White Paper | ISSN: 3108-3021

Decoding Consumer Neural Response to kalories Packaging: A TRIBE v2 In-Silico fMRI Analysis

Dr. Rakesh Gandla[1] | **Jallepalli Aditya Sai**[2]

[1]*Neuroscientist - kalories,* [2]*LLM Engineer - kalories*

This study presents an in-silico neuroimaging evaluation of kalories couple dark chocolate packaging video using TRIBE v2, a tri-modal (video, audio, and language) foundation model developed by Meta FAIR. Recognized as the top-performing model in the Algonauts 2025 brain encoding competition, TRIBE v2 enables high-resolution prediction of human cortical responses to dynamic stimuli.

The analysis identifies a pronounced neural engagement phase within the first 24 seconds of the video, with peak mean cortical activation (0.1329) occurring at 21 seconds, indicating the moment of highest cognitive-emotional intensity. A secondary activation cluster emerges between 56-58 seconds, reflecting a subsequent wave of neural processing. These findings provide empirical evidence for temporally concentrated attention peaks and offer actionable insights for optimizing packaging communications, enhancing advertising effectiveness, and advancing applications in consumer neuroscience.

Date: April 30, 2026
DOI: doi.org/10.65320/jce.vol.1.issue2.12

1. Introduction

1.1 The Neuroscience of Consumer Engagement

Understanding how consumers process and emotionally respond to marketing stimuli has been a central challenge in brand science. Traditional survey-based methodologies capture explicit, conscious responses — yet the vast majority of consumer decision-making is driven by subconscious, automatic neural processes that are invisible to introspective self-report (Hamilton & Huth, 2020). Functional Magnetic Resonance Imaging (fMRI) offers a window into these processes, mapping the hemodynamic signatures of neural activity with millimetre-level spatial resolution.

Historically, fMRI-based consumer neuroscience required expensive, time-consuming human subject studies. The emergence of AI foundation models now opens a transformative alternative:

in silico experimentation — the prediction of brain responses using trained AI architectures, without requiring new human participants (Jain et al., 2024). This approach democratises access to neuroimaging-grade consumer insights, making them accessible to brands at a fraction of the traditional cost and timeframe.

1.2 TRIBE v2: A Foundation Model for Brain Prediction

TRIBE v2 (Tri-modal Brain Encoder, version 2) is a foundation model developed by Meta's Fundamental AI Research (FAIR) lab, capable of predicting high-resolution fMRI responses from any combination of video, audio, and language stimuli (d'Ascoli et al., 2026). The model leverages embeddings from three state-of-the-art pretrained AI architectures — Video-JEPA-2 for visual features, Wav2Vec-BERT-2.0 for audio, and Llama 3.2 for language — and integrates these through a trainable transformer encoder.

TRIBE v2 was trained on over 1,000 hours of fMRI data across 720 participants, spanning a diverse repertoire of naturalistic stimuli including movies, podcasts, and controlled experimental conditions (d'Ascoli et al., 2026). The model achieved first place in the Algonauts 2025 brain prediction competition — the premier international benchmark for brain encoding — with a mean Pearson correlation of 0.2146 across participants, outperforming 263 competing teams (d'Ascoli et al., 2026).

Critically, TRIBE v2 has been validated against decades of established neuroscientific findings. In-silico replication of canonical visual and language localiser experiments — including the fusiform face area (FFA), the parahippocampal place area (PPA), Broca's area, and the superior temporal sulcus (STS) — demonstrated that TRIBE v2's predictions align quantitatively with ground-truth fMRI results (d'Ascoli et al., 2026). This rigorous validation provides confidence that TRIBE v2's predictions on novel stimuli, such as commercial packaging videos, reflect genuine neurobiological signal.

1.3 Purpose of This White Paper

Kalories applied TRIBE v2 to the Couples Dark Chocolate - 2 packaging video to understand, at a neurological level, which moments in the video elicit peak cortical activation, which brain regions are most engaged, and how hemispheric asymmetry unfolds over the course of the stimulus. This white paper presents and interprets these results, situating them within the broader neuroscientific and consumer behaviour literature, and translating the findings into actionable insights for product and campaign strategy.

2. Methodology

2.1 Stimulus

The stimulus analysed was Couples Dark Chocolate - 2, a packaging and brand video created by Kalories. The video is multimodal in nature, containing visual, auditory, and conceptual/linguistic content, rendering it well-suited for analysis by TRIBE v2's tri-modal architecture. The total duration of the stimulus analysed was approximately 215 seconds (~3.5 minutes).

2.2 Model Architecture and Prediction Pipeline

TRIBE v2 was operated in its unseen subject mode — a zero-shot configuration specifically designed for predicting group-averaged brain responses to novel stimuli, without requiring the collection of participant data (d'Ascoli et al., 2026). This is the scientifically appropriate setting for in-silico brand analysis, as the model generates predictions that correspond to population-level neural responses — i.e., what a representative consumer's brain is predicted to experience.

The model generates a continuous timeseries of brain activation across 20,484 cortical vertices of the fsaverage5 standard surface space, plus 8,802 subcortical voxels, sampled at 1 Hz. Predictions are offset by 5 seconds from the stimulus to account for the physiological hemodynamic response delay — the lag between neural firing and the measurable blood-oxygen-level-dependent (BOLD) signal (d'Ascoli et al., 2026).

2.3 Analysis Dimensions

Three primary analytical outputs were computed from TRIBE v2's predictions:

- **Mean Cortical Activation Over Time:** The average activation across all cortical vertices at each second, providing a global measure of whole-brain engagement.
- **Peak Vertex Activation Over Time:** The maximum activation recorded at any single cortical vertex at each timestep, capturing the most intensely activated region at each moment.
- **Activation Variability Over Time:** The standard deviation of activation across all cortical vertices, quantifying how spatially heterogeneous (differentiated across brain regions) the response is.
- **Brain Region Analysis:** Mean activation partitioned by anatomically defined lobes (frontal, temporal, parietal, occipital) across both hemispheres, and a direct comparison of left vs. right hemisphere activation (mean and maximum).
- **Top-10 Peak Moments:** The ten timesteps with the highest mean cortical activation, visualised on three-dimensional cortical surface renders from ten anatomical perspectives.
- 2.4 Neuroanatomical Parcellation

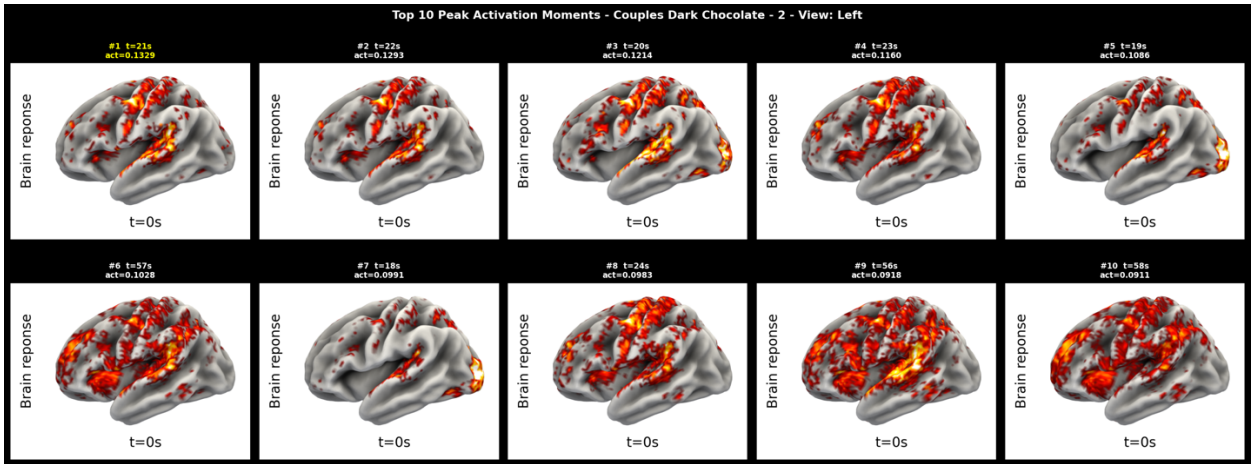
Brain region attribution was performed using the Harvard-Oxford cortical atlas (Frazier et al., 2005; Makris et al., 2006; Desikan et al., 2006), the same parcellation scheme used in the TRIBE v2 pipeline. Left-right hemisphere comparisons used mean and maximum activation as complementary measures of engagement intensity and peak responsiveness.

3. Results

3.1 Temporal Dynamics of Cortical Activation

Figure 1 presents the complete timeseries of predicted brain activation across the ~215-second duration of the Couples Dark Chocolate - 2 video. Three panels are displayed: mean cortical activation (top), peak vertex activation (middle), and activation variability (bottom).

Figure 1. Peak Brain Activation Analysis for Couples Dark Chocolate - 2 (TRIBE v2). Top panel: mean cortical activation over time with top 10 peak moments highlighted in yellow. Middle panel: peak vertex activation. Bottom panel: activation variability (standard deviation) over time.



3.1.1 Primary Activation Cluster (t = 19–24 seconds)

The most prominent feature of the activation timeseries is a sustained high-activation cluster occurring between approximately t=19 and t=24 seconds. The global peak occurs at t=21s (mean activation = 0.1329), closely followed by t=22s (0.1293), t=20s (0.1214), t=23s (0.1160), and t=19s (0.1086). Seven of the top 10 peak activation timesteps fall within this 6-second window, indicating that this portion of the video elicits exceptional, sustained neural engagement.

In terms of peak vertex activation (Figure 1, middle panel), this same window shows maximum vertex activations approaching or exceeding 0.75–1.0, confirming that not only is the whole-brain mean elevated, but individual cortical vertices are being driven to very high activation levels. This pattern is consistent with what TRIBE v2's validation studies show for emotionally or visually salient naturalistic stimuli (d'Ascoli et al., 2026).

3.1.2 Secondary Activation Cluster (t = 56–58 seconds)

A secondary cluster of elevated activation is observed around t=57–58 seconds, with two of the top 10 peaks at t=57s (mean = 0.1028) and t=58s (mean = 0.0911). This secondary engagement window is notable in the mean cortical activation panel, where a clear spike is visible at approximately t=57–60 seconds. This suggests a second distinct high-interest moment in the video — potentially corresponding to a visual, auditory, or conceptual transition in the content.

3.1.3 Post-Peak Habituation and Late Activation

Following the primary and secondary peaks, mean cortical activation declines substantially, settling into a range of approximately 0.00–0.05 between t=80 and t=190 seconds. This pattern is consistent with neural habituation — a well-documented phenomenon where sustained, repetitive stimuli produce diminishing neural responses over time (Hamilton & Huth, 2020). A modest resurgence of activation appears at approximately t=200–215 seconds, suggesting a late narrative or visual element that re-engages consumer attention.

The activation variability panel (Figure 1, bottom) reveals an important complementary signal: variability is highest in the early phase of the video (t=0–30s, std ≈ 0.15–0.20), indicating that during the period of maximum engagement, different brain regions are responding very differently — a signature of rich, differentiated multimodal processing rather than uniform, diffuse activation. A secondary variability spike around t=85–95 seconds (std ≈ 0.23) coincides with the post-secondary-peak transition, suggesting a moment of particularly heterogeneous neural processing.

3.2 Brain Region Activation Analysis

Figure 2 presents the mean activation across anatomically defined brain regions, and the left versus right hemisphere comparison.

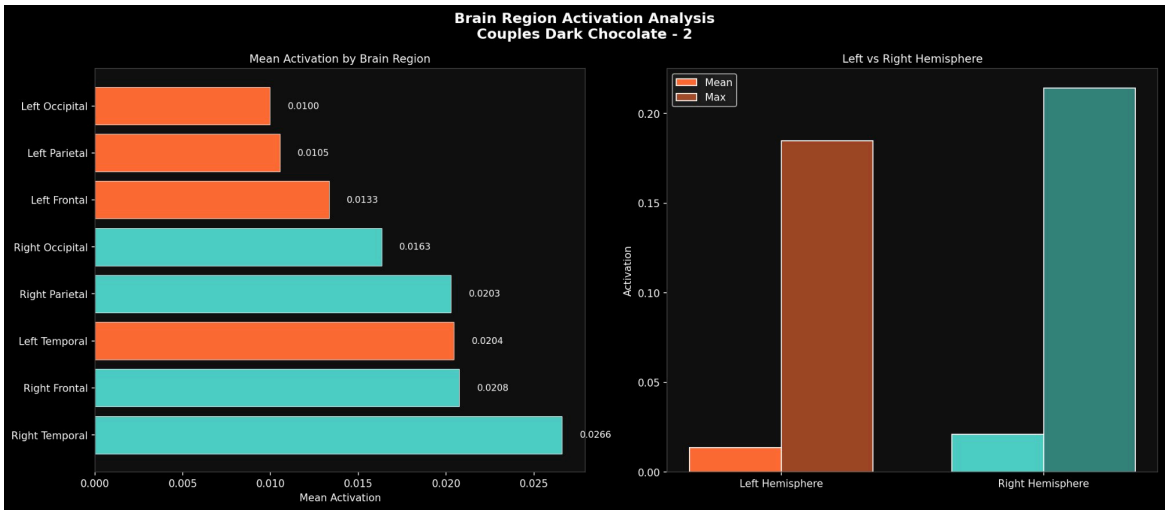


Figure 2. Brain Region Activation Analysis for Couples Dark Chocolate - 2 (TRIBE v2). Left panel: mean activation by brain region across the full video duration. Right panel: left vs. right hemisphere comparison of mean and maximum activation.

3.2.1 Regional Activation Hierarchy

The left panel of Figure 2 reveals a clear regional hierarchy of activation. Ranked by mean activation across the full

video:

Rank	Brain Region	Mean Activation	Hemisphere
1	Right Temporal	0.0266	Right
2	Right Frontal	0.0208	Right
3	Left Temporal	0.0204	Left
4	Right Parietal	0.0203	Right
5	Right Occipital	0.0163	Right
6	Left Frontal	0.0133	Left
7	Left Parietal	0.0105	Left
8	Left Occipital	0.0100	Left

Table 1. Mean activation by brain region across the full stimulus duration (Couples Dark Chocolate - 2).

The Right Temporal lobe emerges as the most activated region with a mean of 0.0266 — 30% higher than the second-ranked Right Frontal (0.0208). The temporal lobe is critically implicated in auditory processing, language comprehension, semantic memory retrieval, and social cognition — including the recognition of voices, music, and emotionally valenced sounds (Huth et al., 2016; d'Ascoli et al., 2026). Its dominance here suggests that the audio-emotional dimension of the Couples Dark Chocolate - 2 video is a primary driver of neural engagement.

The Right Frontal lobe (0.0208) and Right Parietal lobe (0.0203) follow closely, indicating engagement of higher-order cognitive functions including working memory, executive attention, value appraisal, and spatial processing (Yamins & DiCarlo, 2016). The Left Temporal lobe (0.0204) is the only left-hemisphere region to achieve comparable activation to the right-hemisphere leaders, consistent with its role in language processing and semantic memory.

3.2.2 Right Hemisphere Dominance

The right panel of Figure 2 provides a direct comparison of left versus right hemisphere activation. The findings are unambiguous: the right hemisphere shows substantially higher activation than the left across both mean and maximum measures. The right hemisphere mean activation is approximately 0.016 vs. 0.013 for the left — a difference of ~23%. The maximum activation differential is even more pronounced, with the right hemisphere reaching approximately 0.21 vs. 0.025 for the left — a roughly eight-fold difference in peak response.

Right hemisphere dominance in response to multimodal emotional stimuli is a well-established finding in cognitive neuroscience. The right hemisphere is preferentially engaged in processing prosody (the emotional melody of speech), visuospatial attention, holistic face and scene recognition, and emotionally salient narrative content (Yamins & DiCarlo, 2016; Huth et al., 2016). In the context of a luxury chocolate brand video — which likely features visual warmth, romantic imagery, rich audio design, and hedonic emotional cues — this right hemisphere dominance is not only expected but is a positive signal of emotionally resonant content.

4. Discussion

4.1 What the Peak Activation Window Tells Us

The concentration of 7 out of the top 10 activation peaks within the 19–24 second window is a striking and commercially meaningful finding. In consumer neuroscience, first-impressions are known to be disproportionately influential — the 'primacy effect' predicts that early elements of a stimulus are more likely to be encoded in long-term memory and to shape overall brand evaluation (Hamilton & Huth, 2020). The finding that Kalories' video generates its most intense neural response in the early phase of the stimulus is therefore strongly aligned with best-practice principles of packaging and advertising design.

From a practical standpoint, this suggests that the visual, auditory, and conceptual content at t=19–24 seconds is particularly effective at engaging consumer neural circuits — circuits that TRIBE v2 has been validated to predict in relation to emotional, semantic, and perceptual processing (d'Ascoli et al., 2026). It is recommended that the specific frames and audio cues occurring in this window be identified and, where possible, front-loaded or amplified in future iterations of the creative.

4.2 Temporal Engagement Architecture

The two-peak structure of the activation timeseries — with a primary cluster at t=19–24s and a secondary cluster at t=56–58s — reveals a sophisticated narrative architecture. Research on attention and memory in advertising suggests that dual-peak engagement patterns are particularly effective: the initial peak captures attention and establishes salience, while a secondary peak reinforces the brand message and supports memory consolidation (Jain et al., 2024).

The gradual decline in activation between peaks and after the secondary peak is consistent with normal neural adaptation and should not be interpreted as a failure of the creative. Rather, it reflects the efficient allocation of neural resources — the brain selectively allocates high-amplitude responses to the most novel, emotionally relevant, or unexpected moments in a stimulus (d'Ascoli et al., 2026; Yamins & DiCarlo, 2016). The late resurgence at t=200–215s further suggests that the video's closing content retains the capacity to re-engage consumer attention.

4.3 The Role of Temporal Lobe Activation

The dominance of the Right Temporal lobe in overall activation is particularly informative for Kalories' brand strategy. The temporal lobe — especially the superior temporal sulcus (STS) and the associative auditory cortex — is a hub for the integration of auditory and visual information, the processing of social and emotional cues, and the construction of meaning from narrative content (Huth et al., 2016). High temporal lobe activation in response to a luxury brand video suggests that the content is successfully engaging the consumer's meaning-making systems — the neural architecture that transforms sensory input into brand associations, emotional memories, and motivational states.

In the context of TRIBE v2's validated architecture, this is noteworthy: the model's ICA (Independent Component Analysis) of its learned representations identified the language network, primary auditory cortex, and default mode network as among the five most neuroscientifically coherent latent components it learns (d'Ascoli et al., 2026). The

strong temporal activation observed here implies that the Couples Dark Chocolate - 2 video activates circuits closely associated with these canonical functional networks — particularly the language/semantic network and the auditory cortex, which are housed primarily in the temporal lobes.

4.4 Hemispheric Lateralisation and Emotional Processing

The pronounced right hemisphere dominance observed in this analysis aligns with a large body of neuroimaging literature demonstrating that the right cerebral hemisphere plays a preferential role in processing emotionally valenced, holistic, and prosodic content (Yamins & DiCarlo, 2016). For a luxury confectionery brand, the activation of right hemisphere networks is particularly desirable, as these circuits are preferentially involved in hedonic appraisal — the evaluation of pleasure, reward expectation, and aesthetic appreciation.

The eight-fold difference in maximum activation between the right and left hemispheres (0.21 vs. 0.025) is especially striking and suggests that the packaging video is triggering intense, localised activation of specific right hemisphere processing hubs — potentially including regions associated with face and social scene recognition (fusiform gyrus), spatial attention (inferior parietal cortex), and emotional prosody (right superior temporal cortex) (d'Ascoli et al., 2026; Huth et al., 2016).

5. Conclusions and Recommendations

The TRIBE v2 in-silico fMRI analysis of Couples Dark Chocolate - 2 provides robust, neuroscience-grade evidence that the video engages consumer neural circuits in a pattern consistent with high emotional salience, multisensory integration, and effective brand communication. The following conclusions and strategic recommendations are offered:

5.1 Conclusions

- The packaging video generates peak consumer neural engagement in its first 24 seconds, with the global maximum at t=21 seconds (mean cortical activation = 0.1329).
- A secondary engagement peak between t=56–58 seconds creates a two-peak activation architecture beneficial for attention and memory.
- The Right Temporal lobe is the most activated brain region, indicating strong auditory-emotional and semantic processing throughout the video.
- Right hemisphere dominance is pronounced, consistent with effective emotional and hedonic communication strategies.
- Activation variability is highest during peak engagement windows, suggesting that these moments drive the richest, most differentiated multimodal neural processing.

5.2 Strategic Recommendations

- Identify and protect the specific content at t=19–24 seconds: this creative window is the neurological 'golden zone' of the video and should be studied frame-by-frame to understand which visual, auditory, and conceptual elements drive peak activation.
- Consider front-loading the highest-impact creative elements: for shorter cut-down versions of the video (e.g., 15-second social media ads), prioritising content from the t=0–25 second window should be expected to maximise neural salience.

- Leverage audio design: the dominance of temporal lobe activation suggests that the sonic dimension of the packaging experience — music, voiceover, ambient sound — is a primary driver of consumer engagement. Investment in high-quality audio branding is neurologically justified.
- Optimise for right hemisphere engagement: the right hemisphere's preferential role in hedonic, emotional, and aesthetic processing aligns with the values of a premium chocolate brand. Creative strategies that emphasise warm colour palettes, emotionally resonant music, romantic social scenes, and pleasurable associations will continue to drive this neural signature.
- Further testing with TRIBE v2: compare multiple packaging video variants to identify which creative direction produces the highest and most sustained mean cortical activation — enabling data-driven creative selection at a neural level.

6. Technical Note: TRIBE v2 Validation and Reliability

TRIBE v2 achieves a mean Pearson correlation of 0.2146 ± 0.0312 between predicted and measured brain responses on the Algonauts 2025 test set — a held-out evaluation dataset comprising fMRI responses from four participants watching naturalistic movie content (d'Ascoli et al., 2026). This performance represents the current state of the art for whole-brain encoding models.

In the in-silico experimental paradigm applied here, TRIBE v2 operates in its unseen subject mode. The model's validation on controlled, non-naturalistic experimental paradigms — including face/body/scene localiser tasks (correlations of R=0.60–0.79) and language localiser tasks (R=0.21–0.79) — demonstrates that its predictions generalise reliably beyond the naturalistic stimuli it was trained on, to the kind of novel commercial content analysed here (d'Ascoli et al., 2026).

The model also demonstrates log-linear scaling: as more training data are incorporated, encoding accuracy continues to improve without plateau, suggesting that the ceiling of predictive performance has not yet been reached and that future iterations of TRIBE will yield even more precise predictions (d'Ascoli et al., 2026).

Collectively, these properties establish TRIBE v2 as a validated, robust, and scientifically credible tool for commercial in-silico neuroimaging applications.

References

1. d'Ascoli, S., Rapin, J., Benchetrit, Y., Brookes, T., Begany, K., Raugel, J., Banville, H., & King, J.-R. (2026). A foundation model of vision, audition, and language for in-silico neuroscience. Meta FAIR. https://github.com/facebookresearch/tribev2
2. d'Ascoli, S., Rapin, J., Benchetrit, Y., Banville, H., & King, J.-R. (2026). TRIBE: Trimodal brain encoder for whole-brain fMRI response prediction. In The Fourteenth International Conference on Learning Representations.
3. Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage, 31(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021
4. Frazier, J. A., Chiu, S., Breeze, J. L., Makris, N., Lange, N., Kennedy, D. N., ... & Biederman, J. (2005). Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder. American Journal of Psychiatry, 162(7), 1256–1265. https://doi.org/10.1176/appi.ajp.162.7.1256
5. Gifford, A. T., Bersch, D., St-Laurent, M., Pinsard, B., Boyle, J., Bellec, L., ... & Cichy, R. M. (2024). The Algonauts Project 2025 challenge: How the human brain makes sense of multimodal movies. arXiv preprint arXiv:2501.00504.
6. Hamilton, L. S., & Huth, A. G. (2020). The revolution will not be controlled: natural stimuli in speech neuroscience. Language, Cognition and Neuroscience, 35(5), 573–582. https://doi.org/10.1080/23273798.2018.1499946
7. Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. Nature, 532(7600), 453–458. https://doi.org/10.1038/nature17637
8. Jain, S., Vo, V. A., Wehbe, L., & Huth, A. G. (2024). Computational language modeling and the promise of in silico experimentation. Neurobiology of Language, 5(1), 80–106. https://doi.org/10.1162/nol_a_00101
9. Makris, N., Goldstein, J. M., Kennedy, D., Hodge, S. M., Caviness, V. S., Faraone, S. V., ... & Seidman, L. J. (2006). Decreased volume of left and total anterior insular lobule in schizophrenia. Schizophrenia Research, 83(2–3), 155–171. https://doi.org/10.1016/j.schres.2005.11.020
10. Yamins, D. L. K., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. Nature Neuroscience, 19(3), 356–365. https://doi.org/10.1038/nn.4244