NEURAL COMPUTATION

Prof. Alexander Huth

1.20.2021

TODAY

* Overview of class

* Syllabus

* A bit about what I do

BIG QUESTION

* Given a system that computes something,

  * what does it compute?

  * how does it work?

  * what are the ***internal representations*** that support that computation?

WHAT DOES IT MEAN TO UNDERSTAND?

* David Marr (1945 - 1980), *Vision* (1982)

*** Levels of analysis:**

  * Computational

  * Algorithmic/Representational

  * Implementation

COMPUTING SYSTEMS

* Biological: brains

* Artificial: neural networks

COMPUTING SYSTEMS

* Think about this for ~10 seconds, then drop a thought in the chat:

* What are some differences between **artificial** and **biological** neural networks?

BIOLOGICAL VS. ARTIFICIAL NEURAL NETWORKS

* size
* complexity x 2
* generality
* functionality of “neurons”
* analog vs. digital
* information encoding
* interpretability
* backpropagation?
* self-learning
* processing speed
* efficiency
* adaptability
* continual learning
* training
* division of labor
* catastrophic forgetting

TOPICS

* Biological neural networks (brains)
    * What are neurons? How do they compute?
    * How are neurons organized into networks?
    * How do biological neural networks learn?

TOPICS

* (Computational) Neuroscience

  * What methods are used to study biological neural systems?

  * How can/should we design neuroscience experiments?

  * How can we deal with noise and uncertainty in real data?

  * Encoding & decoding models

TOPICS

* System identification (encoding models)

* Given an unknown system $f(x) \rightarrow y$ and measurements of (x,y) pairs, can we create a model of f?

TOPICS

* Understanding artificial neural networks

* Attribution (gradient methods, input ablation)

* Adversarial methods

TOPICS

* Using artificial neural networks to understand biological neural networks

* With system identification

* By getting at underlying principles

COMPUTING SYSTEMS

* Think about this for ~10 seconds, then drop a thought in the chat:

* What are some ***similarities*** between **artificial** and **biological** neural networks?

BIOLOGICAL ❤ ARTIFICIAL NEURAL NETWORKS

* many small stupid things –> big smart thing

* both involve activation

* hard to interpret

* input gives output

* learning

* layered computation

* emergence

* data dependence

* specialized neurons

* break down complex inputs

SYLLABUS

* Online: https://github.com/alexhuth/neuralcomputation-sp2021

* ~3 problem sets

* Final project (alone or in pairs)

  * Proposal due **Mar. 17**

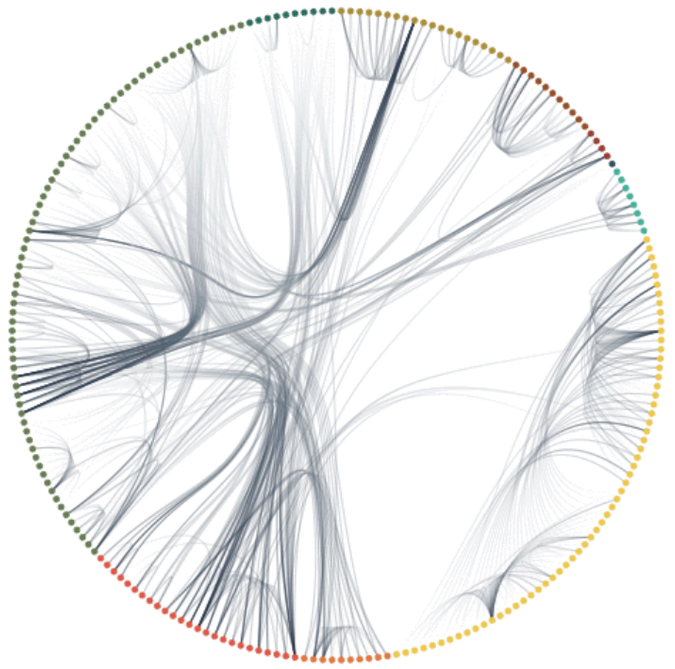
  * In-class presentations **Apr. 26/28 & May 3/5**

  * Write-up due **Apr. 26**

* Student paper presentations (5-10 minutes, 1 slide) every Wednesday

HOW DO WE UNDERSTAND LANGUAGE?

"Close your eyes and let the word paint a thousand pictures..."



NATURAL LANGUAGE EXPERIMENT

Language fMRI data

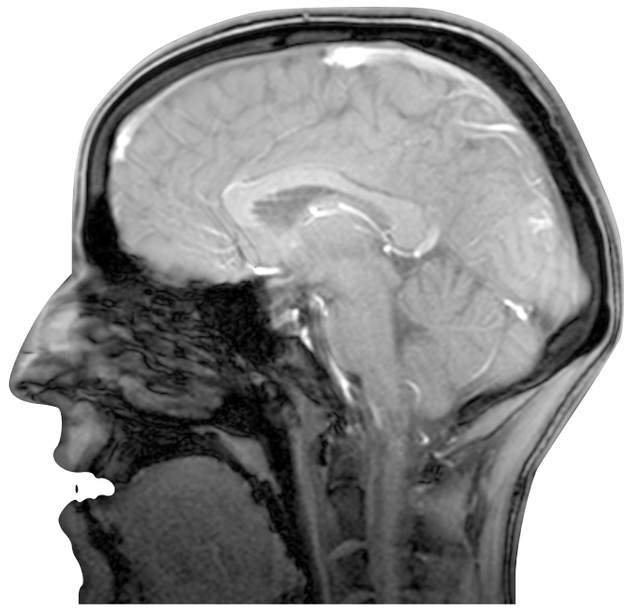
~5h narrative stories from
The Moth Radio Hour

"…she was removing photographs from the walls and placing them in little piles around the house…"

Huth, de Heer, Griffiths, Theunissen, & Gallant. *Nature* (2016) → 2 hours / subject

Jain, LeBel, & Huth (*in prep.*) → 20 hours / subject

VOXELWISE MODELING



NATURAL STORY STIMULI

FEAT. SPACE A

FEAT. SPACE B

FEAT. SPACE C

fMRI DATA

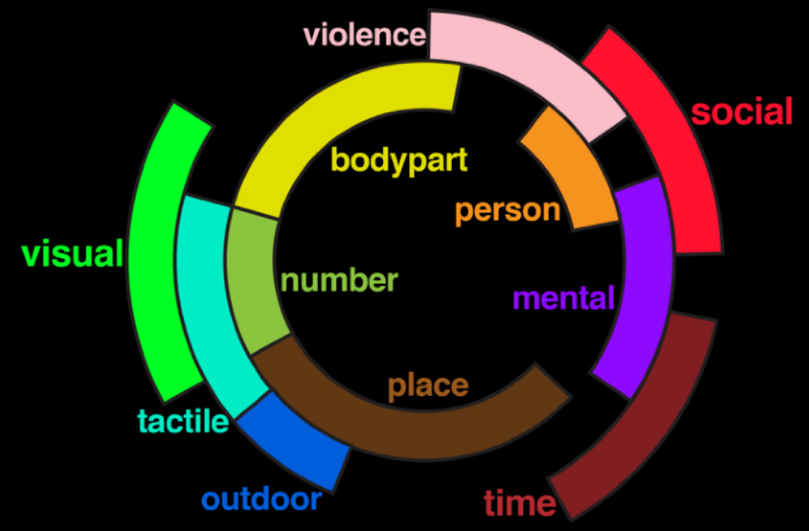
ESTIMATION

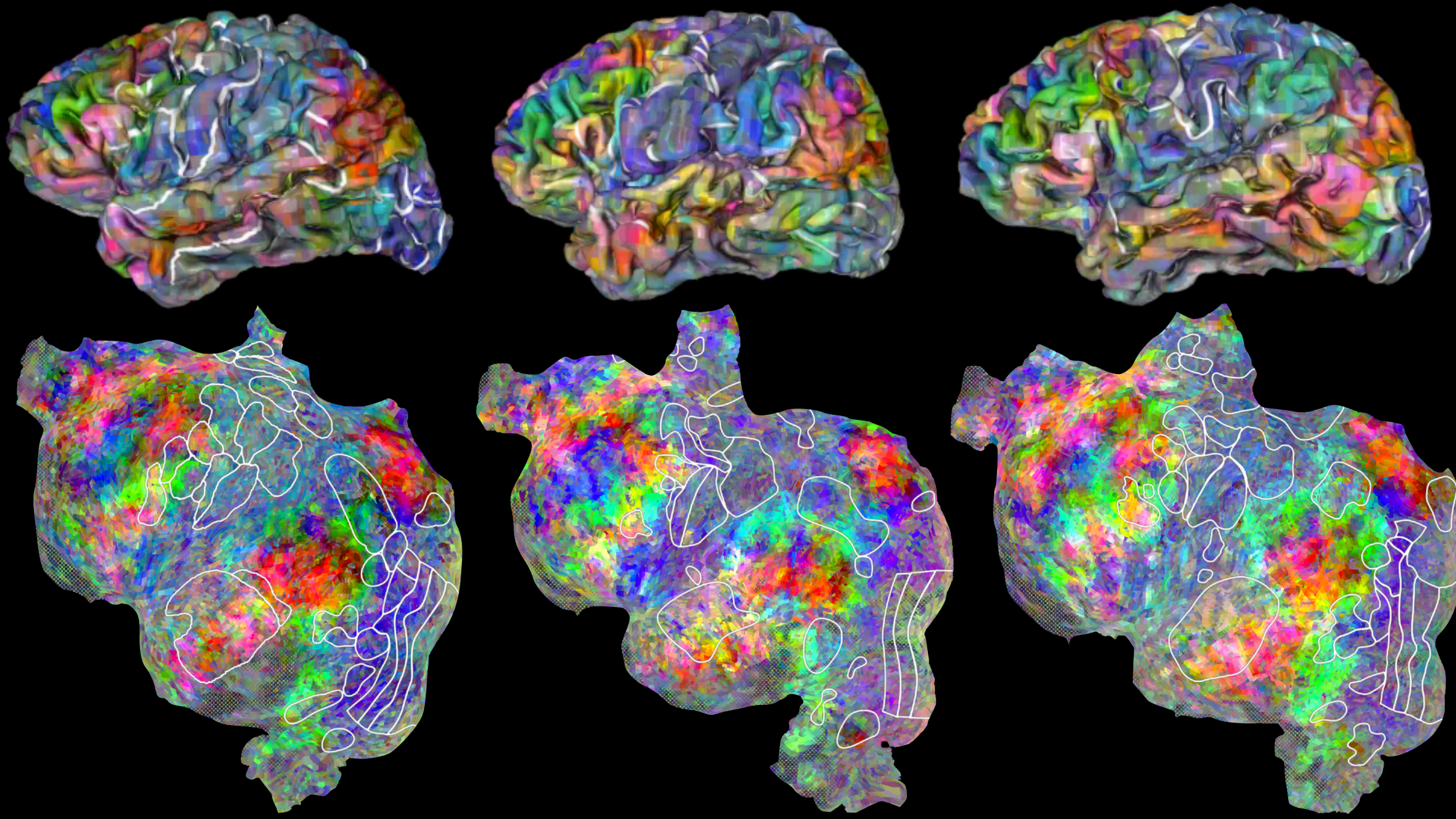
PREDICTION

VOXELWISE REGRESSION MODELS

Huth et al., *Nature* (2016), de Heer et al. *J Neurosci* (2017), Deniz et al., *J Neurosci* (2019)

MAPS OF SEMANTIC SELECTIVITY ACROSS CORTEX





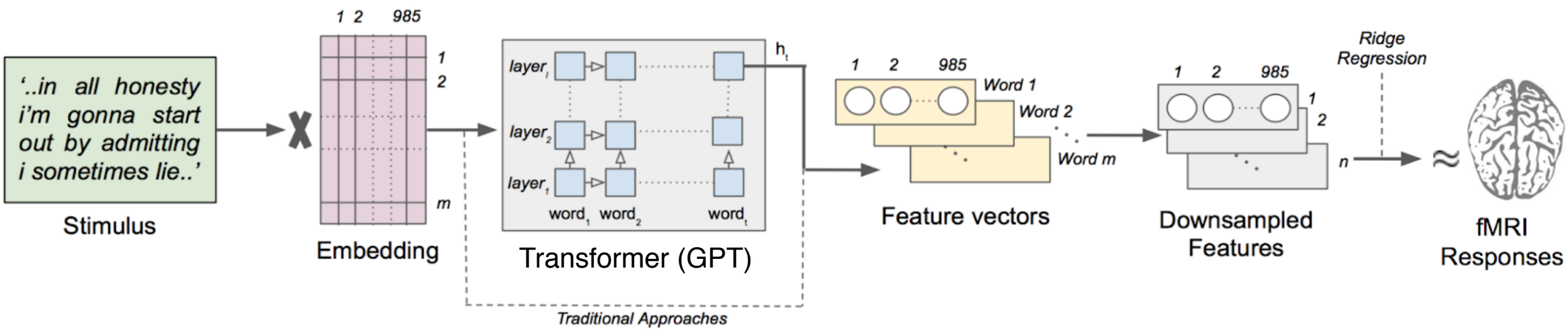
CONTEXT MODEL

Shailee Jain

Strategy:

1. Take **language model** (LM) that is trained to predict the next word from context

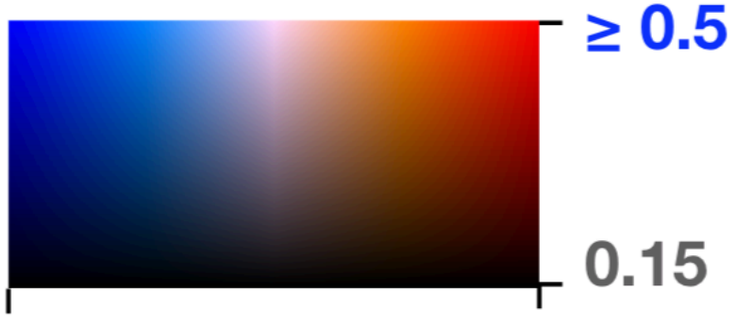
2. Use internal states of LM to model brain

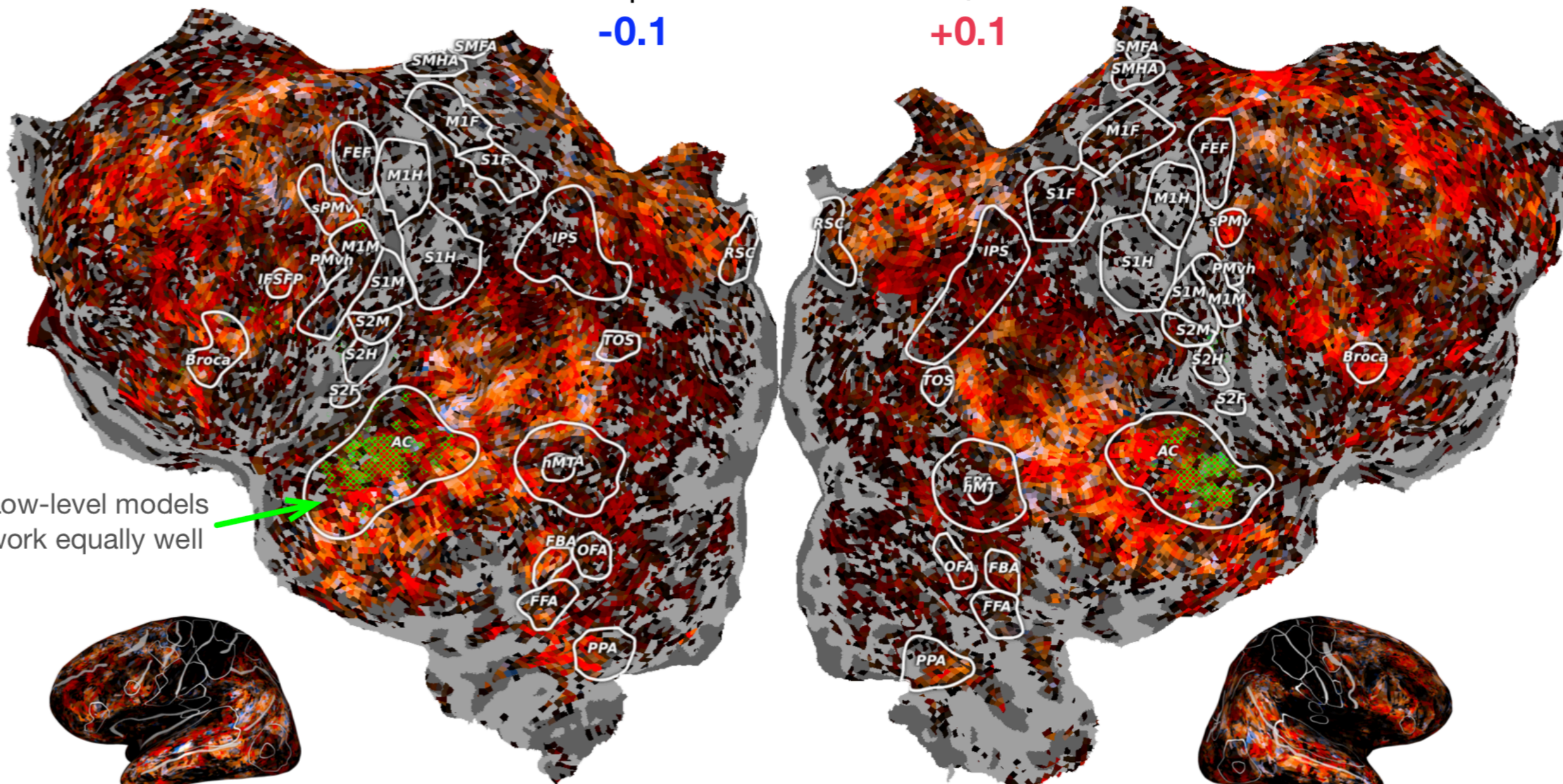


Jain & Huth, *NeurIPS* (2018); Jain et al., *NeurIPS* (2020); Jain et al., (*in prep.*)

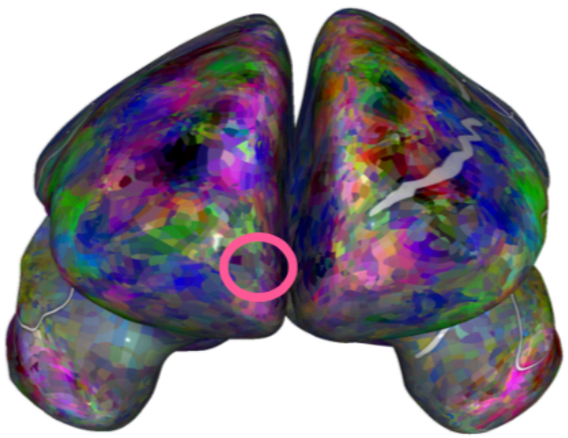
CONTEXT MODEL BEATS WORD MODEL

Old model better ⟷ **Context model better**

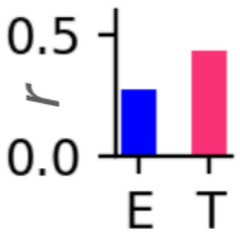




MODEL INTERPRETATION



Voxel: 42436
Medial frontal



Best Transformer Phrases

motioned for me to sit next to him and thanked me

started to ask the lady a question she said excuse me

and she comes out and she says look I'm really sorry

you should probably go check in there so I say thanks

Best Embedding Words

charming

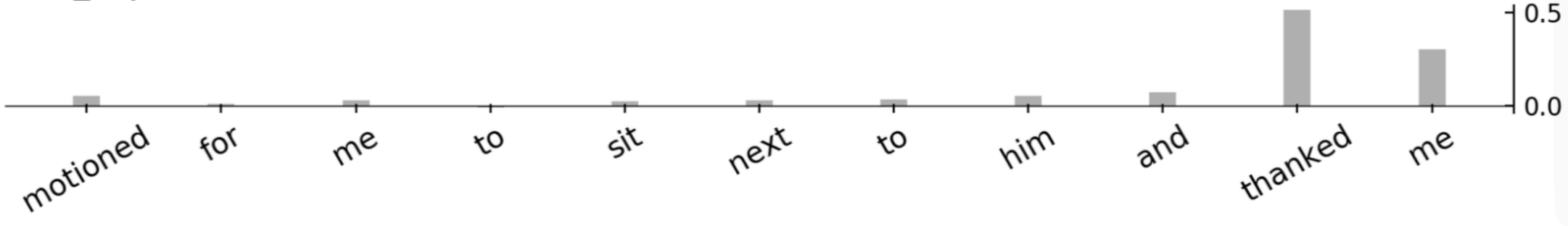
gentleman

pleased

smiled

friendly

kindness



HASTA LA VISTA