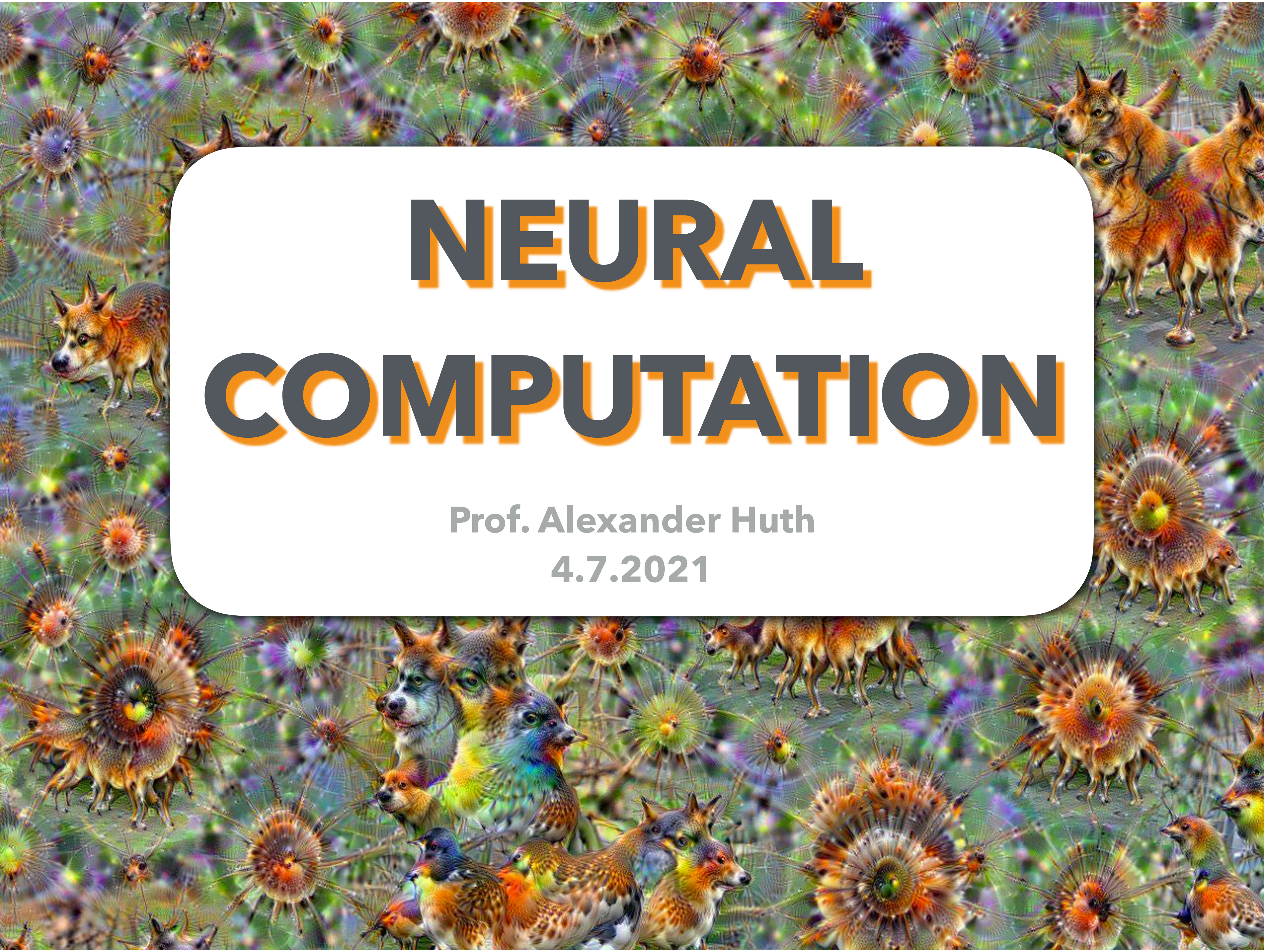


NEURAL COMPUTATION

Prof. Alexander Huth

4.7.2021

COURSE ADMIN / PROJECT

* Final project planning:
  * Post final project ideas on Canvas discussion (**April 12**)
  * Respond to someone else’s idea (**April 13**)
  * Self-organize into groups of 2-4 (??)
  * Work with group to write proposal & submit via email (**April 14**)

COURSE ADMIN / PROJECT

* Writeup (3-4 pages explaining background & what you did) due **May 3**

* In-class presentations (5-10 minutes) **May 3 & 5**

COURSE ADMIN / PROJECT

* Presentations should be live (zoom) and include slides describing the background, your method/approach, and results

* **It is essential** that every person in your group participates in both the project work and the presentation

PROJECT SUGGESTIONS

* **General goal** is to apply or explore something / anything we've talked about in this class
  * could be using real data (e.g. fit some kind of model to a neural dataset)
  * could be theory/methods (e.g. find a better way to do something)

PROJECT SUGGESTIONS

* Apply **system construction** pipeline:
  * Find a neuroscience dataset (visual, auditory, language, etc.)
  * Find (or build) an artificial neural network that solves a related problem
  * Use ANN activations to model neuroscience data

DATASET SUGGESTIONS

* Dataset suggestions:
  * CRCNS: https://crcns.org/data-sets
  * BOLD5000: https://bold5000.github.io/
  * Allen Inst.: http://www.brain-map.org
  * Study Forrest: http://studyforrest.org/
  * Huth 2016: https://github.com/huthlab/speechmodeltutorial

RECAP

* Linking artificial and biological neural networks

* In vision:

  * Train ANN to recognize objects

  * Use its representations to model visual cortex

  * This works pretty well (Eickenberg, Yamins)

RECAP

* How much does the *task* of the ANN matter?
  * Networks trained to classify images learn good representations for modeling brain
  * But random (untrained!) networks also work pretty well!

RECURRENT NEURAL NETWORKS

* Suppose there's a sequence of input output pairs $\{(x_t, y_t)\}_{t=1,2,\ldots,T}$
  * & suppose there are sequential relationships, i.e. y_t is dependent not only on x_t but also on previous x

RECURRENT NEURAL NETWORKS

* Suppose there's a sequence of input output pairs $\{(x_t, y_t)\}_{t=1,2,\ldots,T}$

* To apply a simple neural network we could feed in a bunch of x in parallel

  * But this explodes the number of parameters

RECURRENT NEURAL NETWORKS

* **Recurrent neural networks (RNNs)** can solve this problem

* RNNs can solve these sequence problems *without* a huge number of weights

RECURRENT NEURAL NETWORKS

* (Drawing of basic RNN)

LANGUAGE

* Experimental setting:
  * A subject listens to language (e.g. a narrative story) while BOLD responses are recorded from cortex using fMRI
  * We want to do **system identification**: build some model that predicts BOLD responses from the language stimuli

LANGUAGE

* To apply the “**system construction**” approach:

  * build an artificial neural network that solves **some task**

  * then use its representations (as a **linearizing transform**) to model the brain

* ***What task?***

LANGUAGE

* In vision: **object categorization** seems like a good task (modulo earlier discussion), because it requires the network to learn lots of things

* What is an equivalent for language?

  * Ideally something related to language **meaning**

LANGUAGE MODELS

* One solution may be **language models**, which have taken on a similar role in the NLP field to image classification models (e.g. AlexNet) in computer vision

* The task of a language model is to predict a word from its context

    * e.g. $P(w_i | w_1, \ldots, w_{i-1})$

LANGUAGE MODELS

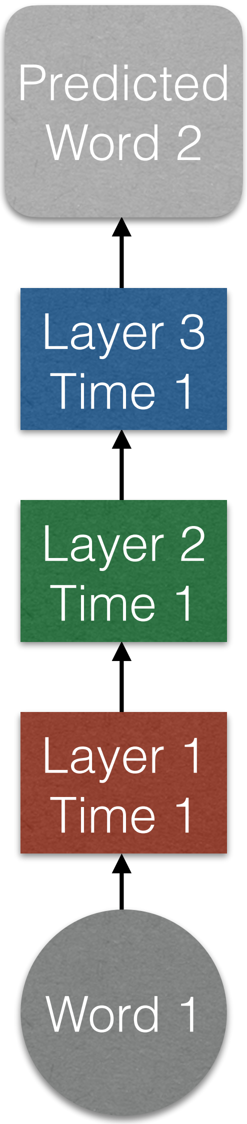
* Language models can use many different architectures

* One is a **recurrent neural network (RNN)**

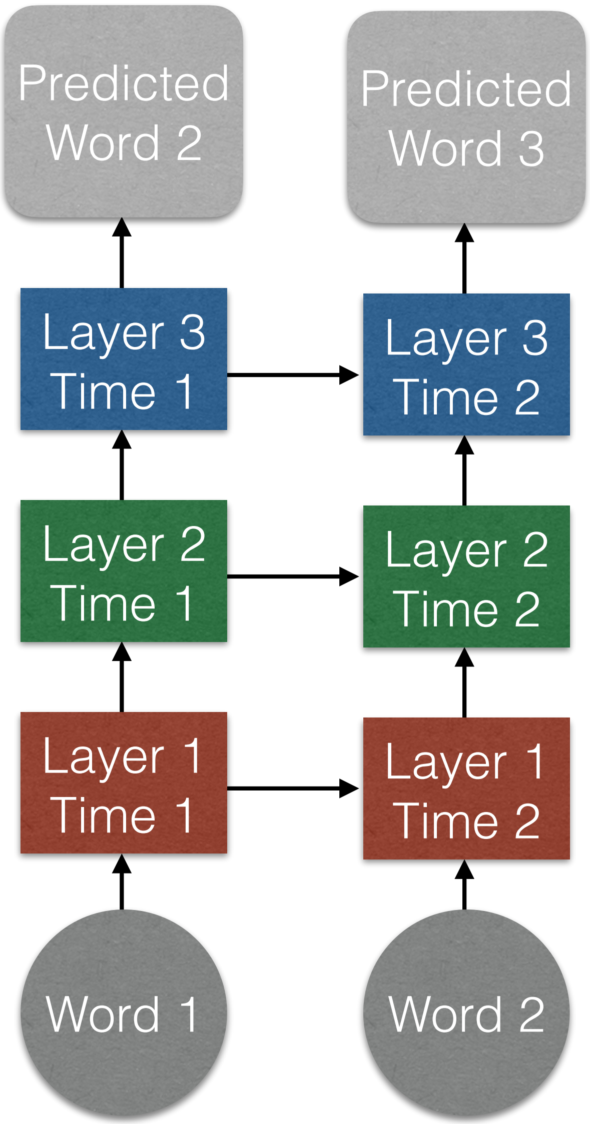
* In particular, a variant RNN called a **long short-term memory (LSTM) network**

  * (We'll talk talk about this network & how it works in detail on Monday)

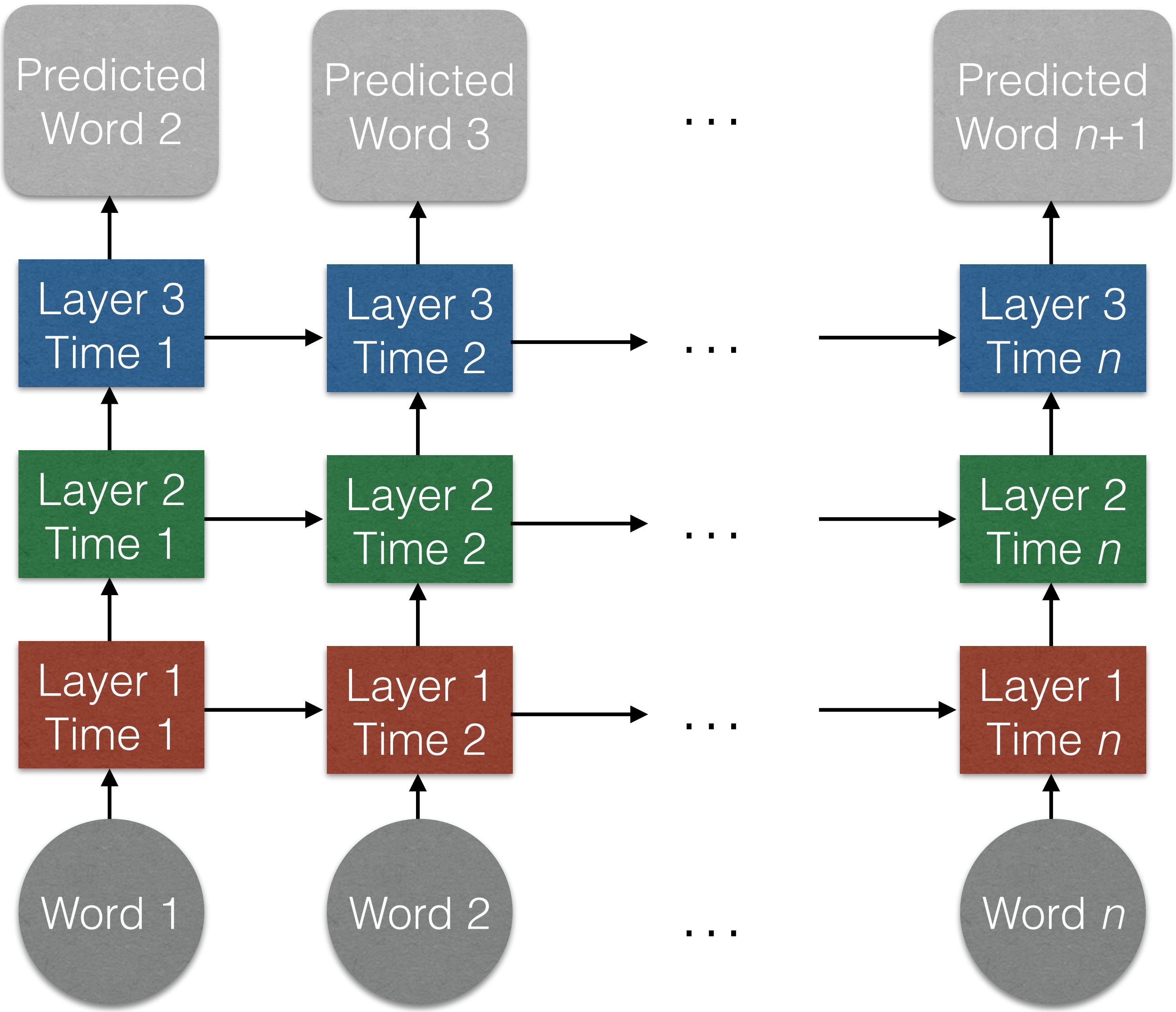
LSTM LANGUAGE MODEL



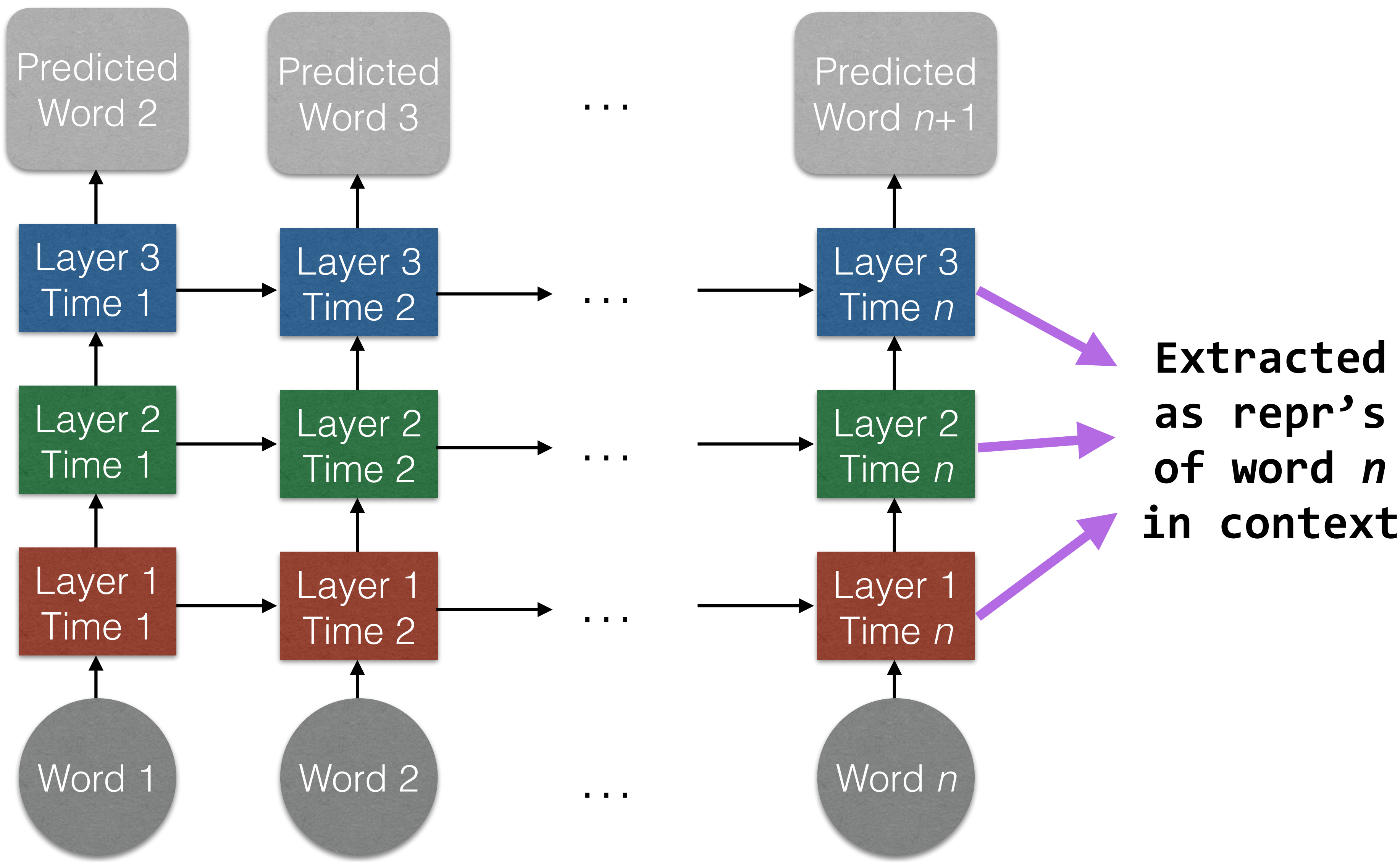
LSTM LANGUAGE MODEL



LSTM LANGUAGE MODEL



LSTM LANGUAGE MODEL



FMRI EXAMPLE USING LSTM LANGUAGE MODEL

* Jain & Huth (2018) Incorporating context into language encoding models for fMRI. *NeurIPS*

FMRI EXAMPLE USING LSTM LANGUAGE MODEL

* Approach:
  * Train LSTM language model on text
  * Do fMRI experiment
  * Use LSTM language model to extract features from fMRI stimuli
  * Build linearized system identification model using LSTM-derived features

FMRI EXAMPLE USING LSTM LANGUAGE MODEL

* In visual models different features can be extracted from different **layers**

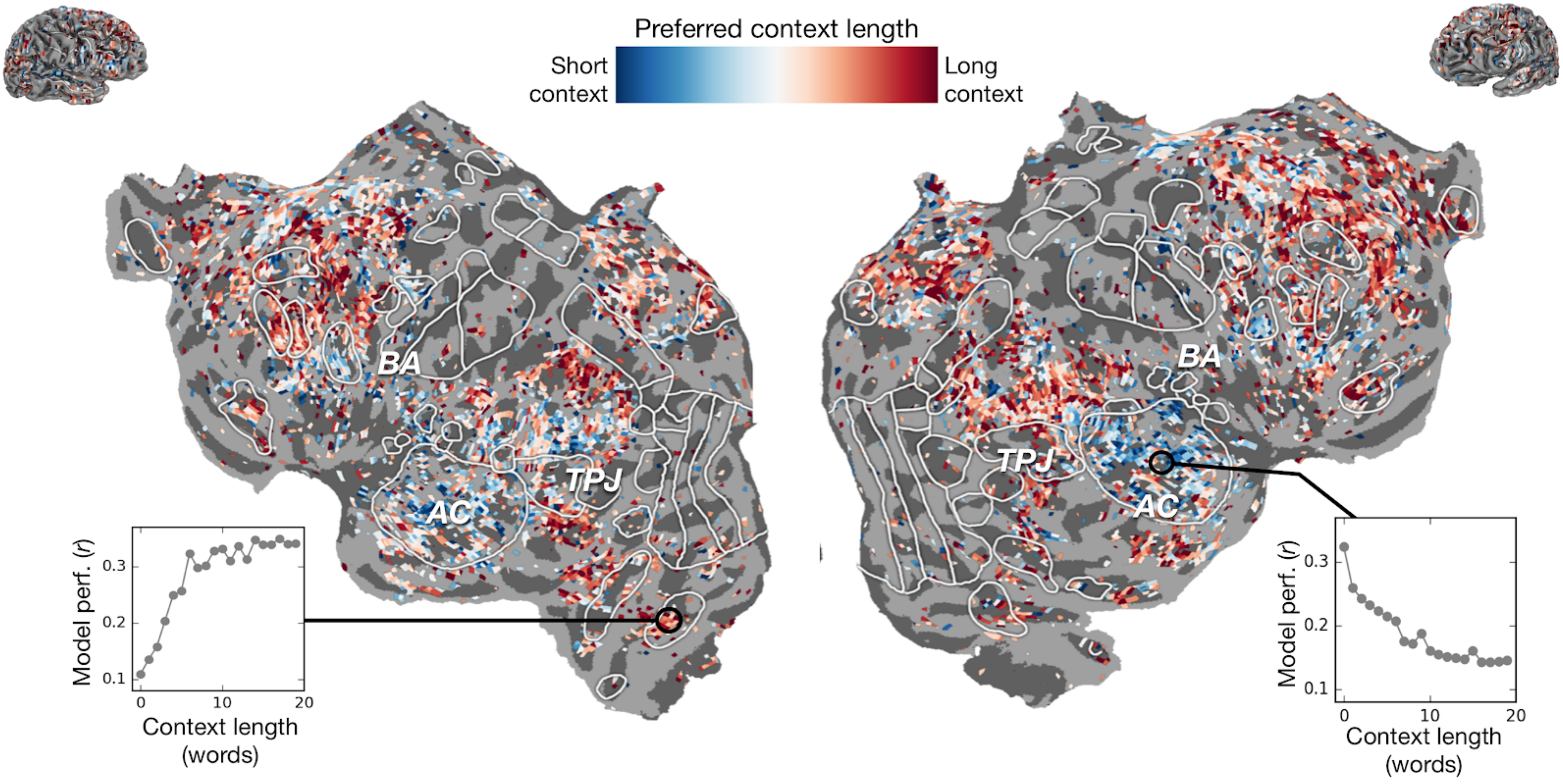
* In language models different features can be extracted from different **layers** *and* with different **amounts of context**

  * e.g. $P(w_i | w_{i-c}, \ldots, w_{i-1})$

    "context length"

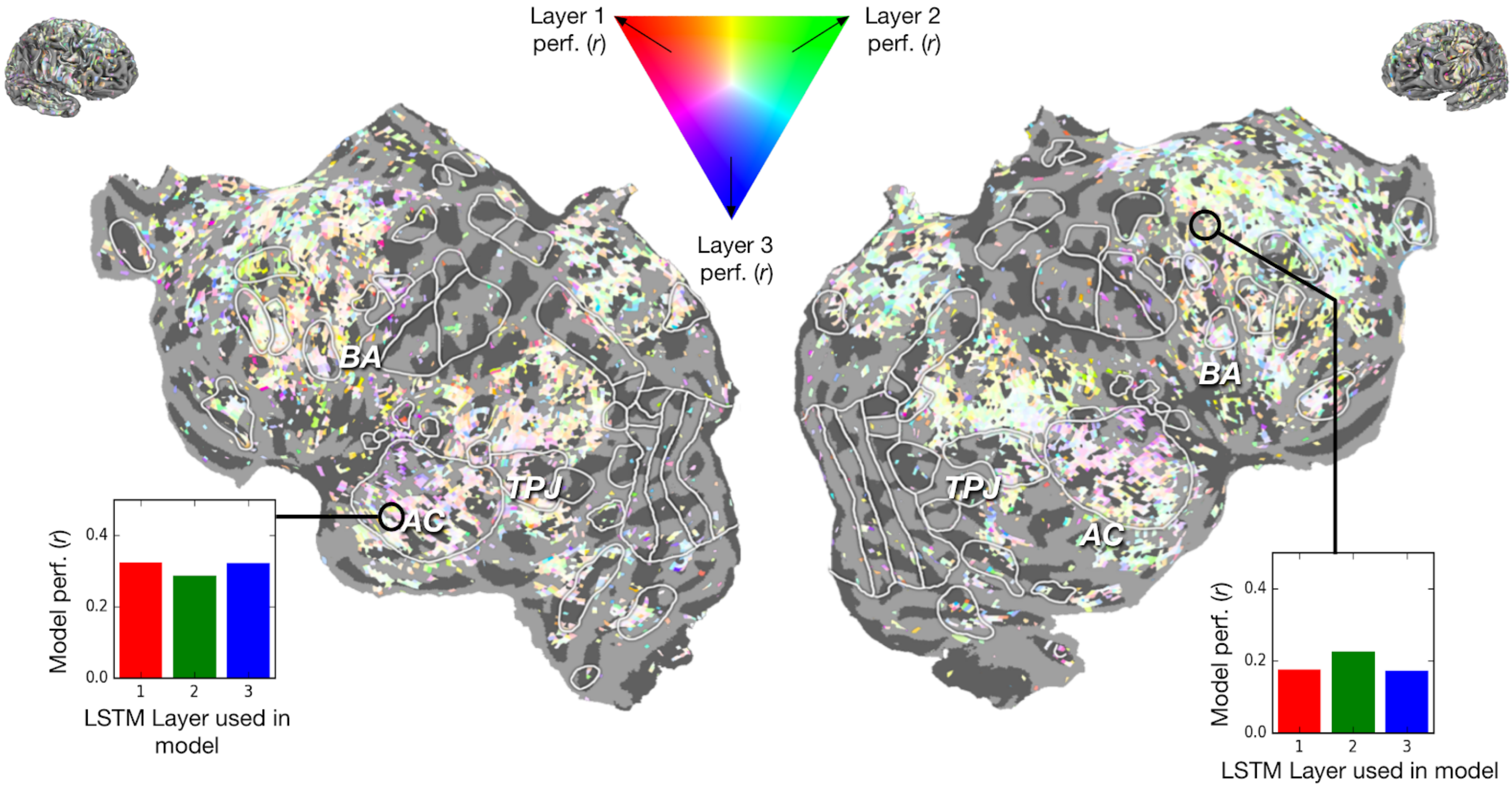
FMRI LSTM EXAMPLE

* Different brain areas prefer different amounts of context



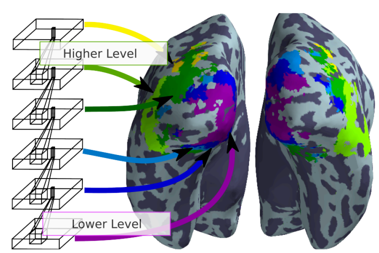
FMRI LSTM EXAMPLE

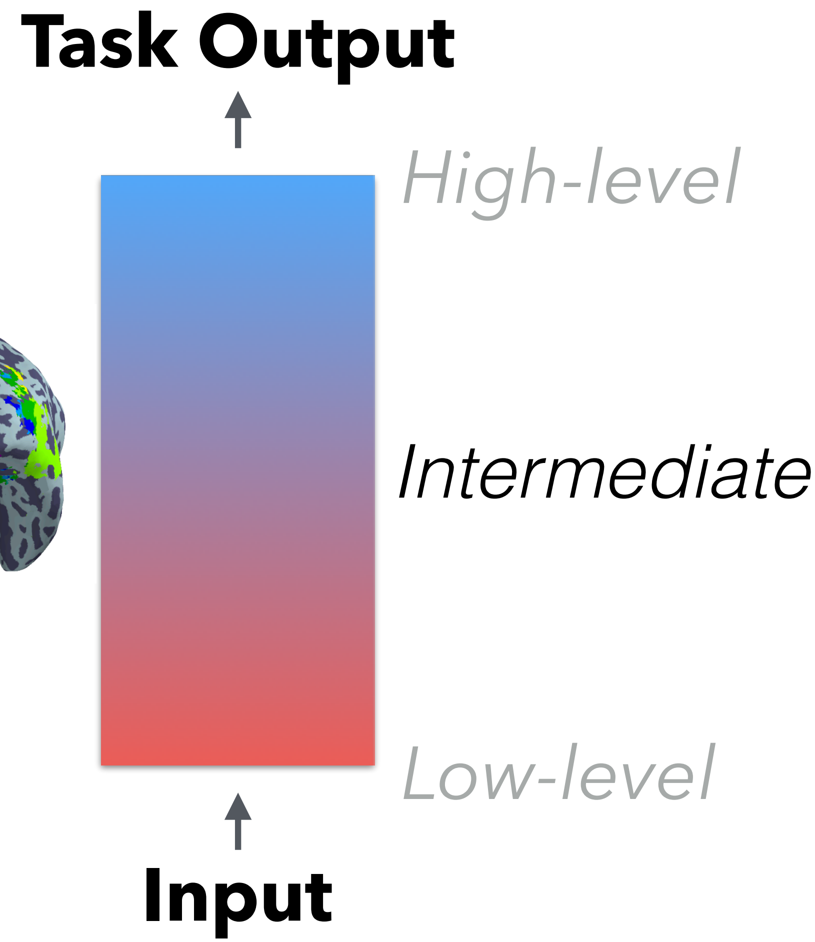
* But “layer preference” does not recapitulate known hierarchies, unlike Eickenberg, etc.



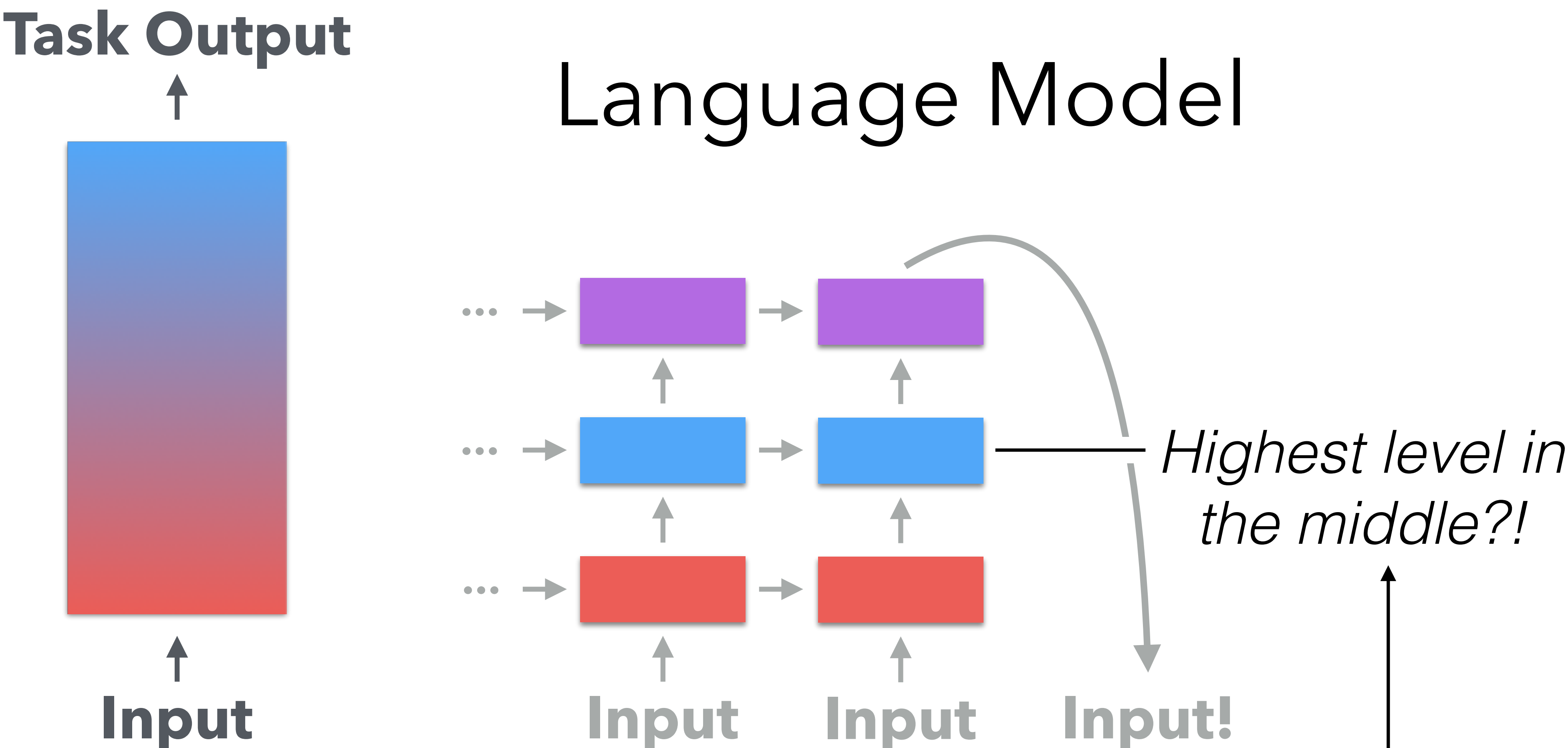
FMRI LSTM EXAMPLE

* In visual models, there is a clear “progression” of representations from low- to high-level





FMRI LSTM EXAMPLE



Also seen in Toneva & Wehbe *NeurIPS 2019*

NEXT TIME

* Problems with simple RNNs: vanishing/ exploding gradients

* Other neural network architectures for language modeling: LSTMs & Transfomers