Poldrack, R.A., Gauthier, B., Bzdok, D., Poline, J.-B., Thirion, B., 2018. Atlases of cognition with large-scale human brain mapping. PLOS Computational Biology 14, e1006565. https://doi.org/10.1371/journal.pcbi.1006565

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. ukasz, Polosukhin, I., 2017. Attention is All you Need, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Walt, S.J. van der, Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., Mulbregt, P. van, 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 1–12. https://doi.org/10.1038/s41592-019-0686-2

Wager, T.D., Lindquist, M., Kaplan, L., 2007. Meta-analysis of functional neuroimaging data: current and future directions. Social Cognitive and Affective Neuroscience 2, 150–158. https://doi.org/10.1093/scan/nsm015

Wager, T.D., Lindquist, M.A., Nichols, T.E., Kober, H., Van Snellenberg, J.X., 2009. Evaluating the consistency and specificity of neuroimaging data using meta-analysis. NeuroImage, Mathematics in Brain Imaging 45, S210–S221. https://doi.org/10.1016/j.neuroimage.2008.10.061

Waskom, M.L., 2021. seaborn: statistical data visualization. Journal of Open Source Software 6, 3021. https://doi.org/10.21105/joss.03021

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nature Methods 8, 665–670. https://doi.org/10.1038/nmeth.1635

Yarkoni, T., Salo, T., Peraza, J.A., Nichols, T.E., 2024. neurostuff/PyMARE: 0.0.8. https://doi.org/10.5281/zenodo.13743687