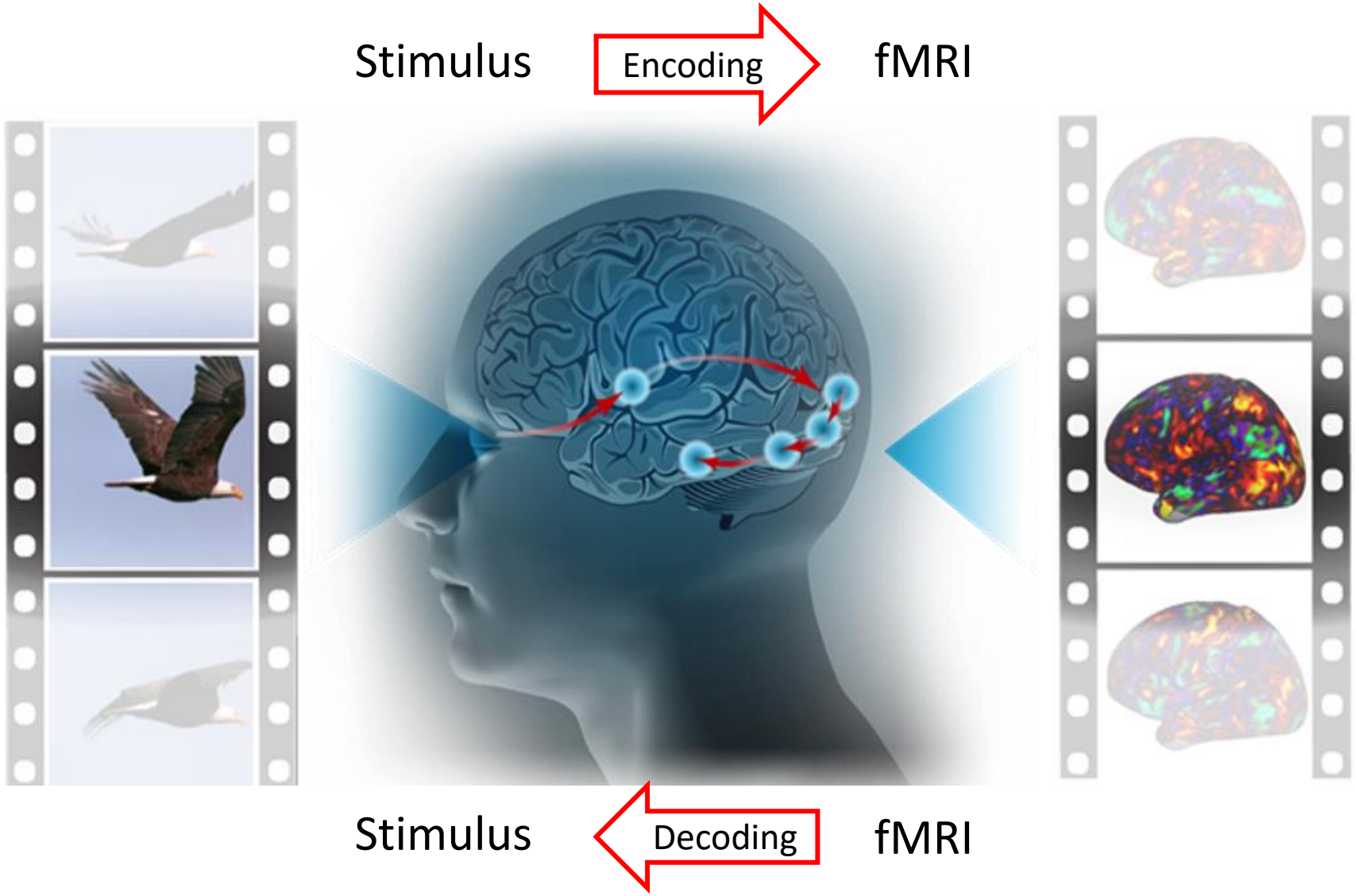
Predicting Brain Activity while Listening to Stories

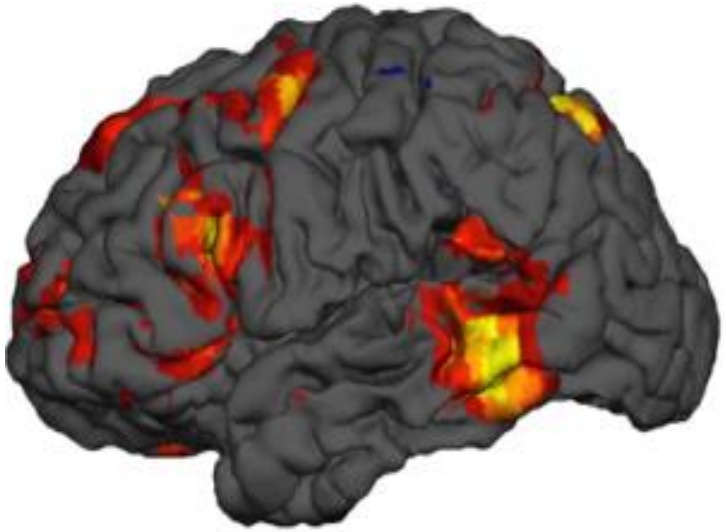
Subba reddy Oota, Frederic Alexandre, Xavier Hinaut

Brain Encoding vs Decoding



Haiguang Wen et al, 2017

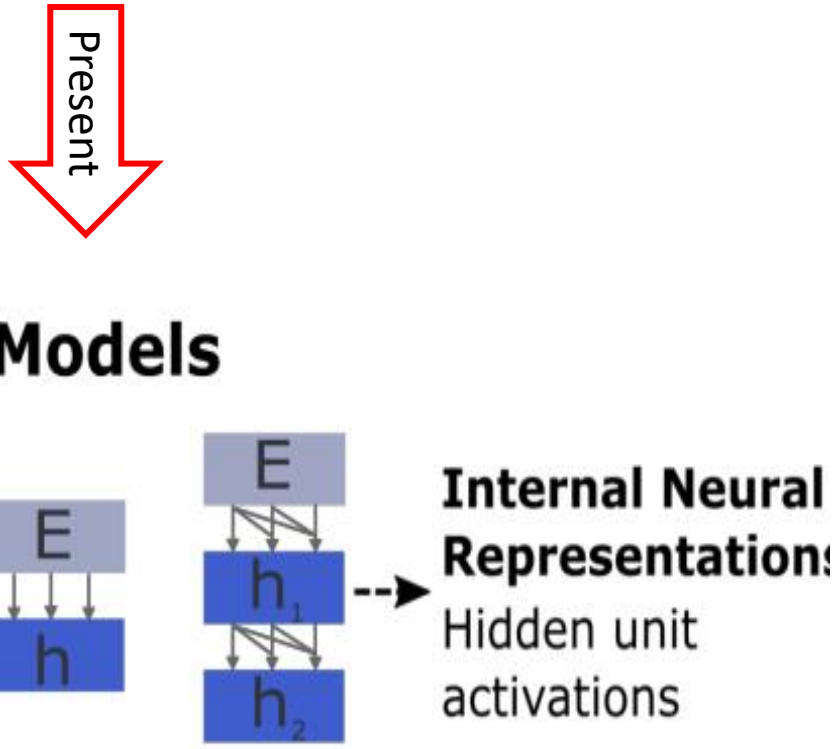
What is fMRI?



https://www.biopac.com/events/fmri-psych/

Brain Encoding

Stimulus



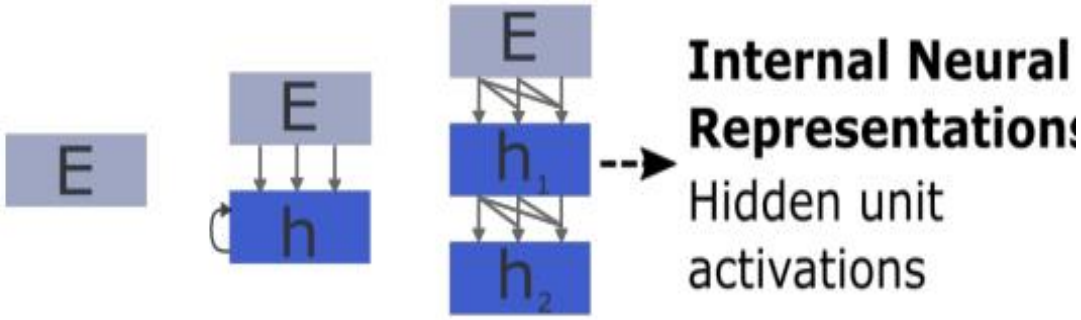
Schrimpf et al. 2021 fMRI

Brain Encoding

Stimulus

Present

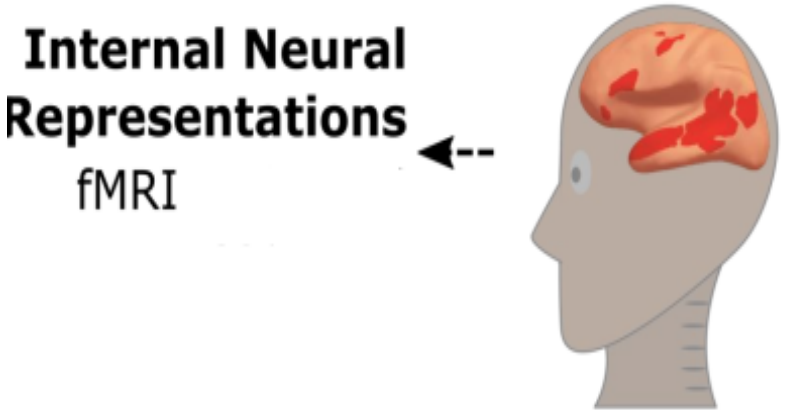
Models



Stimulus

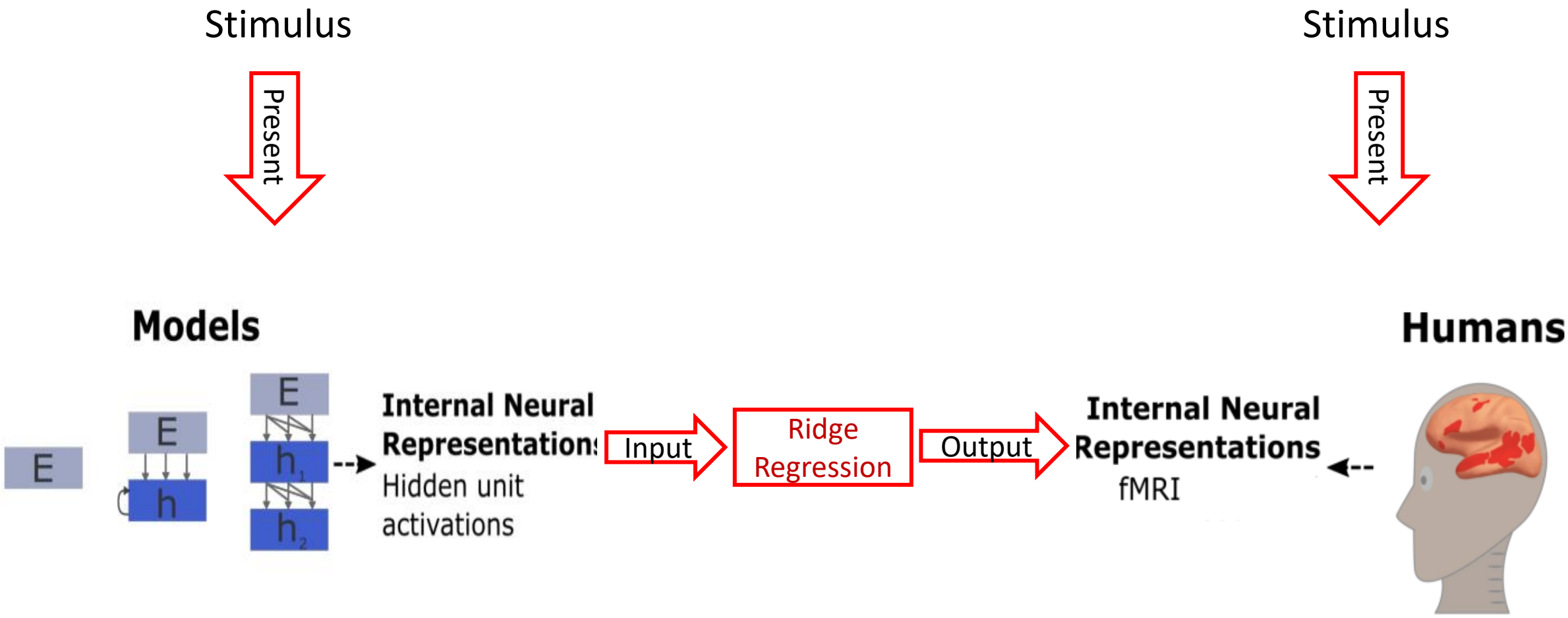
Present

Humans



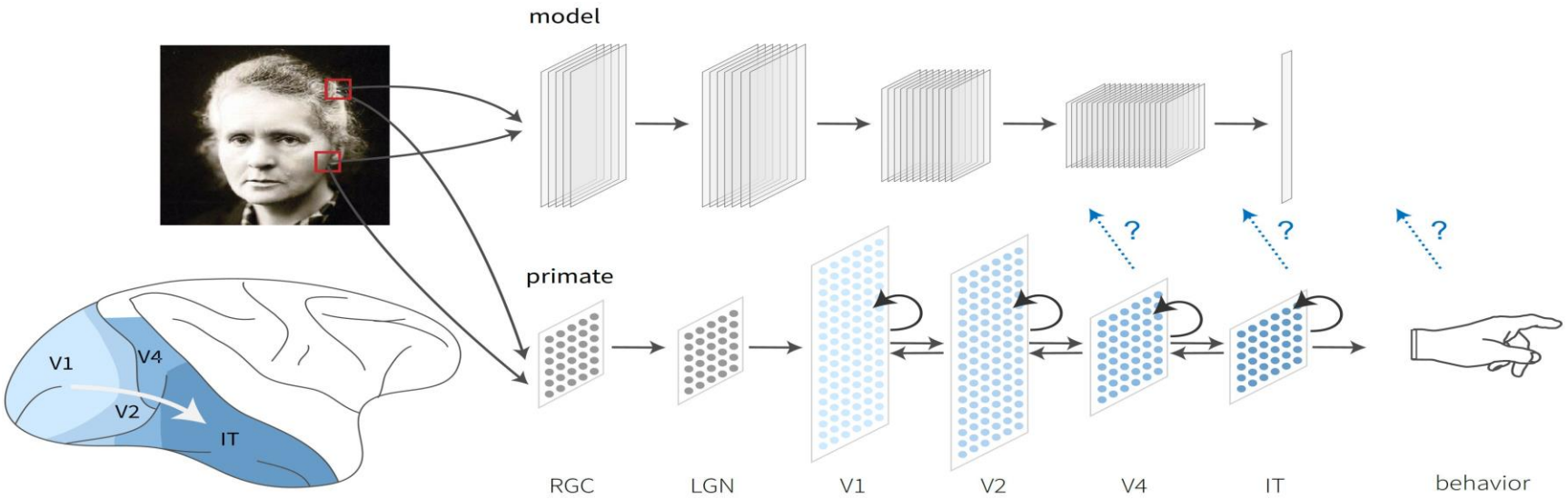
Schrimpf et al. 2021 fMRI

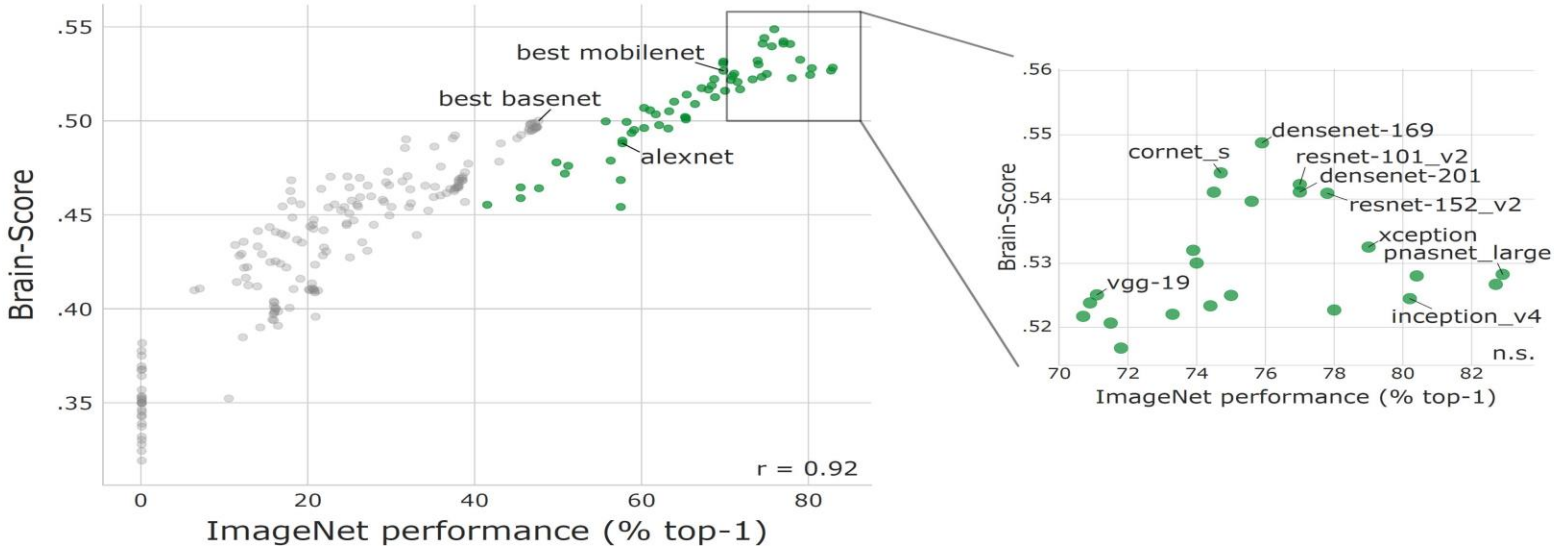
Brain Encoding



Schrimpf et al. 2021 fMRI

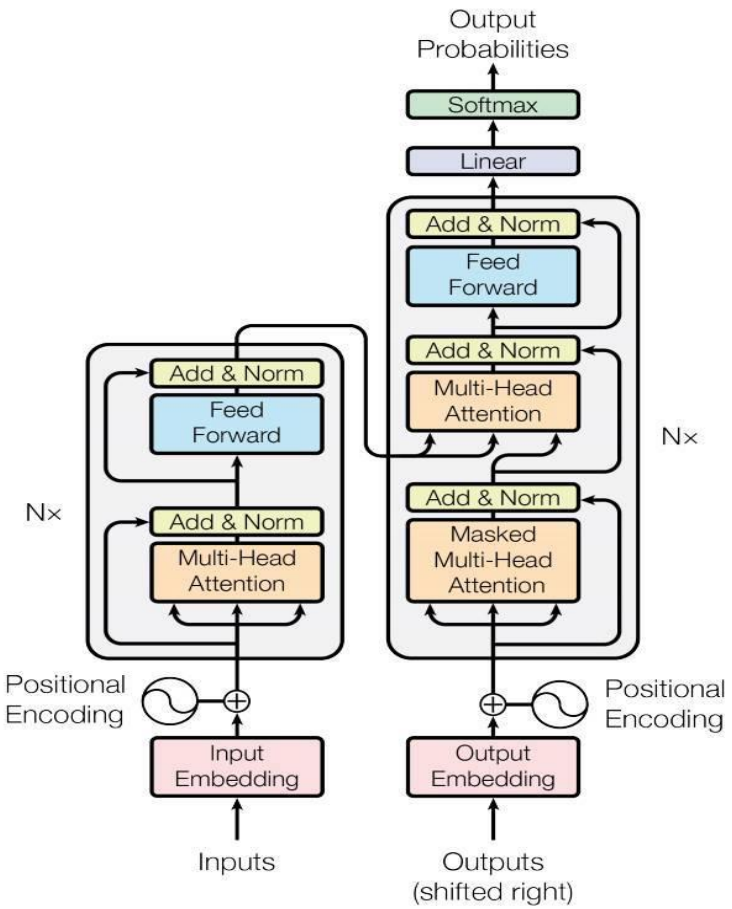
Vision models accurately predict human vision recordings





Schrimpf*, Kubilius*, et al. 2018 | Kubilius*, Schrimpf*, et al. 2019

Most popular language models are Tranformers



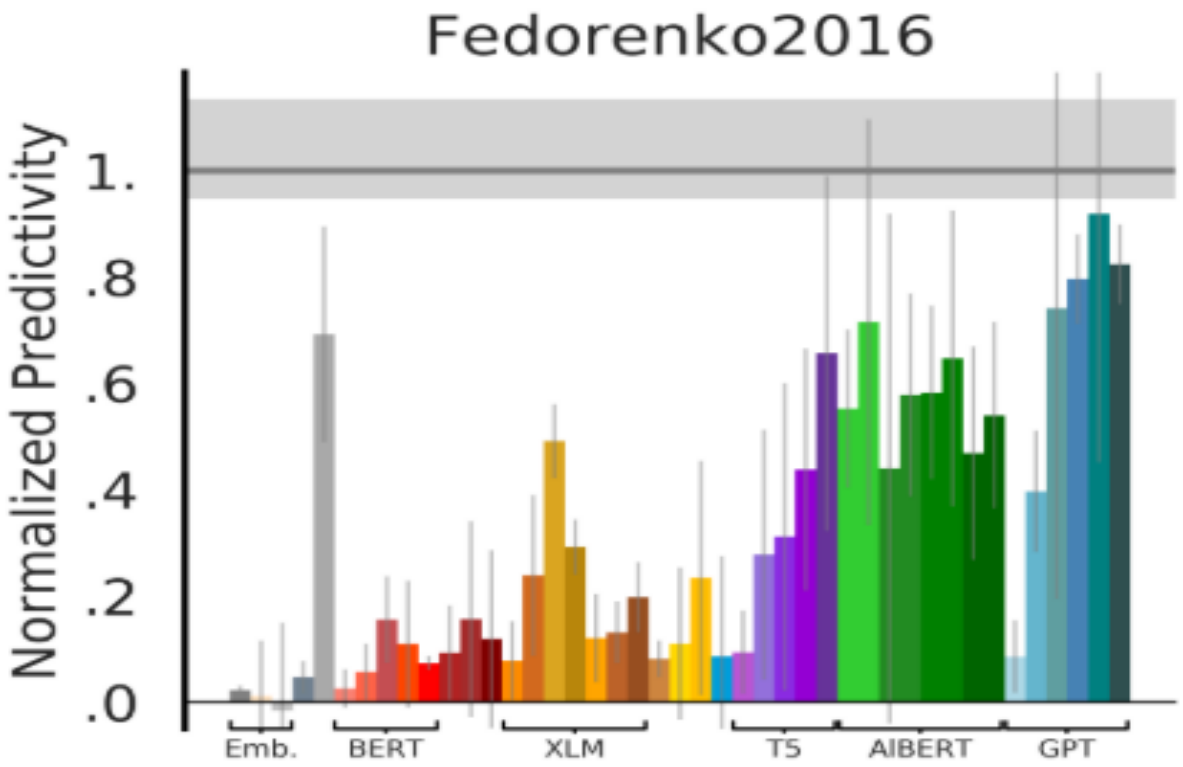
Transformer language models
(BERT, XML, GPT,…)

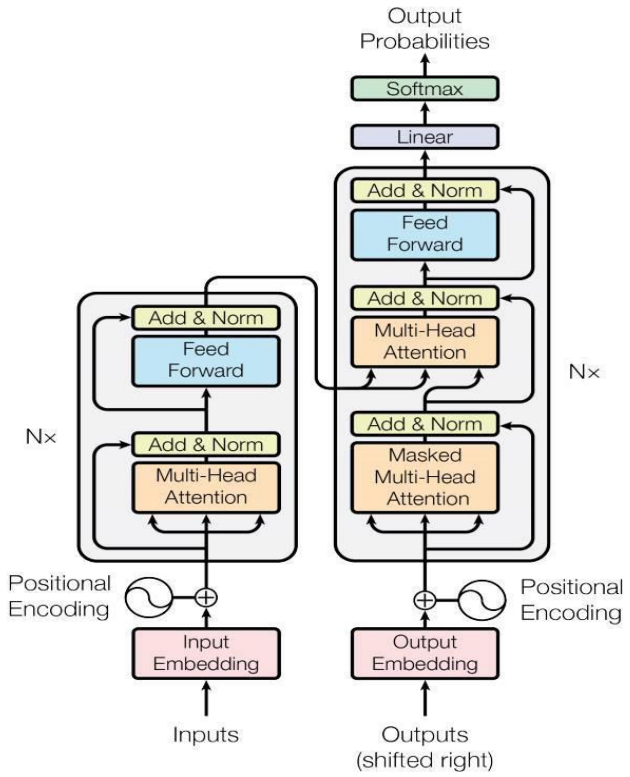
Vaswani et al. 2017

Language models based on words

Words

Stimulus → "Alex was tired so he took a nap."





Transformer language models (BERT, XML, GPT,…)

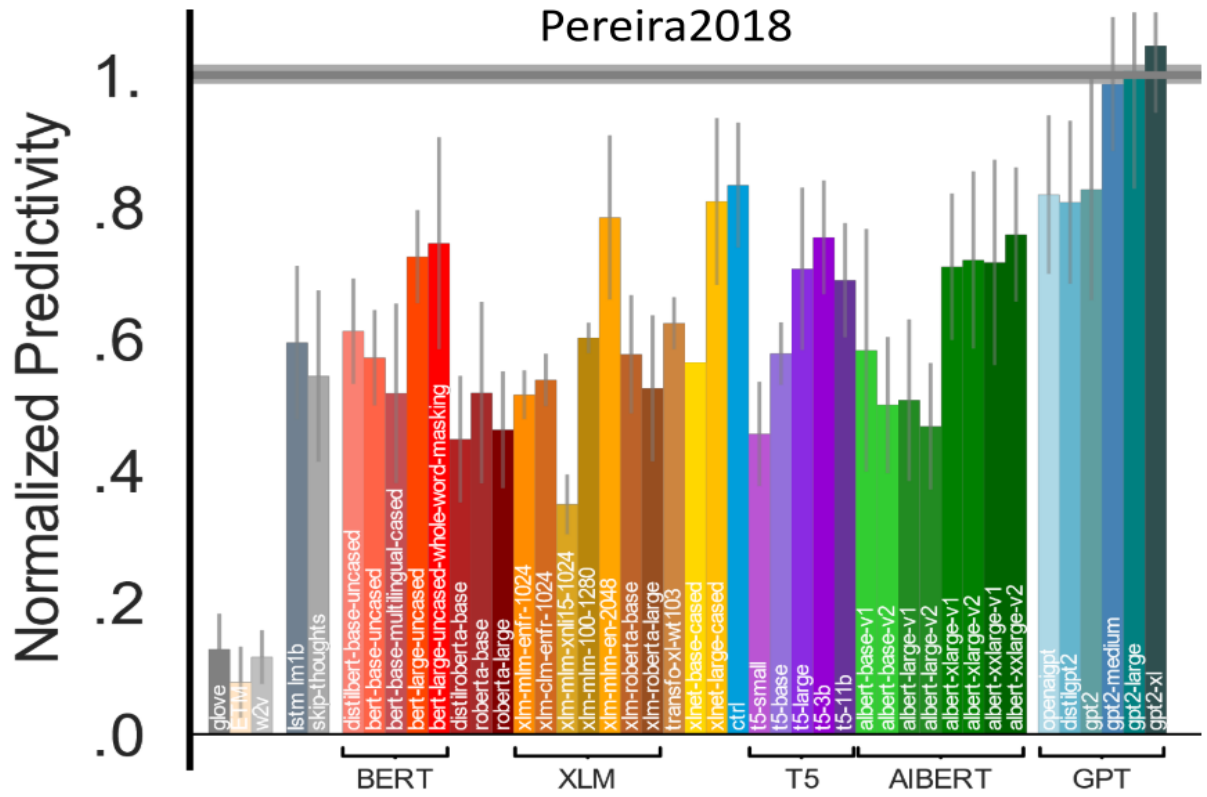
Vaswani et al. 2017, Schrimpf et al. 2021

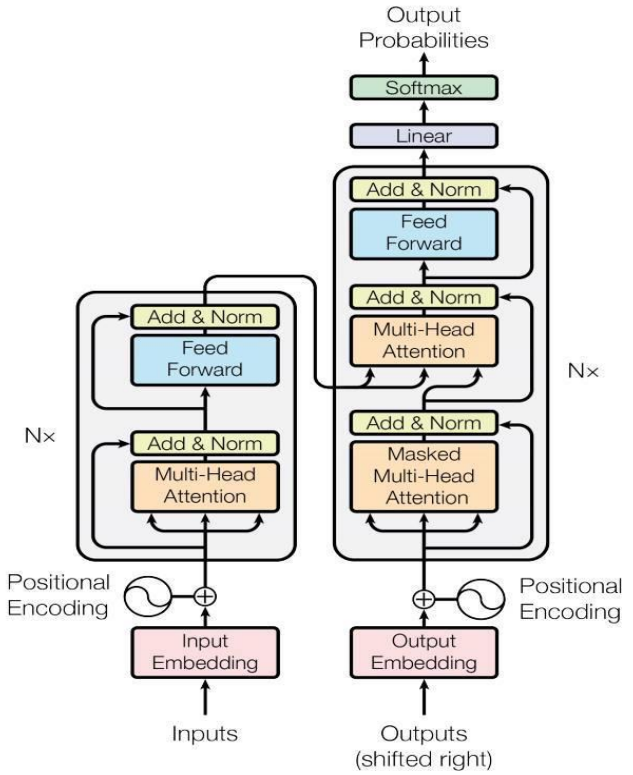
Language models based on sentences

Sentences

Stimulus

"Beekeeping encourages the conservation of local habitats. It is in every beekeeper's interest..."

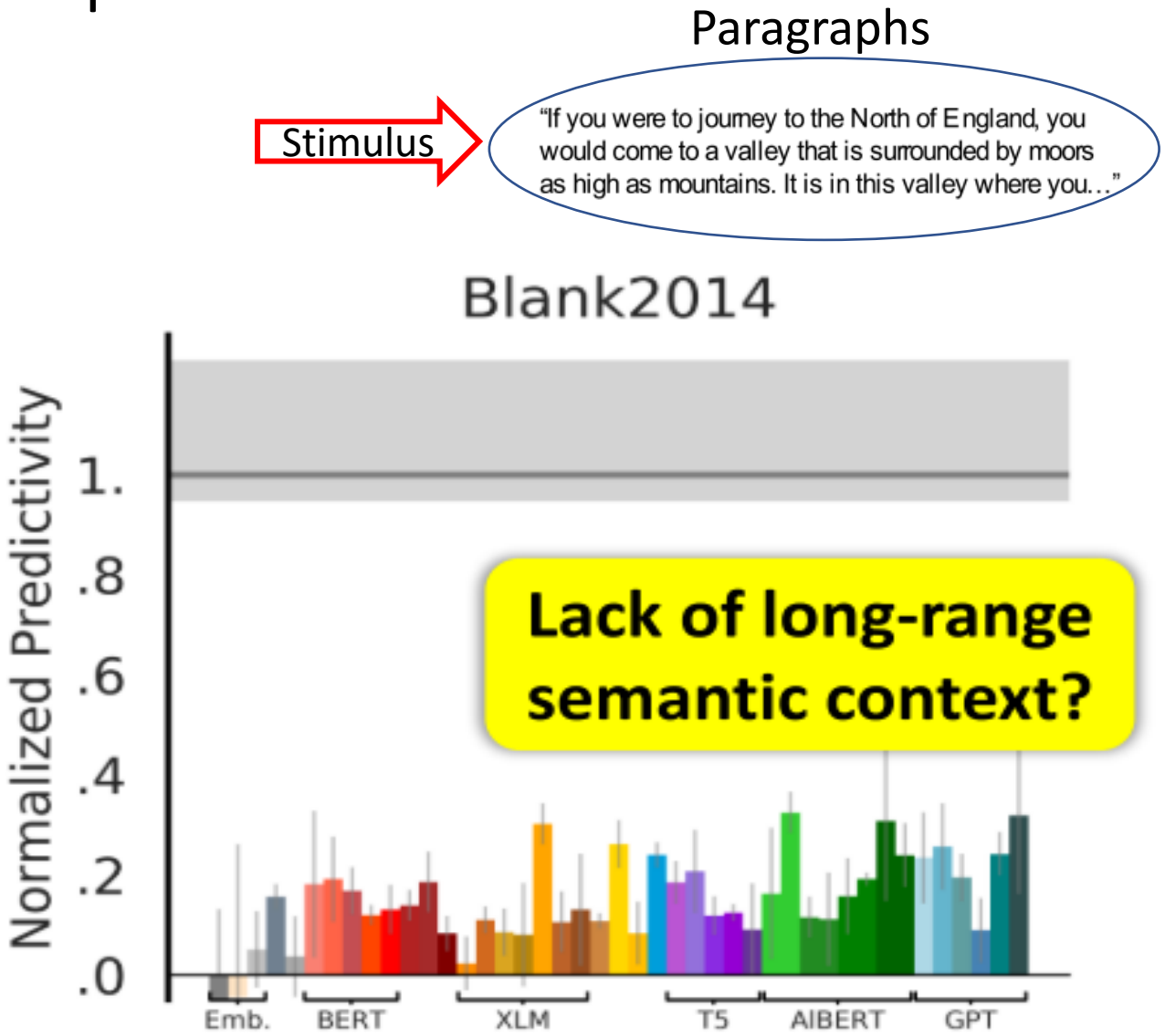




Transformer language models (BERT, XML, GPT,…)

Vaswani et al. 2017, Schrimpf et al. 2021

Can current language models deal with long-term dependencies?



- Transformer language models are unable to handle the long-term dependencies
  - **sequence** length is fixed to 512 words
- LSTMs still lacks investigation of the long-term memory cognitive plausibility and its link to fMRI data

Vaswani et al. 2017, Schrimpf et al. 2021

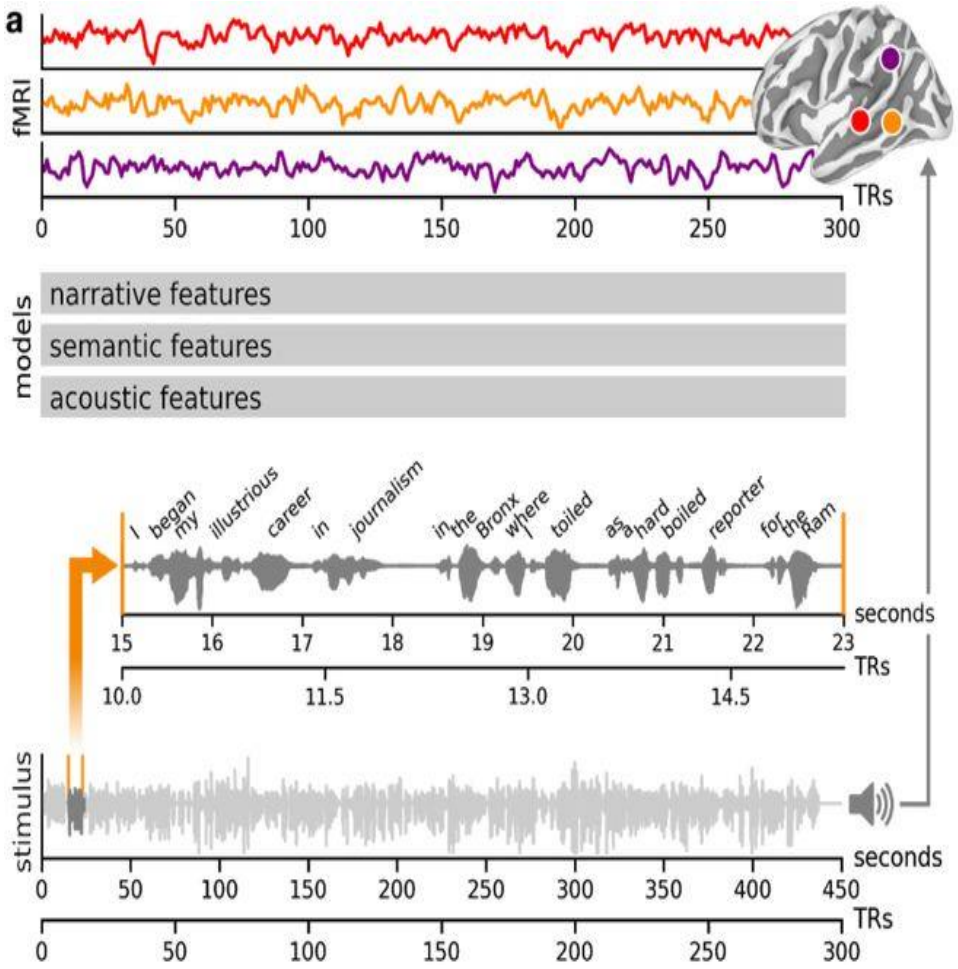
What kind of language models can represent long-term dependencies?

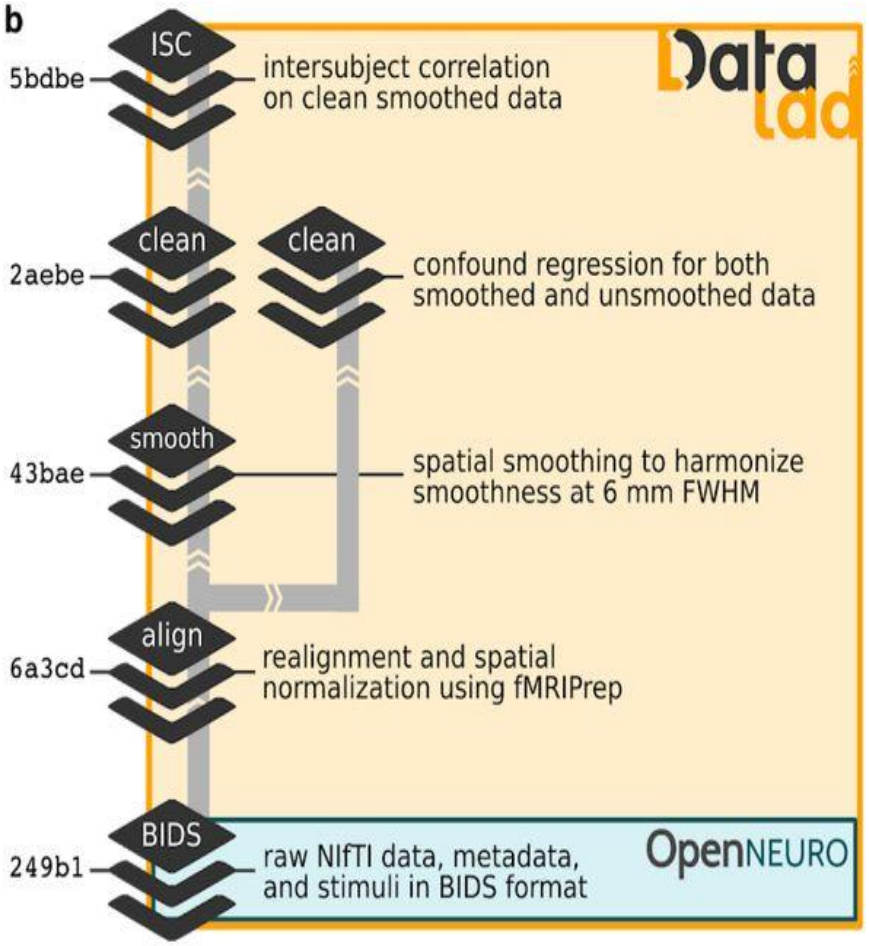
could they also predict higher cognition while subjects are engaged in longer stories ?

The data target: human brain recordings

- We use Pieman story listening:
  - 82 subjects,
  - 259 TRs (repetition time)
  - here it is 1.5 sec.

Example: "I began my illustrious carrier in journalism…"

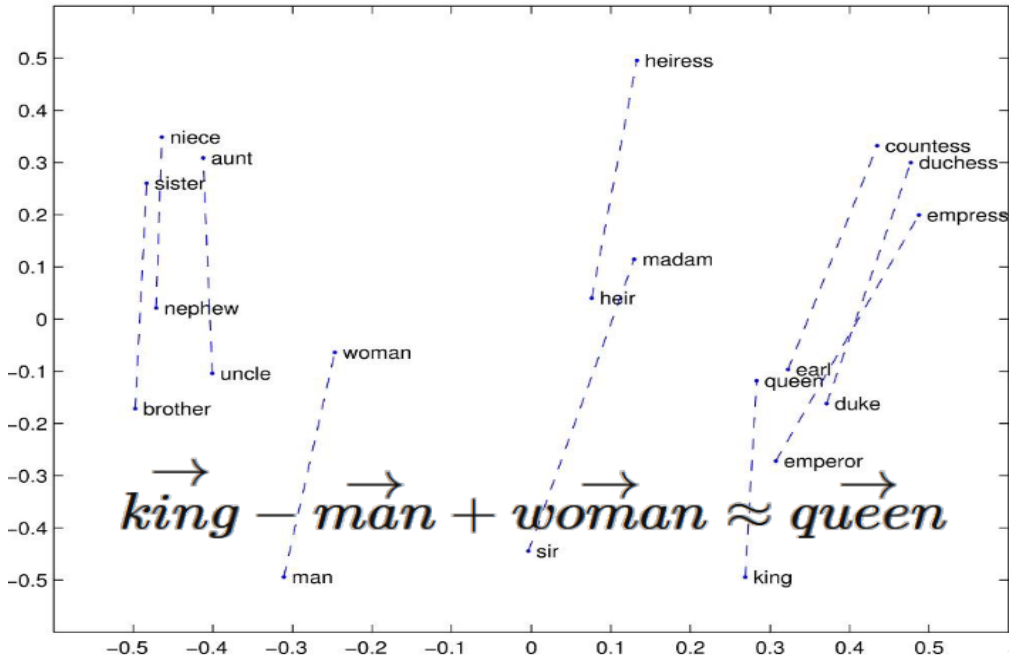




Nastase et al. 2021 fMRI

Models considered (n = 5)

- Embedding type models: GloVe

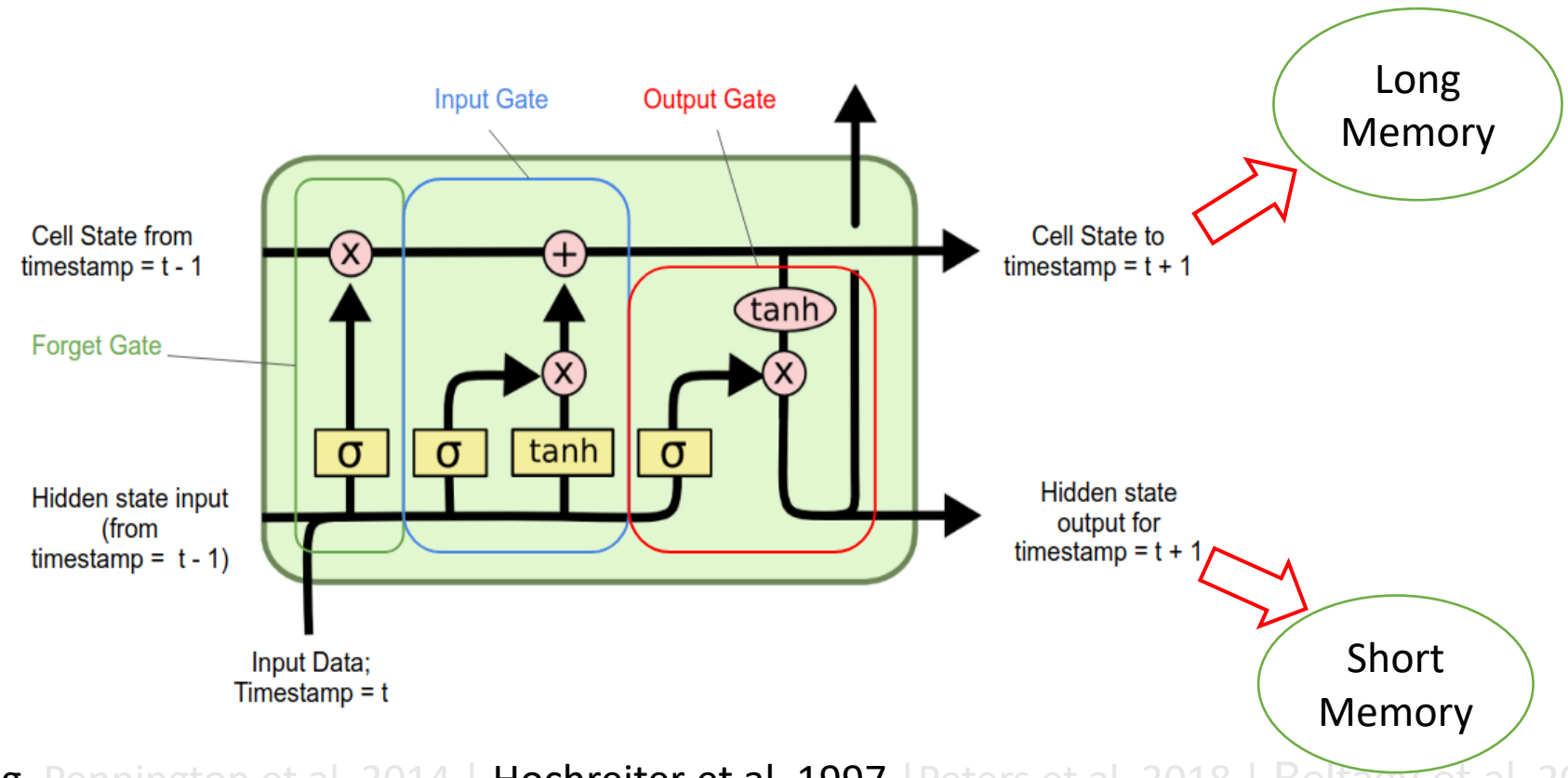


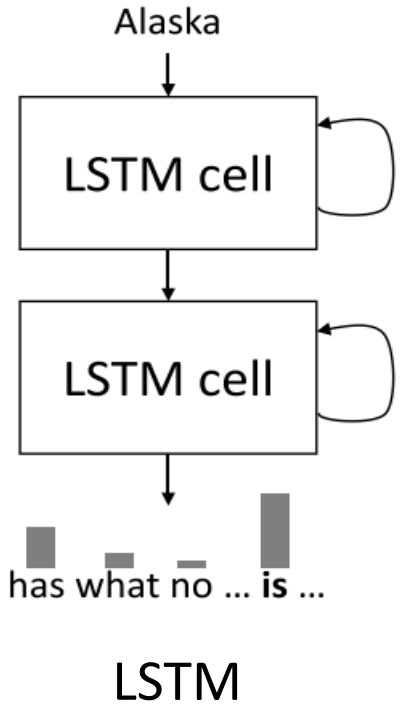
GloVe

e.g. Pennington et al. 2014 | Hochreiter et al. 1997 | Peters et al. 2018 | Beltagy et al. 2020

Models considered (n = 5) Contd.

- Embedding type models: GloVe
- Recurrent type models: LSTM, RandLSTM



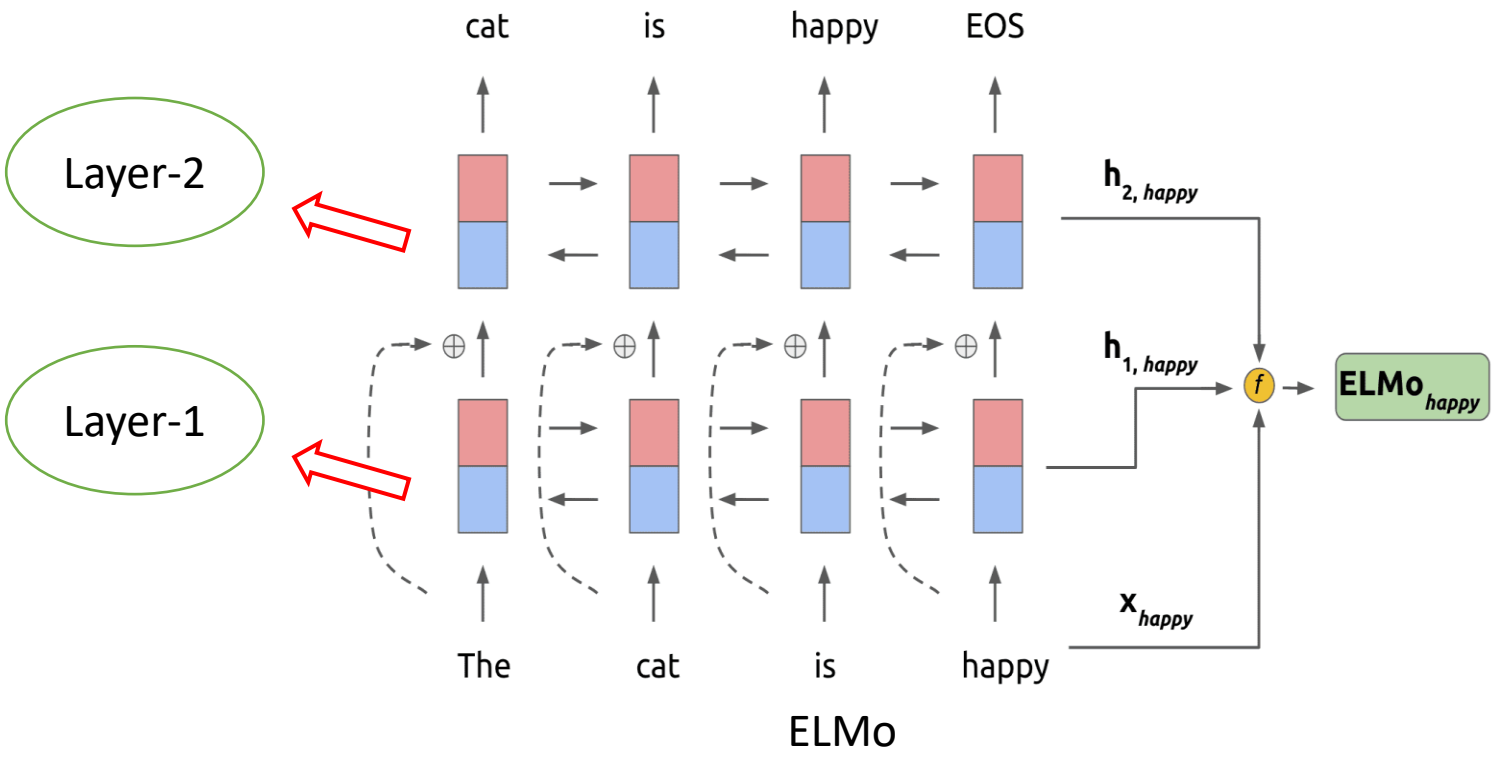


LSTM

e.g. Pennington et al. 2014 | Hochreiter et al. 1997 | Peters et al. 2018 | Beltagy et al. 2020

Models considered (n = 5) Contd.

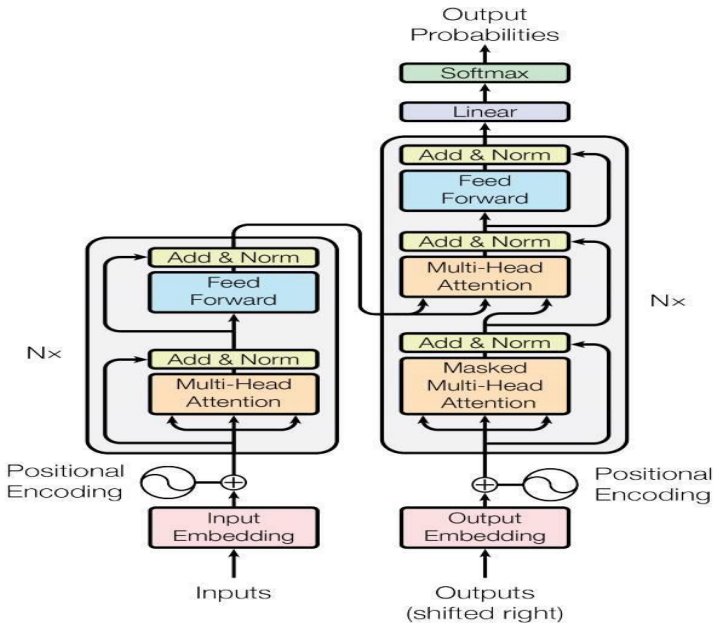
- Embedding type models: GloVe
- Recurrent type models: LSTM, RandLSTM
- Contextualized pretrained models: ELMo, Longformer



e.g. Pennington et al. 2014 | Hochreiter et al. 1997 | Peters et al. 2018 | Beltagy et al. 2020

Models considered (n = 5) Contd.

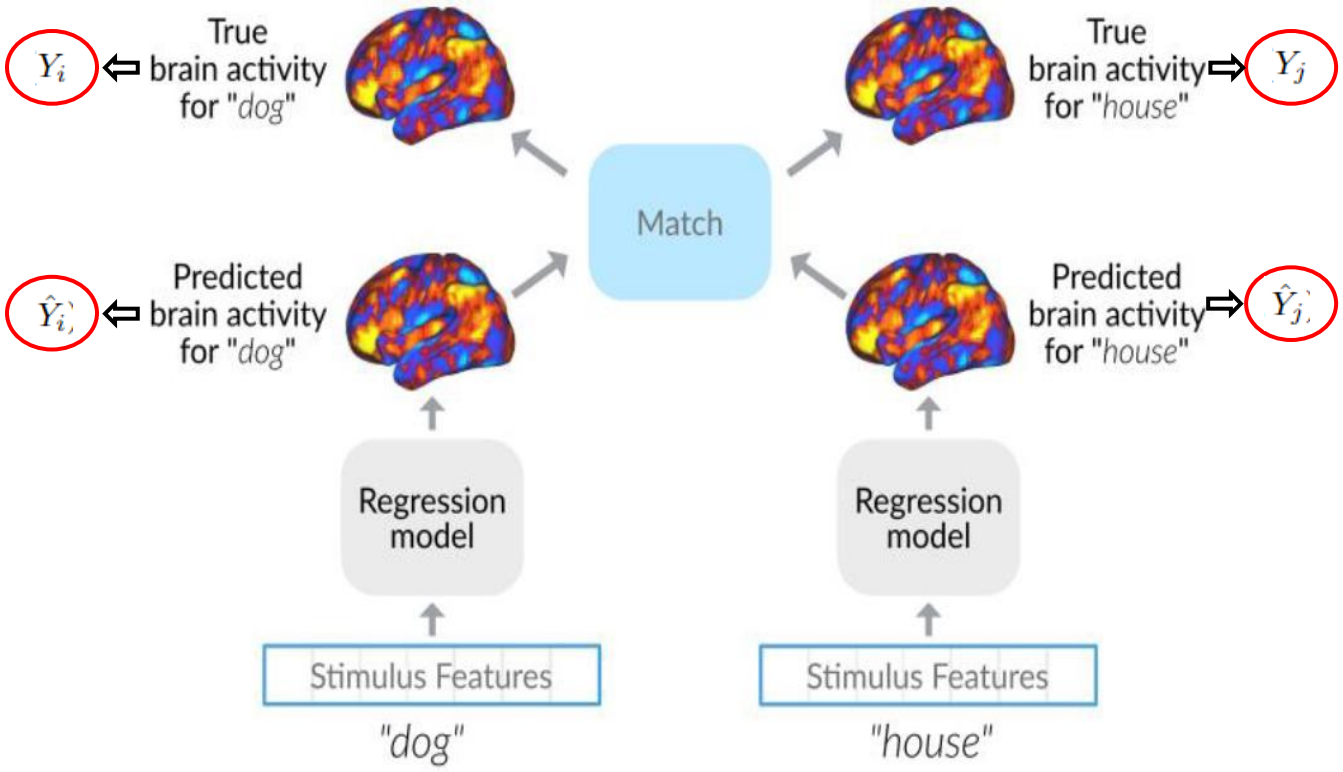
- Embedding type models: GloVe
- Recurrent type models: LSTM, RandLSTM
- Contextualized pretrained models: ELMo, Longformer



Longformer: Transformer with Global + local sliding window

e.g. Pennington et al. 2014 | Hochreiter et al. 1997 | Peters et al. 2018 | Beltagy et al. 2020

Evaluation Metrics: 2V2 and Pearson



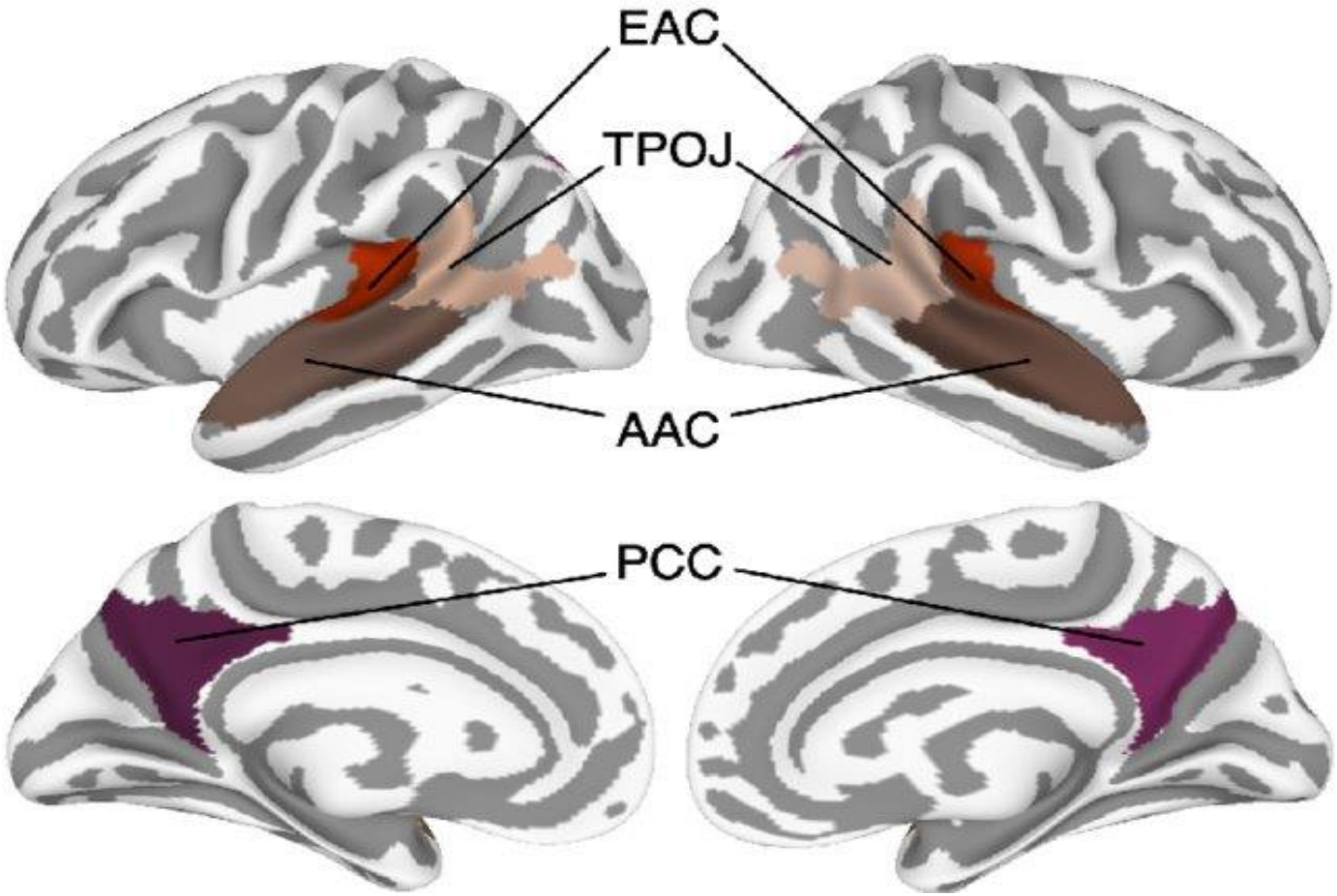
2V2 Accuracy

$$\text{2V2 Accuracy} = \frac{1}{N_{C_2}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} I[\{cosD(Y_i, \hat{Y}_i) + cosD(Y_j, \hat{Y}_j)\} < \{cosD(Y_i, \hat{Y}_j) + cosD(Y_j, \hat{Y}_i)\}]$$

Cosine distance

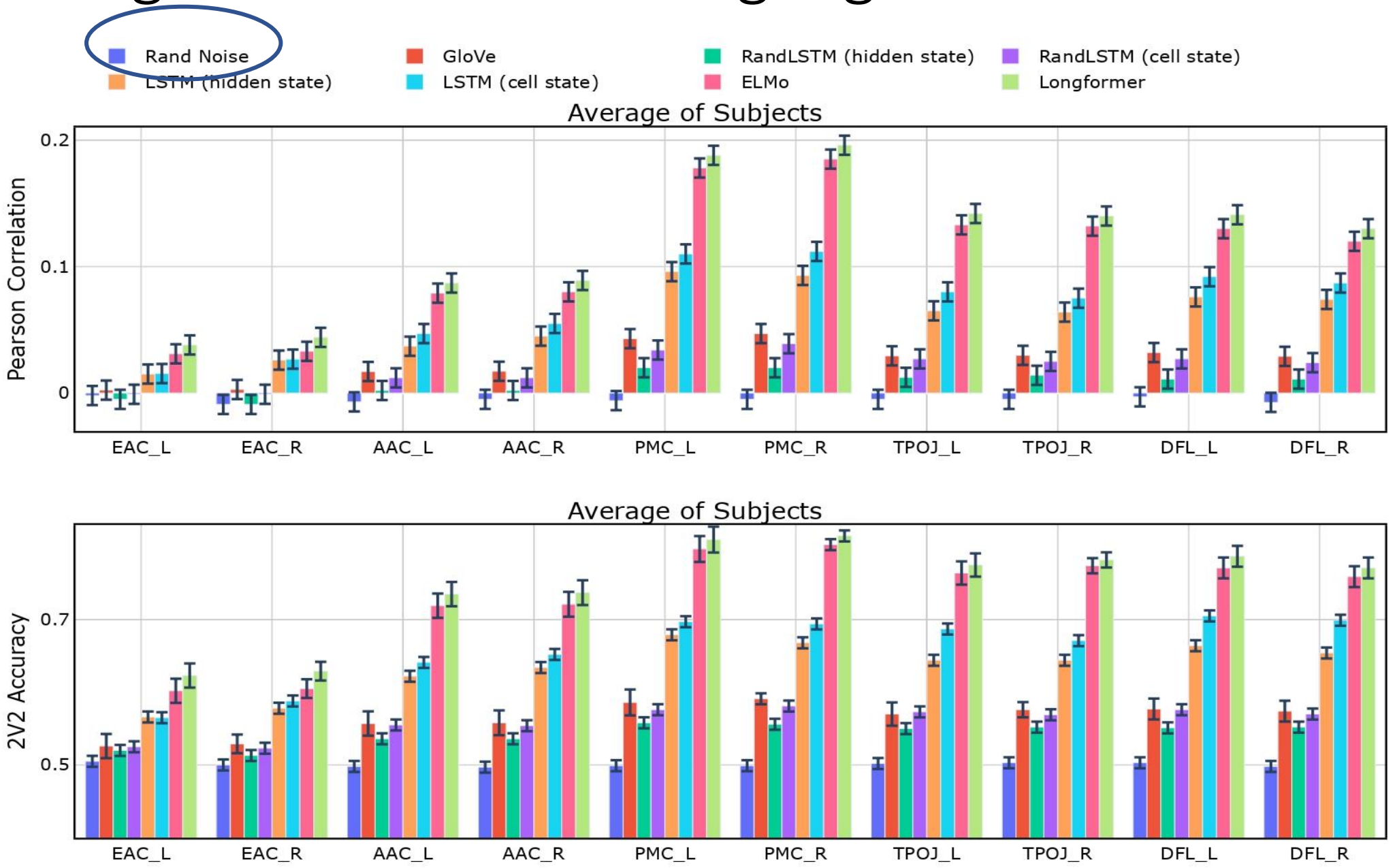
Language Hierarchy in Brain

- EAC – Early auditory cortex
- AAC – Auditory association cortex
- TPOJ – Temporo-parieto-occipital junction
- DFL – Dorsal frontal lobe
- PMC – Posterior medial cortex

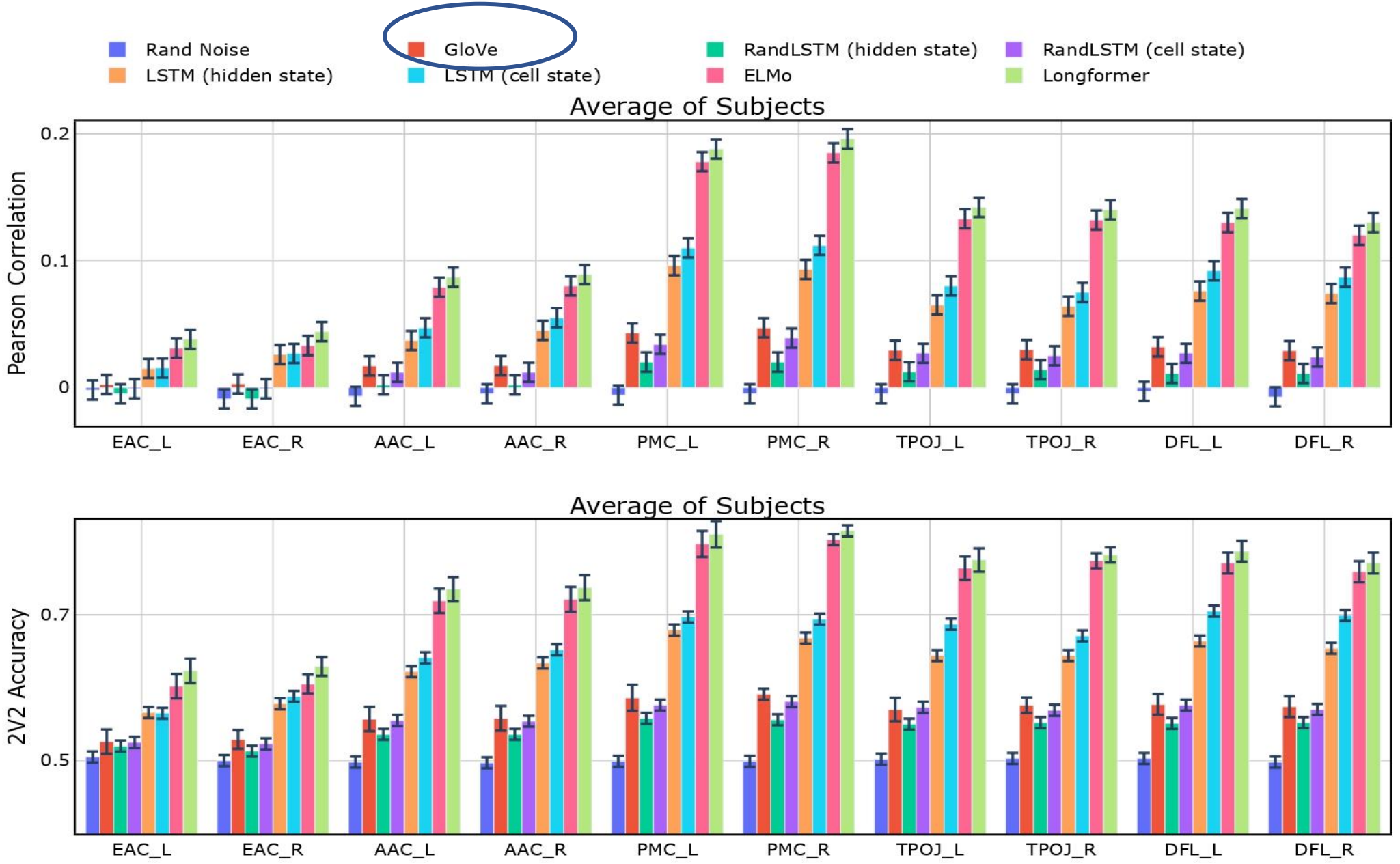


Nastase et al. 2019

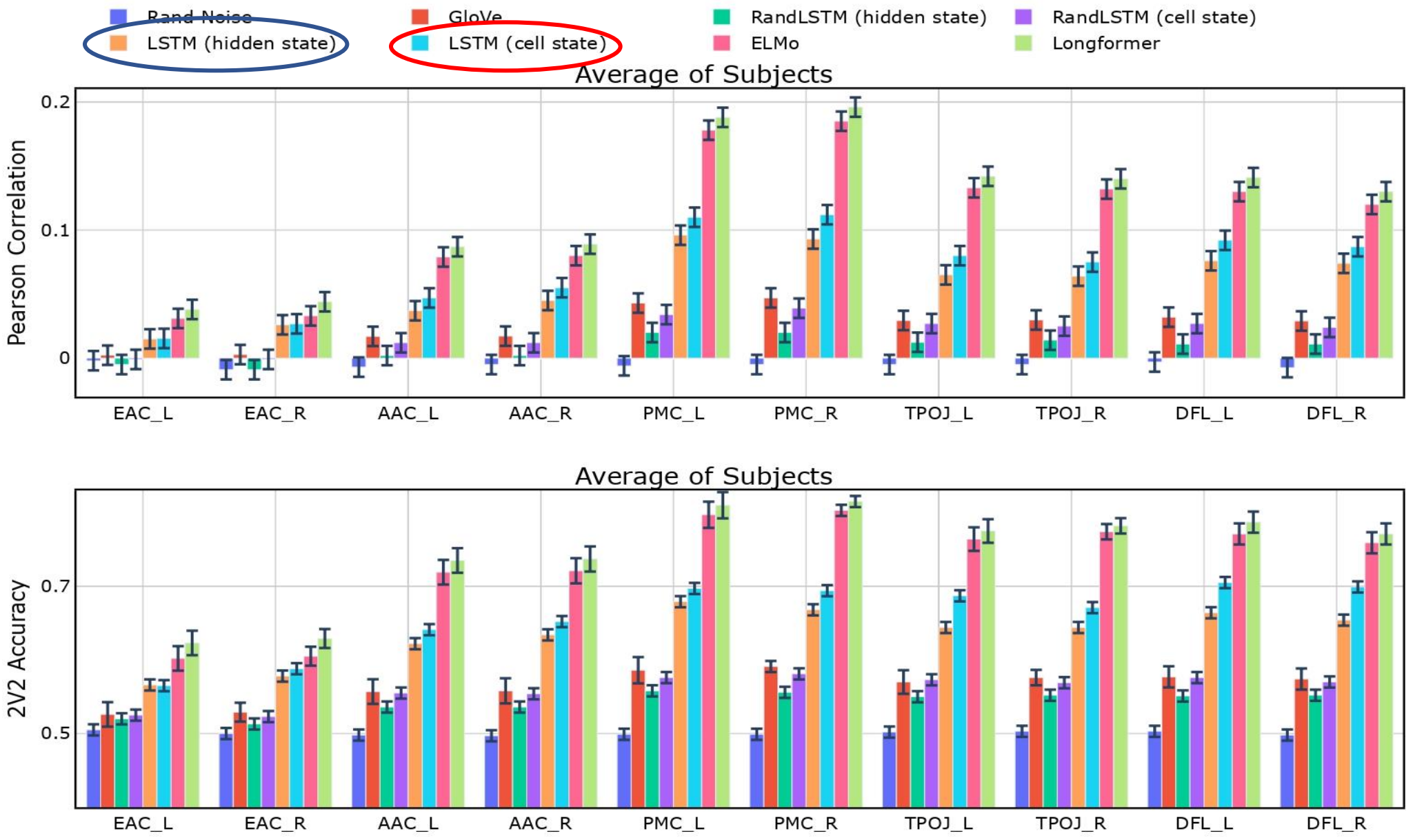
Encoding Performance of language models



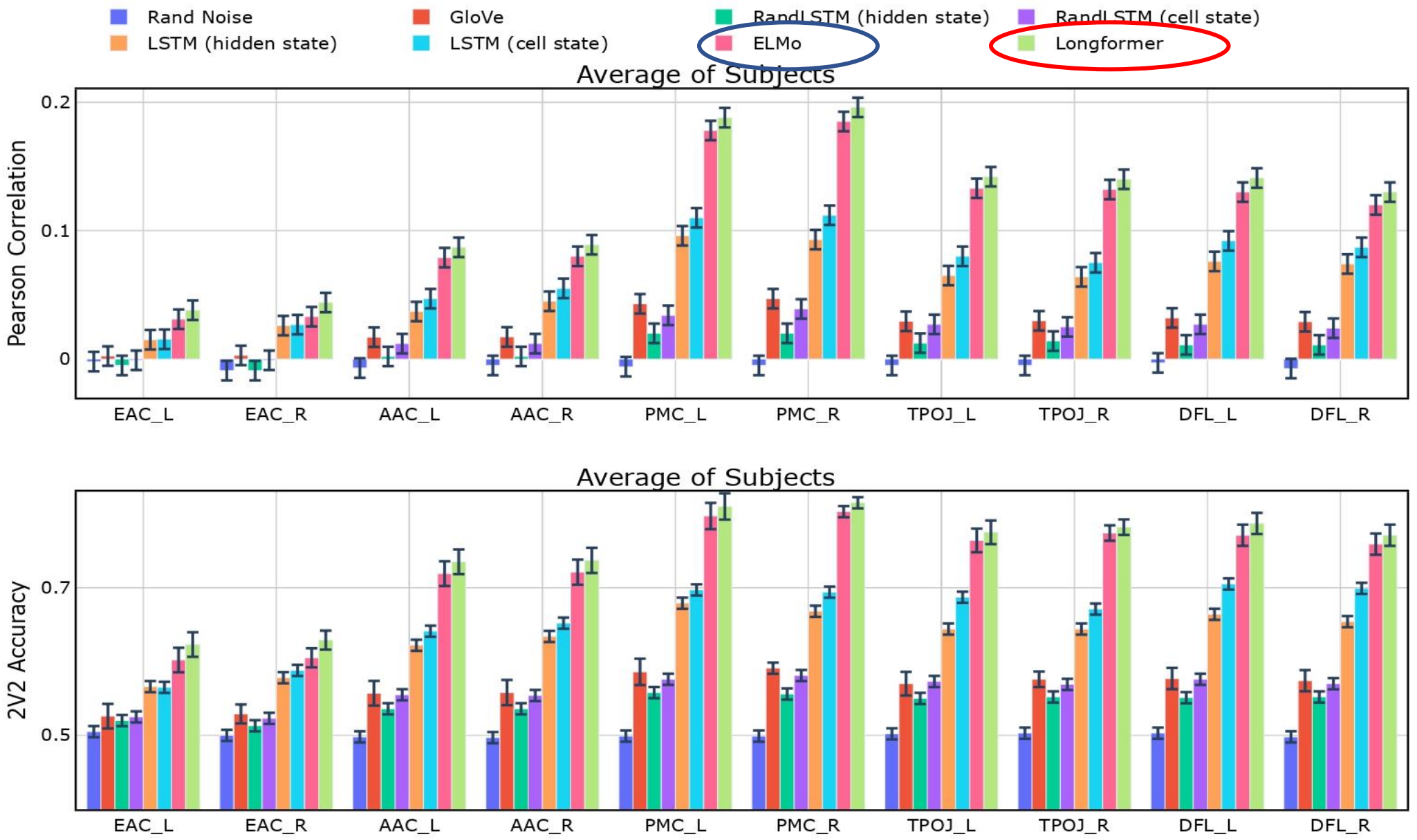
Encoding Performance of language models



Encoding Performance of language models



Encoding Performance of language models



Summary

- Language models with longer context better predict brain activity while subjects listening to stories.
- Future Directions:
  - Use more plausible language models than Transformers
  - Make hierarchical language models
    - Sub-part of the model predict different language levels: phonemes, words, POS, SRL

THANK YOU