

# NeuroConText: Contrastive Learning for Neuroscience Meta-Analysis with Rich Text Representation

Fateme Ghayem,<sup>1\*</sup> Raphaël Meudec,<sup>1</sup> Jérôme Dockès,<sup>1</sup>  
Bertrand Thirion,<sup>1</sup> Demian Wassermann<sup>1</sup>

<sup>1</sup>Université Paris-Saclay, Inria, CEA

Palaiseau, 91120, France

\*Correspondence: fatemeh.ghayyem@inria.fr

July 14, 2025

## Supplementary Material

### A NeuroConText algorithm and its convergence proof

Proposition 1 establishes the convergence of NeuroConText algorithm, followed with the proof. We note that to set the step size  $\mu$ , we monitor the convergence of our approach and continuously track the decrease in InfoNCE and MSE losses, assess model performance on a validation set, and examine the norms of the gradients and the stability of parameter updates. This ensures that each optimization step contributes effectively to the overall training objective.

**Proposition 1** (Monotonic decrease of the total loss). *Consider the objective functions  $\mathcal{L}_{\text{InfoNCE}}(\theta_{\text{text}}, \theta_{\text{img}})$ ,  $\mathcal{L}_{\text{MSE}}(\theta_{\text{text}}, \theta_{\text{dec}})$ , and define*

$$\mathcal{L}(\theta_{\text{text}}, \theta_{\text{img}}, \theta_{\text{dec}}) = \mathcal{L}_{\text{InfoNCE}}(\theta_{\text{text}}, \theta_{\text{img}}) + \mathcal{L}_{\text{MSE}}(\theta_{\text{text}}, \theta_{\text{dec}}).$$

*Assume  $\mathcal{L}_{\text{InfoNCE}}$  is  $L_1$ -smooth in  $(\theta_{\text{text}}, \theta_{\text{img}})$ , and  $\mathcal{L}_{\text{MSE}}$  is  $L_2$ -smooth in  $(\theta_{\text{text}}, \theta_{\text{dec}})$ . Then, for a sufficiently small step size  $\mu$ , the update scheme in **Algorithm 1** produces a nonincreasing sequence*

$\{\mathcal{L}(\theta_{\text{text}}^{(k)}, \theta_{\text{img}}^{(k)}, \theta_{\text{dec}}^{(k)})\}$ , which is therefore convergent to a local minimum since  $\mathcal{L}$  is non-negative and thus lower-bounded.

---

**Algorithm 1:** Alternating Minimization of  $\mathcal{L}_{\text{InfoNCE}}(\theta_{\text{text}}, \theta_{\text{img}}) + \mathcal{L}_{\text{MSE}}(\theta_{\text{text}}, \theta_{\text{dec}})$ 


---

**Data:** Training data  $(\mathbf{X}, \mathbf{Y})$ , step size  $\mu > 0$ , initial parameters  $\theta_{\text{text}}^{(0)}, \theta_{\text{img}}^{(0)}, \theta_{\text{dec}}^{(0)}$

**Result:** Final parameters  $\theta_{\text{text}}^{(k)}, \theta_{\text{img}}^{(k)}, \theta_{\text{dec}}^{(k)}$

**while** stopping criterion not met **do**

/\* Step 1: Intermediate update of  $\theta_{\text{text}}$  w.r.t  $\mathcal{L}_{\text{InfoNCE}}$  \*/

$$\tilde{\theta}_{\text{text}}^{(k+1)} \leftarrow \theta_{\text{text}}^{(k)} - \mu \nabla_{\theta_{\text{text}}} \mathcal{L}_{\text{InfoNCE}} \left( \theta_{\text{text}}^{(k)}, \theta_{\text{img}}^{(k)} \right);$$

/\* Step 2: Update of  $\theta_{\text{img}}$  w.r.t  $\mathcal{L}_{\text{InfoNCE}}$  \*/

$$\theta_{\text{img}}^{(k+1)} \leftarrow \theta_{\text{img}}^{(k)} - \mu \nabla_{\theta_{\text{img}}} \mathcal{L}_{\text{InfoNCE}} \left( \theta_{\text{text}}^{(k)}, \theta_{\text{img}}^{(k)} \right);$$

/\* Step 3: Final update of  $\theta_{\text{text}}$  w.r.t  $\mathcal{L}_{\text{MSE}}$  \*/

$$\theta_{\text{text}}^{(k+1)} \leftarrow \tilde{\theta}_{\text{text}}^{(k+1)} - \mu \nabla_{\theta_{\text{text}}} \mathcal{L}_{\text{MSE}} \left( \tilde{\theta}_{\text{text}}^{(k+1)}, \theta_{\text{dec}}^{(k)} \right);$$

/\* Step 4: Update of  $\theta_{\text{dec}}$  w.r.t  $\mathcal{L}_{\text{MSE}}$  \*/

$$\theta_{\text{dec}}^{(k+1)} \leftarrow \theta_{\text{dec}}^{(k)} - \mu \nabla_{\theta_{\text{dec}}} \mathcal{L}_{\text{MSE}} \left( \theta_{\text{text}}^{(k+1)}, \theta_{\text{dec}}^{(k)} \right);$$

$$k \leftarrow k + 1;$$

/\* Check stopping criterion \*/

**end**

---

**Proof.** Let's define  $\theta_1 := \theta_{\text{text}}$ ,  $\theta_2 := \theta_{\text{img}}$ ,  $\theta_3 := \theta_{\text{dec}}$  and two sub-losses  $f(\theta_1, \theta_2) = \mathcal{L}_{\text{InfoNCE}}(\theta_1, \theta_2)$  and  $g(\theta_1, \theta_3) = \mathcal{L}_{\text{MSE}}(\theta_1, \theta_3)$ . Then, the total objective becomes:

$$\mathcal{L}(\theta_1, \theta_2, \theta_3) = f(\theta_1, \theta_2) + g(\theta_1, \theta_3).$$

We divide the proof into two parts: (1) showing that the  $\theta_1, \theta_2$  updates (Steps 1–2 in Algorithm 1) reduce  $f$ ; and (2) showing that the  $\theta_1, \theta_3$  updates (Steps 3–4 in Algorithm 1) reduce  $g$ . We then argue that these reductions are not undone by the intervening updates, provided  $\mu$  is sufficiently small.

### 1. Steps 1–2 decrease $f(\theta_1, \theta_2)$ :

#### Step 1 (gradient descent on $\theta_1$ wrt $f$ ):

$$\tilde{\theta}_1^{(k+1)} = \theta_1^{(k)} - \mu \nabla_{\theta_1} f(\theta_1^{(k)}, \theta_2^{(k)}).$$

Under  $L_f$ -smoothness of  $f$  and for  $\mu \leq 1/L_f$ , standard gradient-descent theory Bertsekas, 1997; Nesterov et al., 2018 guarantees

$$f(\tilde{\theta}_1^{(k+1)}, \theta_2^{(k)}) \leq f(\theta_1^{(k)}, \theta_2^{(k)}) - \Delta_{f,1}^{(k)},$$

where  $\Delta_{f,1}^{(k)} > 0$  depends on  $\mu$  and  $\|\nabla_{\theta_1} f(\cdot)\|^2$ .

**Step 2 (gradient descent on  $\theta_2$  wrt  $f$ ):**

$$\theta_2^{(k+1)} = \theta_2^{(k)} - \mu \nabla_{\theta_2} f(\theta_1^{(k)}, \theta_2^{(k)}).$$

By  $L_f$ -smoothness in  $\theta_2$ , we get

$$f(\theta_1^{(k)}, \theta_2^{(k+1)}) \leq f(\theta_1^{(k)}, \theta_2^{(k)}) - \Delta_{f,2}^{(k)},$$

for some positive  $\Delta_{f,2}^{(k)}$ . Combining these sub-steps yields

$$f(\tilde{\theta}_1^{(k+1)}, \theta_2^{(k+1)}) \leq f(\theta_1^{(k)}, \theta_2^{(k)}) - \Delta_f^{(k)}, \quad \Delta_f^{(k)} = \Delta_{f,1}^{(k)} + \Delta_{f,2}^{(k)} > 0.$$

Hence, after Steps 1–2,  $f$  is strictly decreased from its value at the start of iteration  $k$ .

2. Steps 3–4 Decrease  $g(\theta_1, \theta_3)$ .

**Step 3 (gradient descent on  $\theta_1$  wrt  $g$ ):**

$$\theta_1^{(k+1)} = \tilde{\theta}_1^{(k+1)} - \mu \nabla_{\theta_1} g(\tilde{\theta}_1^{(k+1)}, \theta_3^{(k)}).$$

If  $g$  is  $L_g$ -smooth in  $\theta_1$  and  $\mu \leq 1/L_g$ , then

$$g(\theta_1^{(k+1)}, \theta_3^{(k)}) \leq g(\tilde{\theta}_1^{(k+1)}, \theta_3^{(k)}) - \Delta_{g,1}^{(k)}.$$

**Step 4 (gradient descent on  $\theta_3$  wrt  $g$ ):**

$$\theta_3^{(k+1)} = \theta_3^{(k)} - \mu \nabla_{\theta_3} g(\theta_1^{(k+1)}, \theta_3^{(k)}).$$

By the same  $L_g$ -smoothness argument,

$$g(\theta_1^{(k+1)}, \theta_3^{(k+1)}) \leq g(\theta_1^{(k+1)}, \theta_3^{(k)}) - \Delta_{g,2}^{(k)},$$

so combining both sub-steps gives

$$g(\theta_1^{(k+1)}, \theta_3^{(k+1)}) \leq g(\tilde{\theta}_1^{(k+1)}, \theta_3^{(k)}) - \Delta_g^{(k)}, \quad \Delta_g^{(k)} = \Delta_{g,1}^{(k)} + \Delta_{g,2}^{(k)} > 0.$$

Thus, after Steps 3–4,  $g$  is strictly decreased from its value at the end of Step 2.

### 3. Ensuring the second update of $\theta_1$ (Step 3) cannot undo the decrease in $f$ .

A critical concern is that  $\theta_1$  is changed *twice* in the same iteration: once w.r.t.  $f$  (Step 1), once w.r.t.  $g$  (Step 3). In the following, we prove that by choosing  $\mu$  sufficiently small, the second update does not cause  $f$  to increase again. In particular, a common sufficient condition is

$$0 < \mu < \frac{1}{L_f + L_g} \quad \text{or} \quad 0 < \mu < \frac{1}{2 \times \max(L_f, L_g)}.$$

#### 3.1 First-order vs. second-order effects.

1. *Direct update on  $f$*  (Step 1) is a *first-order* improvement. By a first-order Taylor expansion Rudin et al., 1964:

$$f(\theta_1 - \mu \nabla f(\theta_1)) \approx f(\theta_1) - \mu \|\nabla f(\theta_1)\|^2,$$

indicating a change proportional to  $\mu \|\nabla f\|$ .

2. *Change in  $f$  due to a step on  $g$*  (Step 3) is a *second-order* Lipschitz effect. If  $\Delta\theta_1 = -\mu \nabla g$ , then by  $L_f$ -smoothness:

$$|\Delta f| = |f(\theta_1 + \Delta\theta_1) - f(\theta_1)| \leq \frac{L_f}{2} \|\Delta\theta_1\|^2 = \mathcal{O}(\mu^2).$$

Hence, the first-order decrease from Step 1 (of order  $\mu$ ) dominates the second-order bump (of order  $\mu^2$ ) for sufficiently small  $\mu$ . In practice, one often takes  $\mu < 1/(L_f + L_g)$  so that  $\mu^2 \ll \mu$ . Similarly, Steps 1–2 cannot undo progress on  $g$  from Steps 3–4.

**4. Conclusion: Non-increase of  $\mathcal{L}$  in one full iteration.** Therefore, if  $\mu$  is small enough,

$$f(\theta_1^{(k+1)}, \theta_2^{(k+1)}) + g(\theta_1^{(k+1)}, \theta_3^{(k+1)}) \leq f(\theta_1^{(k)}, \theta_2^{(k)}) + g(\theta_1^{(k)}, \theta_3^{(k)}).$$

As a result,

$$\mathcal{L}(\theta_1^{(k+1)}, \theta_2^{(k+1)}, \theta_3^{(k+1)}) \leq \mathcal{L}(\theta_1^{(k)}, \theta_2^{(k)}, \theta_3^{(k)}).$$

Since  $\mathcal{L}$  is lower-bounded (non-negative), this non-increasing sequence converges to a finite limit.  $\square$

## B Visualizing the distribution shift between short and long text

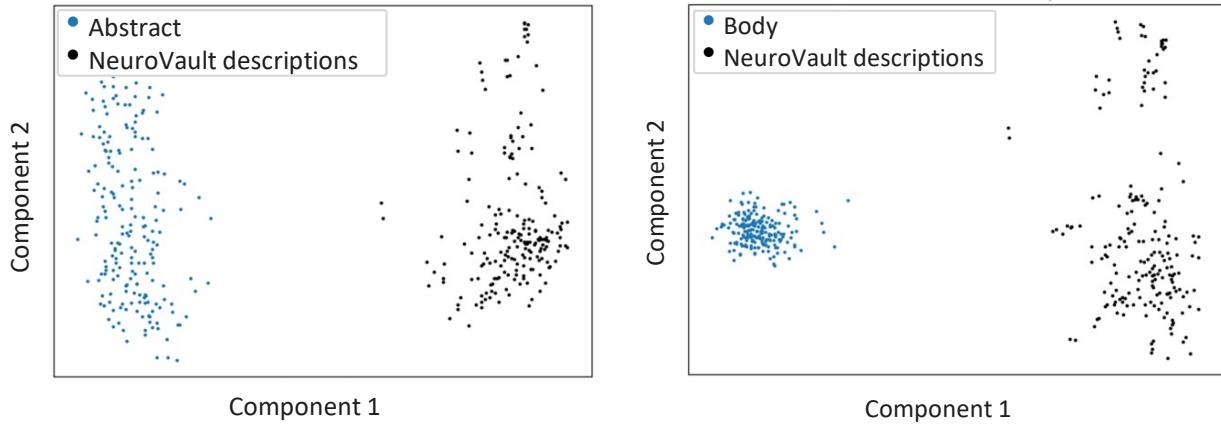


Figure 1: PCA decomposition of LLM embeddings comparing NeuroVault contrast definitions (short text) with abstracts and full-body texts (long text). The left plot shows a clear distribution shift along the first component between NeuroVault descriptions and abstracts, while the right plot shows a more pronounced shift along both the first and second components between NeuroVault descriptions and full-body texts. This indicates a significant distribution shift between short NeuroVault descriptions and full-length articles.

## C Prompt for augmenting NeuroVault contrast descriptions

Below is the prompt we used to extend NeuroVault contrast descriptions with GPT-4o:

I have 169 descriptions of the NeuroVault contrasts dataset for the brain. On the other hand, I have trained an autoencoder to encode the brain text and images from neuroscientific articles and to decode images from the text's latent (encoded) representation. Now, I want to use my trained autoencoder to encode the NeuroVault descriptions and decode (estimate) the NeuroVault brain contrasts. However, the NeuroVault descriptions are very short, while my autoencoder networks have been trained on lengthy neuroscientific articles. Therefore, I want to extend the NeuroVault descriptions to resemble the body of an article, including sections like introduction, framework/method, experiments, discussion, conclusion, etc., similar to an 8-page or longer article. I will provide you with the NeuroVault descriptions, and I would like you to extend them for me. I want the text to be as long as possible and to include all the relevant key terms related to the description.

## D The impact of LLM-based text augmentation on the performance of the retrieval task

Table 1: Retrieval task on NeuroVault: the impact of NeuroConText contrastive approach and LLM-based text augmentation. Augmented descriptions consistently improve the performance of NeuroConText (both in latent and decoded forms) across all metrics—Recall@10, Recall@100, and Mix&Match—indicating that LLM-based text augmentation enhances the model’s ability to retrieve relevant brain maps. For NeuroConText, augmented descriptions result in significant gains in Recall@10 (+9%), Recall@100 (+16%) and Mix&Match (+11%) compared to original descriptions. Similarly, NeuroConText (decoded) benefits from augmentation, showing improvements of +10%, +7%, and +8% across the three metrics, respectively. In contrast, NeuroQuery and Text2Brain do not show significant improvements, with Text2Brain showing slight declines across all metrics when using augmented descriptions.

Method	Description Type	Metric [%] ↑		
		Recall@10	Recall@100	Mix&Match
<b>NeuroConText</b> (in latent)	Augmented description	13	71	64
	Original description	4	55	53
	Difference (aug. - org.)	9	16	11
<b>NeuroConText</b> (decoded)	Augmented description	20	70	66
	Original description	10	63	58
	Difference (aug. - org.)	10	7	8
<b>NeuroQuery</b>	Augmented description	20	70	65
	Original description	20	67	63
	Difference (aug. - org.)	0	3	2
<b>Text2Brain</b>	Augmented description	12	61	58
	Original description	13	67	62
	Difference (aug. - org.)	-1	-6	-4

## E Prompt used for article summarization

We used a customized prompt to extract essential information about brain regions and associated cognitive processes to generate summaries of neuroscientific articles. The prompt instructed the latest version of the Mistral language model to act as a neuroscientific expert, asking relevant questions about the article and providing answers. Below is the exact prompt used:

You are a neuroscientific expert. Your task is to write a summary of a neuroscientific article. When presented with the article, come up with interesting questions to ask and answer each question. Afterward, combine all the information and write a summary.

article: {essay}

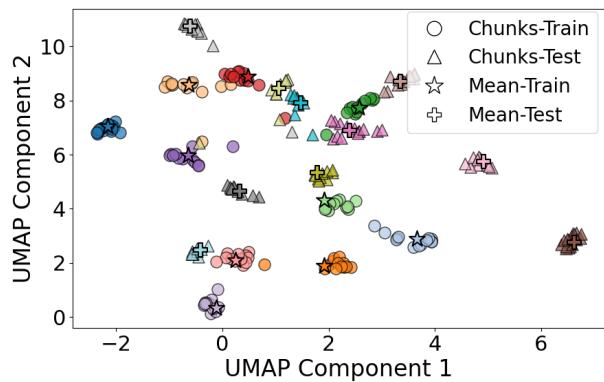
Instructions:

\*Summarize: In clear and concise language, summarize the key points and themes presented in the article.

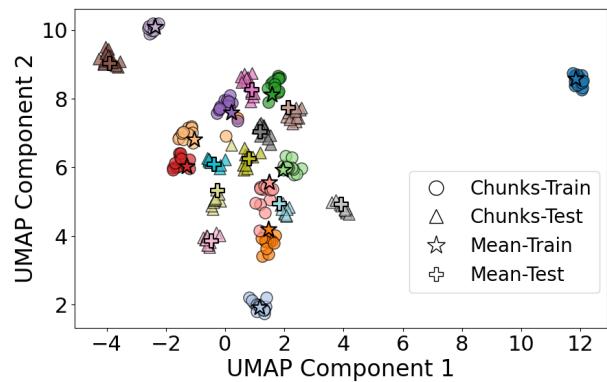
\*Write a short summary: Write a short summary in 2 sentences.

\*The brain regions: Write the brain regions and any cognitive or affective processes related to brain activation.

## F UMAP visualizations of the chunks' LLM embeddings



(a) Original Mistral-7B embeddings space



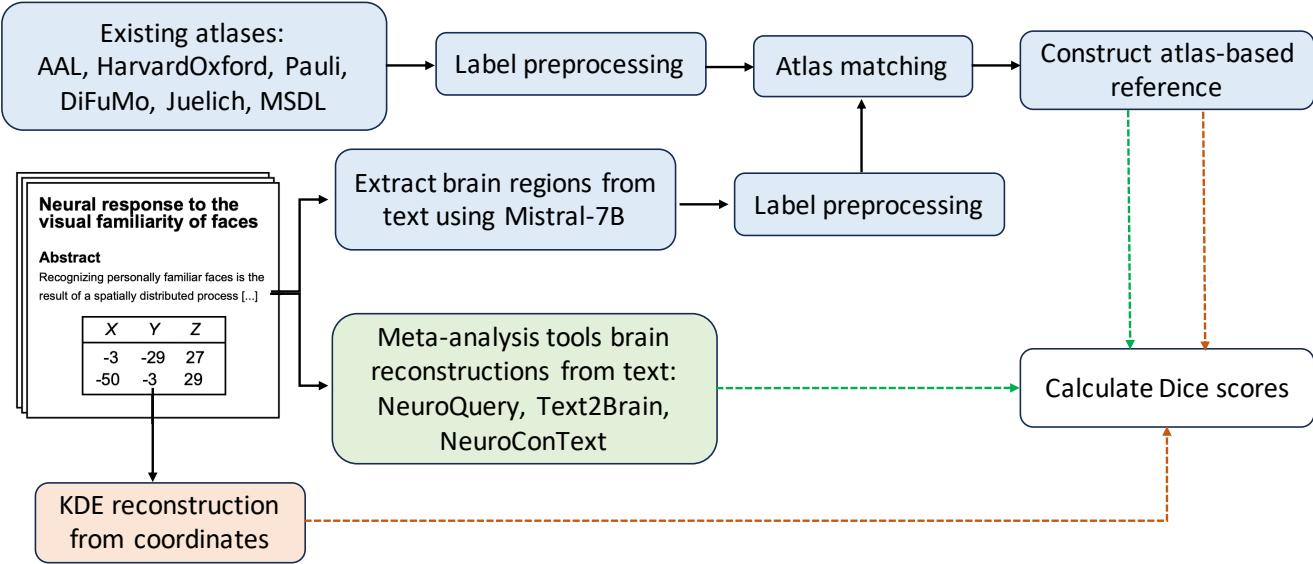
(b) NeuroConText text latent space

**Figure 2: UMAP visualizations of the top 10 articles with the highest number of chunks.** The left figure shows the original Mistral-7B embeddings and the right figure shows the encoded text embeddings in the shared latent space of NeuroConText model. Each color corresponds to the chunks of a given article. In both representations, chunks corresponding to the same article are closely clustered around their mean, while the means of different articles remain well-separated. This pattern supports the effectiveness of the averaging strategy, as it preserves proximity within an article's chunks to their mean while maintaining clear separation from the chunks of other articles in the latent space.

## **G Pipeline and results to assess meta-analysis tool's ability to mitigate sparse coordinate reporting challenges in CBMA**

To evaluate how well meta-analysis tools reconstruct brain activation patterns discussed in the literature, we first construct a reference map for each article based on anatomical labels. Using Mistral-7B, we extract all brain regions mentioned in each article. We then match these extracted regions to labels from six atlases available in Nilearn (AAL, Harvard-Oxford, DiFuMo, Juelich, MSDL, and Pauli), applying some preprocessing and normalization to both sets of labels (atlases' labels and extracted articles' regions) to improve alignment (Table 2). Hemisphere sensitivity is also considered during matching, e.g. regions like "left Hippocampus" are only compared to atlas components from the corresponding hemisphere, according to the rules in Table 3. Note that we consider more directional restrictions when matching article region names to atlas labels; our pipeline considers that all terms (words) from the article region must appear in the atlas label after standardization. This includes directional terms like "superior", "inferior", "anterior", and "posterior". If such a term is present in the article region but missing in the atlas label, the match fails. Table 4 provides an example for matching with directional terms. With this preprocessing, we obtained 30.21% coverage of the articles' regions by the 6 atlas labels. Fig. 4 presents the distribution of matched brain region labels across different atlases, showing that the majority of matches (over 90%) originate from the DiFuMo atlas. Also, the list of 30 most frequently matched labels (all associated with DiFuMo atlas), and the 30 least frequent matched labels are shown in Figure 5 and Table 5, respectively.

After matching the atlas labels with the article regions, we average the matched atlas components to form a reference brain map that represents the reported regions for each article. We compute the Dice score between this atlas-based reference map and the reconstructed maps obtained from different meta-analysis tools, as well as the KDE representation of activation coordinates. Meta-analysis tools (NeuroConText, NeuroQuery, and Text2Brain) consistently achieve higher Dice scores than KDE, suggesting their superior ability to capture activation patterns and mitigate limitations due to sparse coordinate reporting in CBMA.



**Figure 3: Atlas-based reference map framework to assess the reconstruction performance of meta-analysis tools:** This framework illustrates the pipeline for evaluating brain map reconstructions from meta-analysis tools using atlas-based reference maps. Brain regions are first extracted from article text using Mistral-7B. After label preprocessing, these extracted regions are matched to components from six brain atlases (AAL, Harvard-Oxford, DiFuMo, Juelich, MSDL, Pauli). The matched components are averaged to construct a reference map for each article. Separately, reconstructed maps are obtained from meta-analysis tools (NeuroQuery, Text2Brain, NeuroConText) as well as KDE. Dice scores are then computed between each reconstructed map and the corresponding atlas-based reference. The blue blocks highlight the steps involved in constructing the atlas-based reference maps.

Table 2: Text preprocessing rules for atlas labels and article regions.

Before	After
LH, lh, _lh, _l	left
RH, rh, _rh, _r	right
hemi, h, hemisphere	hemisphere
gyri	gyrus
cortex	(removed)
lobe	(removed)
hemisphere	(removed)
Extra spaces	Removed
All text	Lowercased

In Fig. 7, we show three qualitative examples to compare the atlas-based reference maps with the KDE map and the reconstructed map generated by NeuroConText. These examples correspond to the articles for which NeuroConText achieved the best, median, and worst Dice scores. The first row displays KDE-based reconstructions, the second row shows the atlas-based reference maps averaged over six atlases, the third row shows the reference using only the DiFuMo atlas, and the final row presents NeuroConText-generated

Table 3: Hemisphere matching between article regions and atlas components

Article region hemisphere	Assigned atlas component hemisphere.
right	right
left	left
left and right	left, right, right & left, non
non	left, right, right & left, non

Table 4: Examples of region matching with directional terms.

Article Region	Atlas Label	Match?
superior frontal gyrus	superior frontal gyrus	Yes
superior frontal gyrus	frontal gyrus	No
frontal gyrus	superior frontal gyrus	Yes
posterior cingulate	posterior cingulate cortex	Yes
posterior cingulate	cingulate cortex	No
inferior parietal lobule	inferior parietal lobule	Yes
inferior parietal lobule	parietal lobule	No

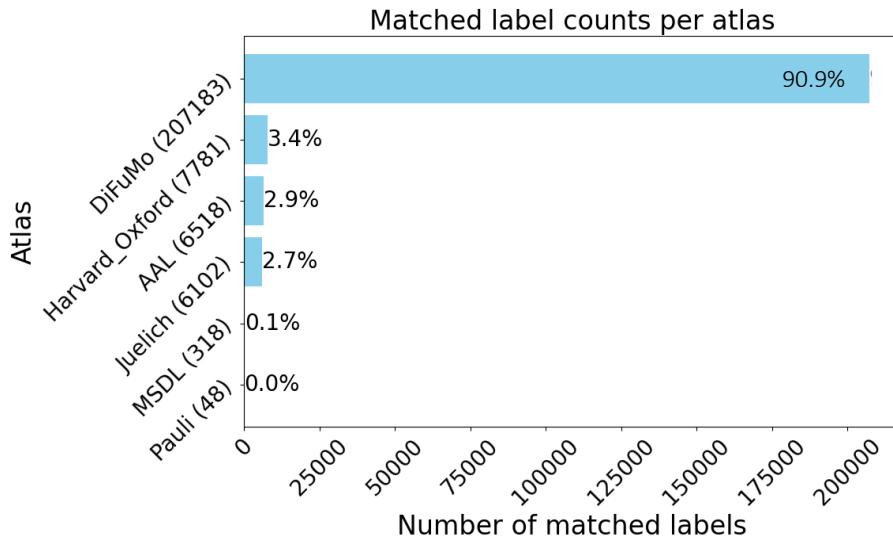


Figure 4: Number of matched brain regions per atlas across all articles.

maps. Tables 6, 7, and 8 report the brain regions extracted from the articles using a large language model (LLM), followed by the matched atlas labels for Articles 1, 2, and 3, respectively. The matching was performed across six different atlases. The full lists of atlas-matched regions are available in the NeuroConText GitHub repository.

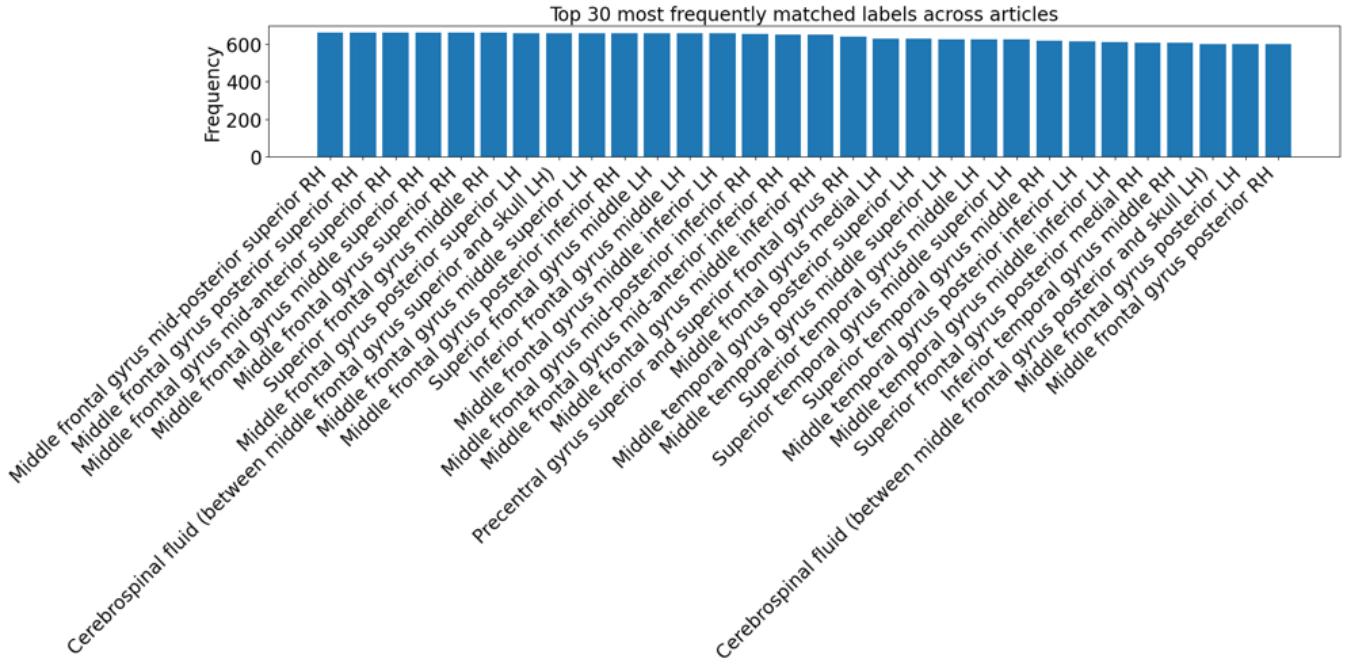
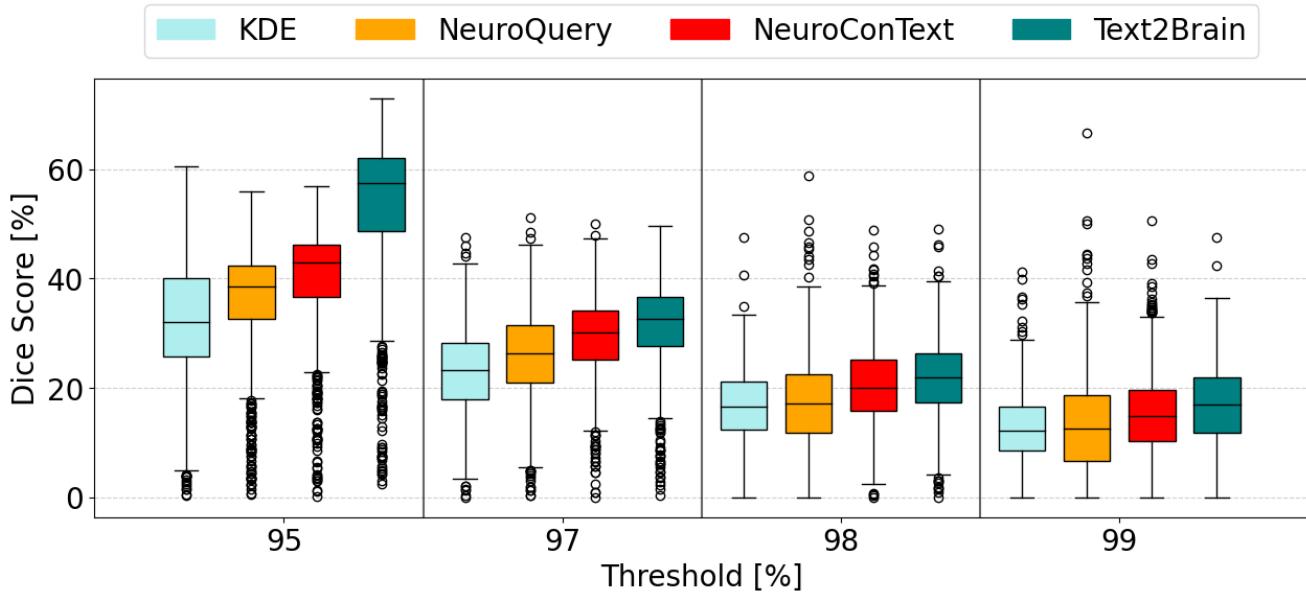


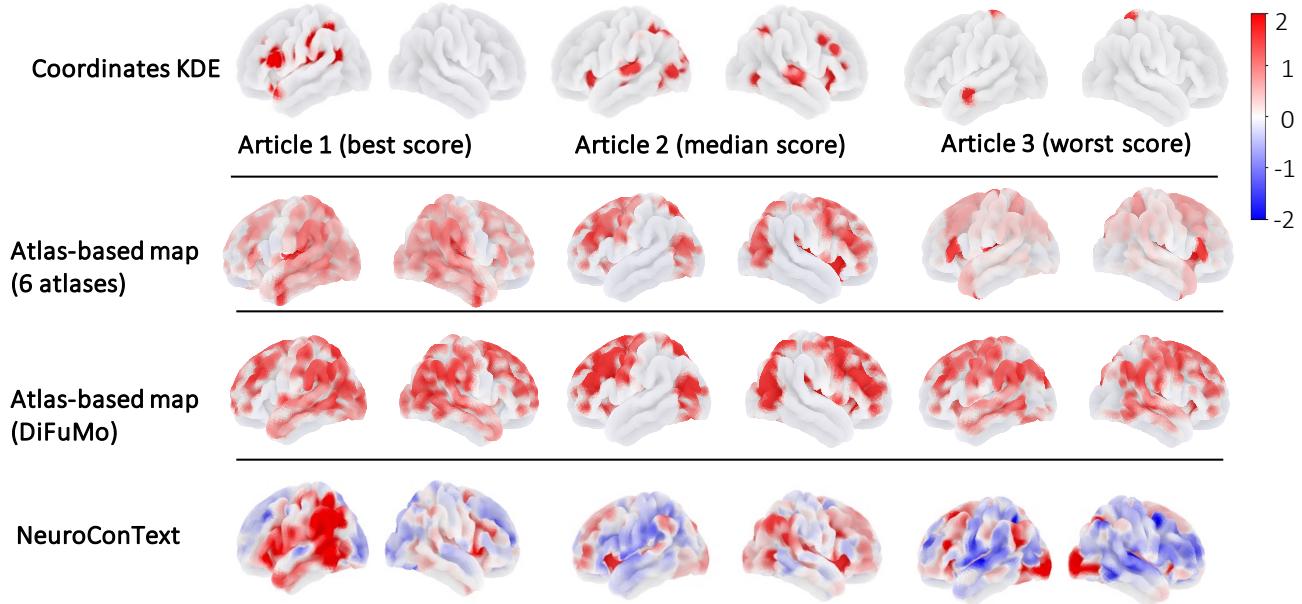
Figure 5: List of 30 most frequently matched labels.



**Figure 6: Dice similarity between atlas-based reference maps and reconstructed brain maps:** Dice scores are computed between atlas-based reference maps (constructed by matching article-extracted region names to atlas components) and reconstructed brain maps generated by KDE, NeuroQuery, Text2Brain, and NeuroConText. Scores are shown across different threshold levels (95–99%) applied to the reconstructed maps. Meta-analysis tools consistently yield higher Dice scores than KDE, especially at lower thresholds. This indicates meta-analysis tools provide better overlap with the region-based reference.

Table 5: 30 least frequent matched labels with their corresponding atlases

Label	Count	Atlas
Vermis_10	2	AAL
Vermis_1_2	2	AAL
Vermis_3	2	AAL
Vermis_4_5	2	AAL
Vermis_6	2	AAL
Vermis_7	2	AAL
Vermis_8	2	AAL
Vermis_9	2	AAL
Suborbital sulcus	2	DiFuMo
WM Acoustic radiation	2	Juelich
WM Uncinate fascicle	2	Juelich
Broca	2	MSDL
Med DMN	2	MSDL
STH	2	Pauli
GM Amygdala_centromedial group	3	Juelich
GM Amygdala_laterobasal group	3	Juelich
GM Amygdala_superficial group	3	Juelich
Basal	3	MSDL
Cereb	3	MSDL
Cing	3	MSDL
Front DMN	3	MSDL
Occ post	3	MSDL
Striate	3	MSDL
Sup Front S	3	MSDL
Vis	3	MSDL
Ca	3	Pauli
EXA	3	Pauli
GPe	3	Pauli
GPi	3	Pauli
HN	3	Pauli



**Figure 7: Qualitative analysis to asses NeuorConText's ability to mitigate the sparse coordinate limitation of CBMA:(a) Visualization of reconstructed brain maps for three representative articles with the best, median, and worst Dice scores. The first row shows KDE-based reconstructions, the second row presents the atlas-based reference maps averaged over 6 different atlases, the third row presents the atlas-based reference maps based on only DiFuMo atlas, and the last row displays NeuroConText-generated maps. The comparison highlights that NeuroConText captures broader activation patterns compared to KDE, aligning more closely with the atlas-based reference.**

Table 6: Brain regions and atlas label matching for **Article 1 - (best Dice score)**. The full list is available in the NeuroConText GitHub repository for [PMID 30210434](#).

Source	Sample Regions
Article regions extracted via Mistral LLM	<p><b>Left Ventro-dorsal Stream:</b> Inferior parietal lobe (incl. Supramarginal gyrus), Inferior frontal gyrus (ventral premotor cortex, opercular, triangular), Precentral gyrus, Rolandic operculum, Insula.</p> <p><b>Left Frontal Cortex:</b> Precentral gyrus, Superior/Middle/Inferior frontal gyri (incl. opercular, triangular, orbital), Rolandic operculum.</p> <p><b>Left Parietal Cortex:</b> Postcentral gyrus, Superior and Inferior parietal lobules (incl. Supramarginal and Angular gyri).</p> <p><b>Left Temporal Cortex:</b> Superior and Middle temporal gyri, Superior temporal pole.</p> <p><b>Subcortical Regions:</b> Putamen, Pallidum, Thalamus.</p> <p><b>Other:</b> Left Insula, Left Angular Gyrus.</p>
AAL	Pallidum_L, Pallidum_R, Putamen_L, Putamen_R, Thalamus_L, Thalamus_R
Harvard-Oxford	Angular Gyrus, Frontal Medial Cortex, Intracalcarine Cortex, Middle Temporal Gyrus (anterior division), Planum Temporale, Superior Frontal Gyrus
DiFuMo (1024)	Angular gyrus anterior LH, Frontal pole medial, Fusiform gyrus posterior RH, Lingual gyrus posterior LH, Middle frontal gyrus mid-anterior RH, Thalamus superior. <i>Full region mapping available in NeuroConText GitHub repository.</i>
Juelich	GM Parietal operculum OP1–OP4, GM Superior parietal lobule 7A, 7PC

Table 7: Brain regions and atlas label matching for **Article 2 - (median Dice score)**. The full list is available in the NeuroConText GitHub repository for *PMID 36166644*.

Source	Sample Regions
<b>Article regions extracted via Mistral LLM</b>	<p><b>Frontal Regions:</b> Frontal to central areas, Medial frontal area, Middle frontal gyrus (incl. right), Inferior frontal gyrus, Anterior insula (incl. dorsal anterior and right dorsal anterior insula), Precentral gyrus, Rolandic operculum.</p> <p><b>Parietal Regions:</b> Inferior and Superior parietal lobules, General parietal areas.</p> <p><b>Temporal Regions:</b> Superior temporal gyrus (bilateral and posterior), Superior temporal plane (STP), Planum temporale (PT), Temporo-parieto-occipital junction (TPO).</p> <p><b>Occipital Regions:</b> Occipital areas, Inferior occipital gyrus, Visual cortex.</p> <p><b>Cingulate Regions:</b> Anterior cingulate cortex (ACC), Dorsal ACC (dACC), Midcingulate cortex (MCC).</p> <p><b>Other:</b> Fusiform gyrus, Optic nerve, Optic chiasm, Salience network, Superior olivary nucleus, Right/Left-/Bilateral hemispheres, Contralateral and Ipsilateral attention systems.</p>
<b>AAL</b>	Insula_L, Insula_R
<b>DiFuMo (1024)</b>	Anterior cingulate cortex anterior LH, Anterior insula (multiple subdivisions), Middle frontal gyrus (multiple subdivisions), Superior parietal lobule (multiple subdivisions), Dorsolateral prefrontal cortex LH, Fusiform gyrus RH. <i>Full region mapping available in the repository.</i>

Table 8: Brain regions and atlas label matching for **Article 3 - (worst Dice score)**. The full list is available in the NeuroConText GitHub repository for *PMID 30885230*.

Source	Sample Regions
<b>Article regions extracted via Mistral LLM</b>	<b>Limbic and Subcortical:</b> Amygdala, Hippocampus (bilateral), Anterior cingulate cortex (ACC), Middle cingulate cortex. <b>Frontal Regions:</b> Dorsolateral prefrontal cortex (DLPFC), General prefrontal cortex, Frontal areas. <b>Parietal Regions:</b> Intraparietal sulcus (IPS), Paracentral gyrus (bilateral). <b>Temporal and Occipital Regions:</b> Visual number form (posterior inferior temporal gyrus), Left temporal gyrus. <b>Other:</b> Insula.
<b>AAL</b>	Amygdala_L, Amygdala_R, Hippocampus_L, Hippocampus_R, Insula_L, Insula_R, Caudate_L, Caudate_R, Cuneus_L, Cuneus_R, Precuneus_L, Precuneus_R, Putamen_L, Putamen_R, Thalamus_L, Thalamus_R
<b>Harvard-Oxford</b>	Angular Gyrus, Insular Cortex, Lingual Gyrus, Middle Frontal Gyrus, Occipital Fusiform Gyrus, Paracingulate Gyrus, Postcentral Gyrus, Precentral Gyrus, Superior Frontal Gyrus
<b>DiFuMo (1024)</b>	Amygdala (anterior/posterior), Anterior insula (multiple subdivisions), Caudate (multiple subdivisions), Hippocampus posterior RH, Cuneus (anterior/posterior/superior), Middle and Superior frontal gyri (multiple subdivisions), Angular gyrus, Fusiform gyrus posterior RH. <i>Full region mapping available in the repository.</i>
<b>Juelich</b>	GM Hippocampus (cornu ammonis, dentate gyrus, subiculum), GM Insula (Id1, Ig1, Ig2), GM Hippocampus entorhinal cortex, GM Hippocampus-amygdaloid transition area