# Linguistic properties and model scale in brain encoding from small to compressed language models

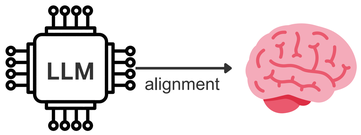
🏆 Spotlight Paper

Subba Reddy Oota[1], Vijay Rowtula[2*], Satya Sai Namburi[3*], Khushbu Pahwa[4], Anant Khandelwal[5], Manish Gupta[5], Tanmoy Chakraborthy[6], Bapi S. Raju[2]

[1]Technische Universität Berlin, Berlin, Germany, [2]IIIT Hyderabad, India, [3]GE Healthcare, USA, [4]Amazon AWS, USA, [5]Microsoft, [6]IIT-Delhi, India

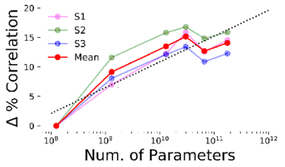
subbareddyoota@gmail.com

## Introduction




LMs predict brain activity, and that alignment improves as models scale, suggesting neural scaling laws [1, 2, 3]

Parameter Count [1]




⚠ Larger models, better brain alignment, but harder to interpret and costly to run

**"What is the minimal model capacity required to achieve brain alignment comparable to larger LLMs?"**

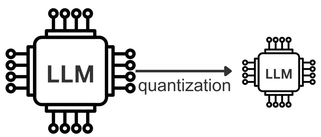
## Datasets

### Naturalistic fMRI dataset [4]

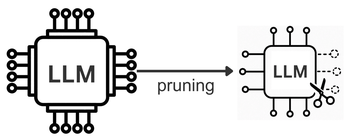
| Dataset | Modality | #Subj | Language | #TRs |
|---|---|---|---|---|
| Subset-Moth-Radio-Hour | Listening | 9 | English (en) | 3737 |
| Subset-Moth-Radio-Hour | Reading | 9 | English (en) | 3737 |

### Transformer-based language models

| Model name | Model scale | Type |
|---|---|---|
| DeepSeek-R1 | 1.5B, 3B, 7B, 14B | Encoder |
| LLaMA-3.2 | 1B, 3B, 8B, 14B | Encoder |
| Qwen-2.5 | 1.5B, 3B, 7B, 14B | Encoder |

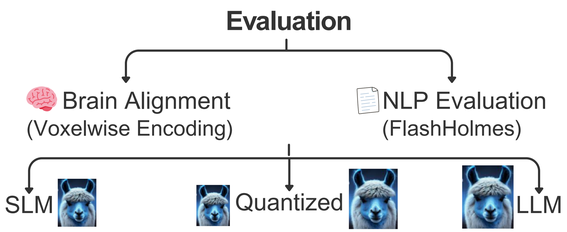



AWQ, GPTQ, SmoothQunat

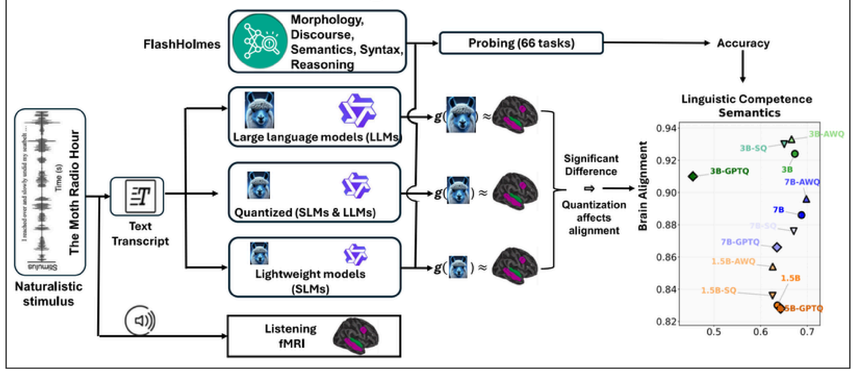



Unstructured: 10%, 20%, 50%

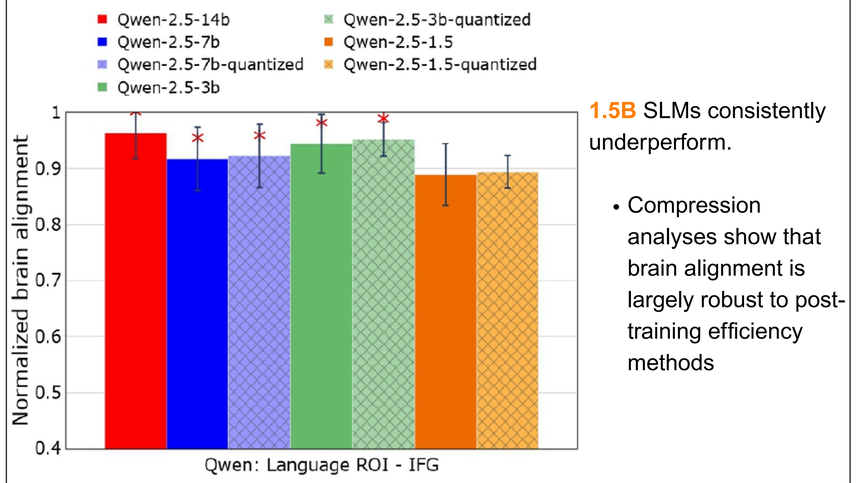
### Evaluation




## Does linguistic competence drive brain–model alignment?




## 3B SLMs match brain alignment with 7B-14B LLMs



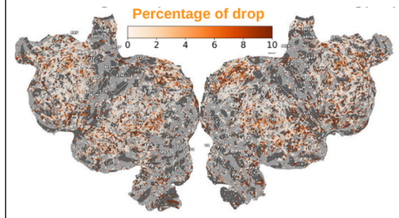

1.5B SLMs consistently underperform.

- Compression analyses show that brain alignment is largely robust to post-training efficiency methods

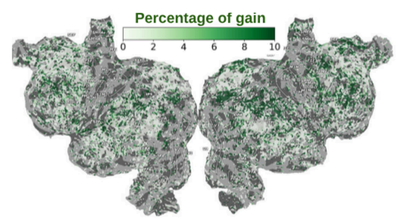
## Conclusions

- Brain alignment saturates early (~3B): 3B SLMs match 7B–14B LLMs, while 1B–1.5B models reduce alignment, especially in semantic regions
- Alignment is robust to compression--except GPTQ: AWQ and SmoothQuant preserve alignment; GPTQ consistently degrades it.
- Linguistic competence and brain alignment dissociate

## Brain alignment is robust to compression, except GPTQ

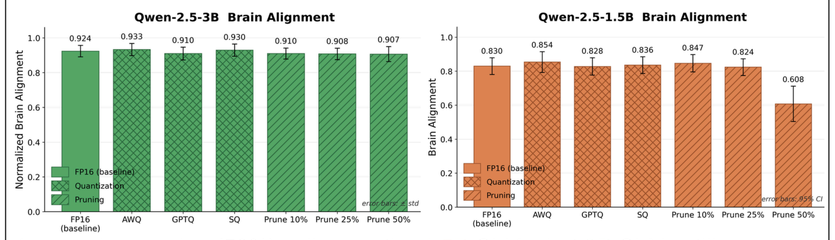



Percentage drop (3b → 3b-gptq)



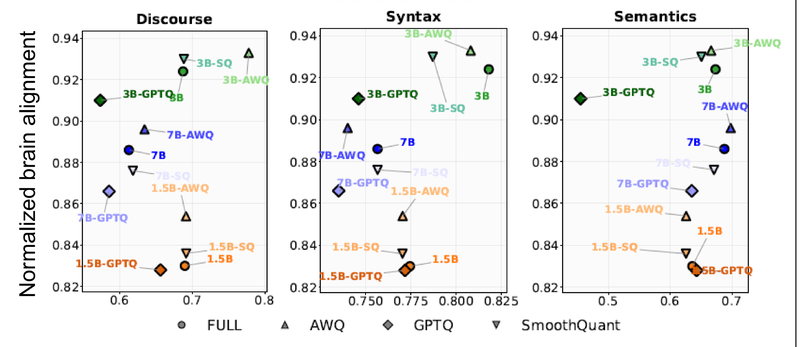

Percentage gain (3b → 3b-awq)

### Moderate pruning preserves alignment at both scales (1.5B-3B)




Aggressive pruning (50%) collapses the 1.5B SLMs, but 3B SLMs still preserves

## Linguistic competence and brain alignment dissociate




- 3B and 7B models cluster at the top of brain alignment, competence and alignment come apart
- Tiny SLMs (1.5B) can keep the linguistic properties yet lose brain-relevant representations

**References**

[1] Antonello, Richard and Vaidya, Aditya and Huth, Alexander.Scaling laws for language encoding models in fMRI. NeurIPS 2023.
[2] Matsuyama, Takuya and Sasaki, Kota S and Nishimoto, Shinji. Applicability of scaling laws to vision encoding models. arXiv 2023.
[3] AlKhamissi, Badr, et al. From language to cognition: How llms outgrow the human language network. EMNLP 2025.
[4] Deniz, Fatma, et al. The representation of semantic information across human cerebral cortex during listening versus reading is invariant to stimulus modality. Journal of Neuroscience 2019.

**Acknowledgment**

This work was partially supported by a travel grant from FICO.

ICML International Conference On Machine Learning