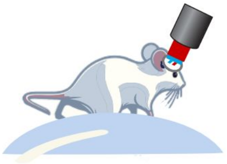
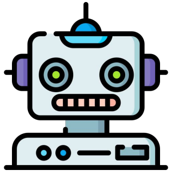
Developing Models for the Next Generation of Neuro Inspired AI Research





Antonis Antoniades (@anton_iades)
University of California, Santa Barbara

Who am I? 👀

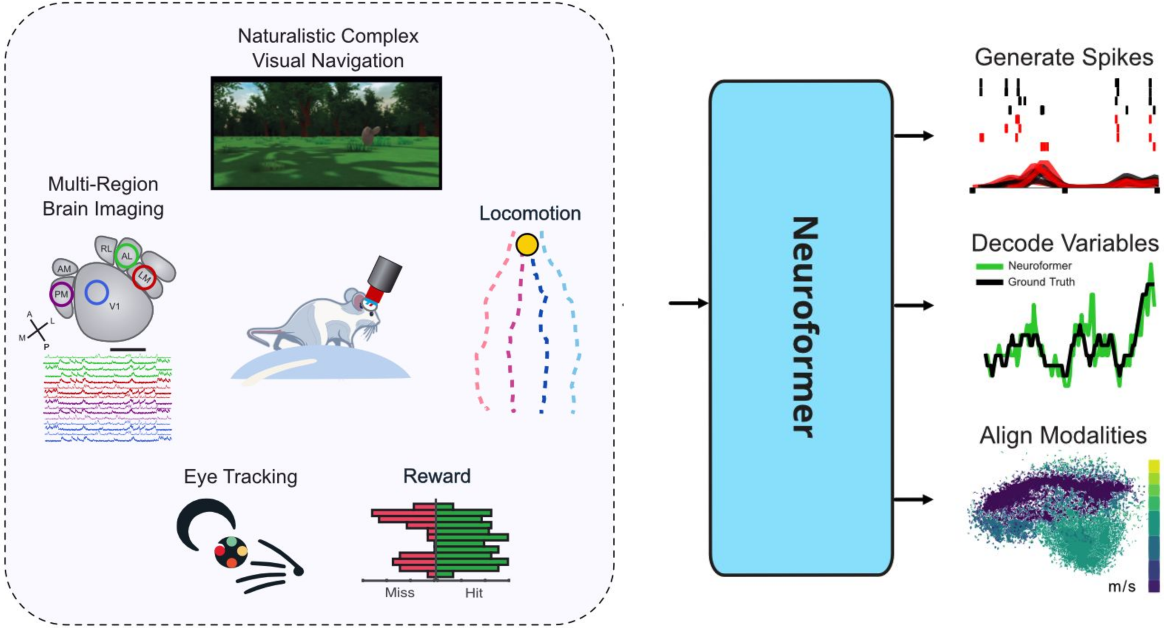
- Originally from the island of Cyprus
- Served 2 years in Cyprus Special Forces
- UCSB Grad 21’ (Physics)
- Interned at Odyssey CS, NodeDistrict, Leela AI
- Co-founded a health tech company called Calibrex
- Interested in using AI to make scientific discoveries (particularly in Neuroscience)
  - Was working for the Smith Neuro Lab as an undergraduate at UCSB
- Currently a 2nd year PhD within UCSB NLP Group (CS), working across machine learning (Dr. William Wang) and neuroscience (Dr. Spencer Smith)
- I like to play the guitar and greek bouzouki

Multimodal and Multitask Neuroscience Experiments

Joseph Canzano, Smith LAB (UCSB)

www.slslab.org

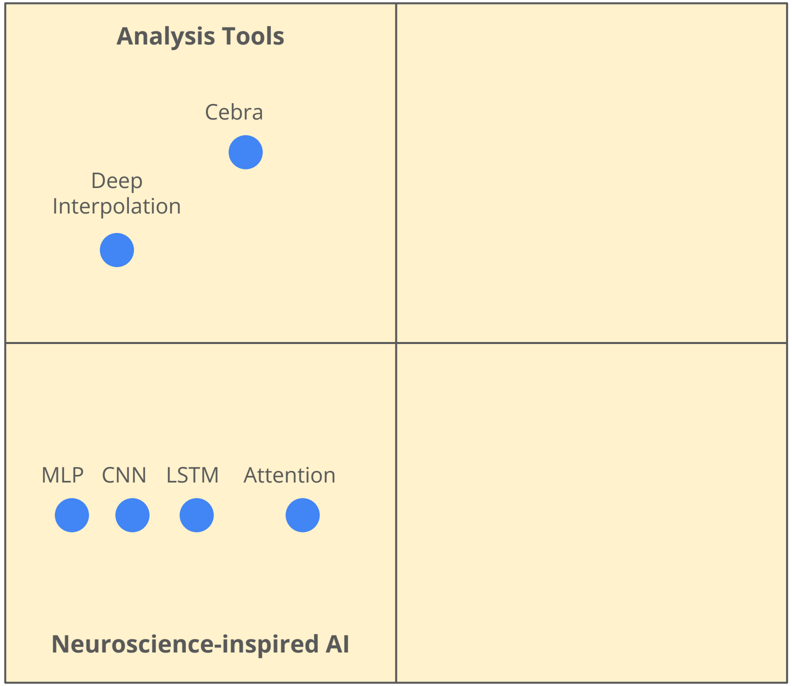
Multimodal and Multitask Neuroscience Experiments



Goal: Build a **data-driven**, **flexible** framework to inspire new, large-**scale**, **multimodal** Neuroscience research

Neuroscience + AI Landscape

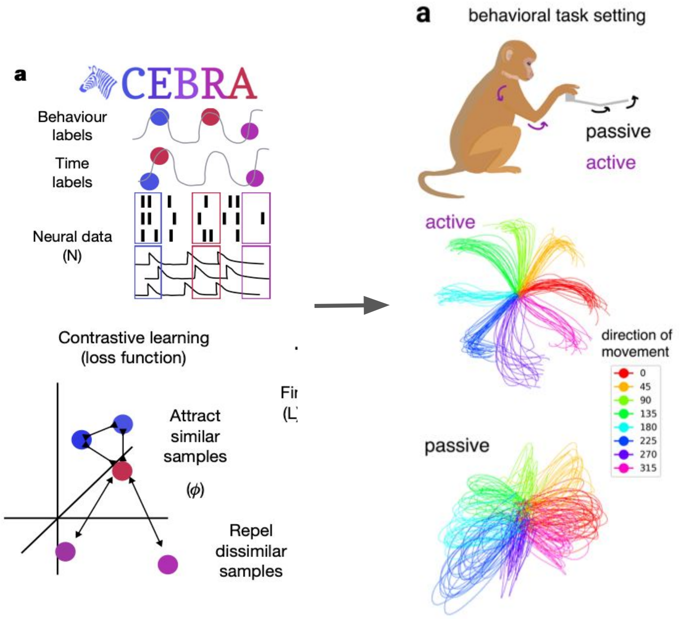
The current Neuroscience + AI landscape



Representation Learning of Population Activity using Contrastive Learning

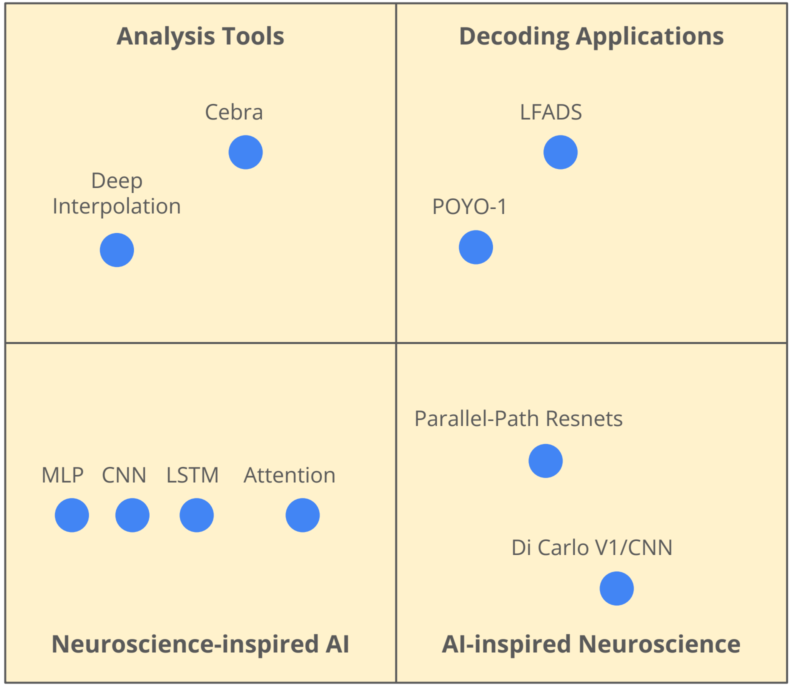
Learn multimodal low-dimensional latents

- **Strong** vs. **Weak** Principle
  - Raw vs. Compressed data
- Even **latents** are high-dimensional
- Low-Dimensional Latents, while interpretable, **do not provide any extensive insights.**



Learnable latent embeddings for joint behavioral and neural analysis, Schneider, Lee et al., (2022)

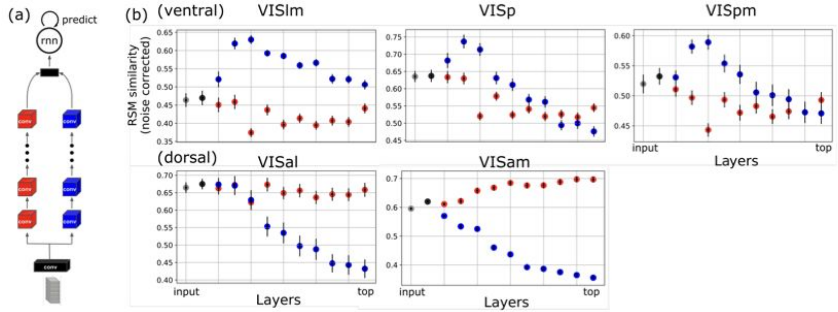
The current Neuroscience + AI landscape



Representational Similarities between ANNs and Mammalian Brain

Validating Brain Principles in code

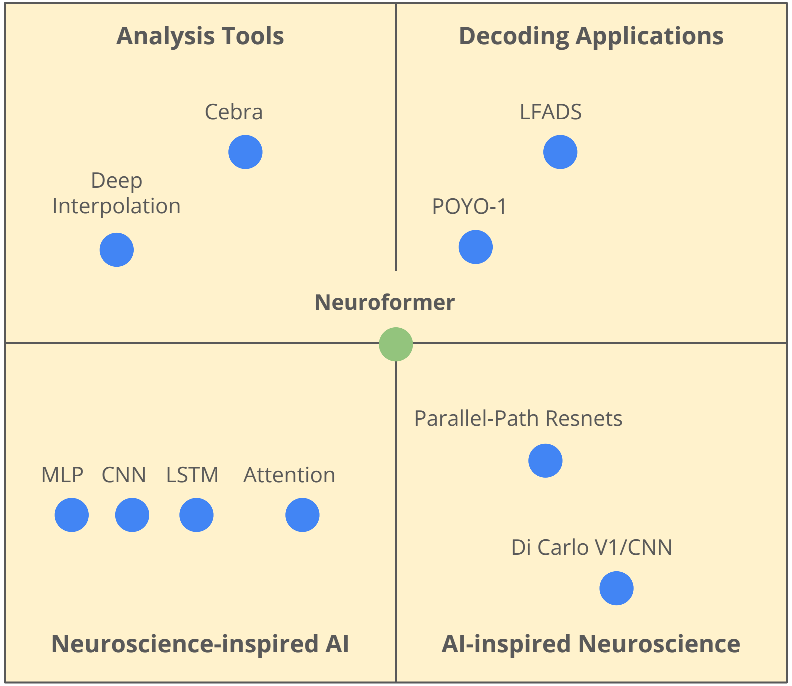
- Train **2-path resnet network**
- Compare representations to **dorsal/ventral stream** data
- 2-path ResNet **splits into representations** that resemble **ventral and dorsal streams**



The functional specialization of visual cortex emerges from training parallel pathways with self-supervised predictive learning. Bakhtiari et al., (2021)

Opportunity: How about we constrain model using the brain data itself?

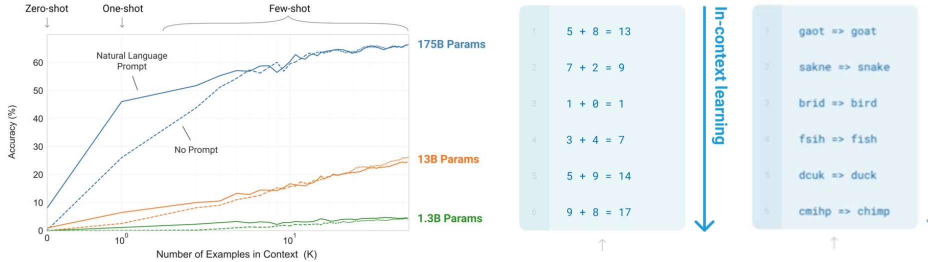
The current Neuroscience + AI landscape



The Bitter Lesson: ***Scale***

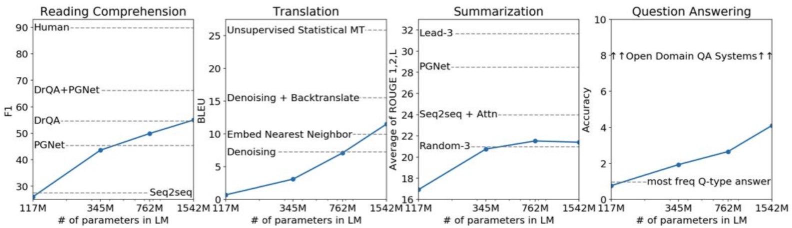
Compute + ***Data*** drives the vast majority of AI progress (Richard Sutton, seminal RL Professor)

GPT-3



Larger models **can learn new patterns** and tasks with **no training = *in-context learning***
(Brown et al., 2020)

GPT-2



Self-supervised **pre-training** on **lots of data**, yields language models that perform well on **various downstream tasks.**
(Radford et al., 2019)

Performance

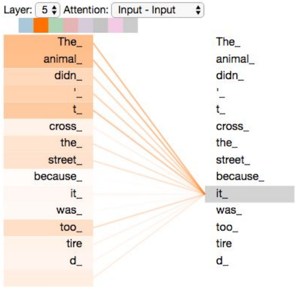
Parameters + Data *(log scale)*

Transformers

Generalization of an MLP

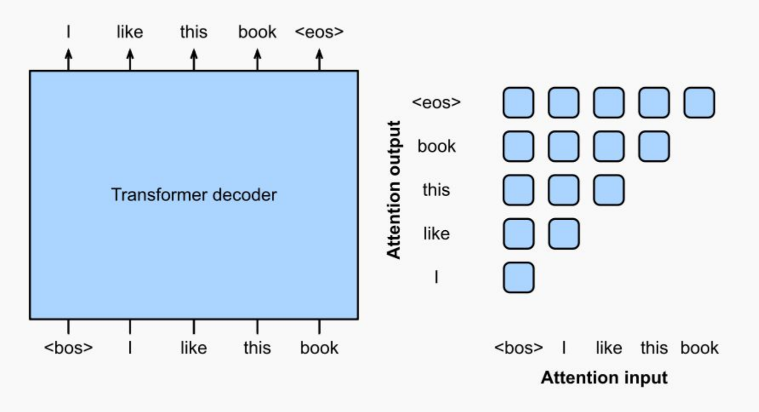
- Discretize data into **tokens**

- Process feature representations by iteratively unraveling the **relationships** between the discrete tokens using **attention**



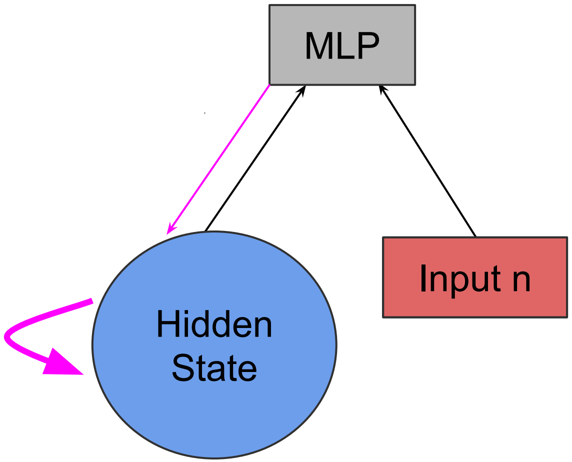
Vawani et al., 2017

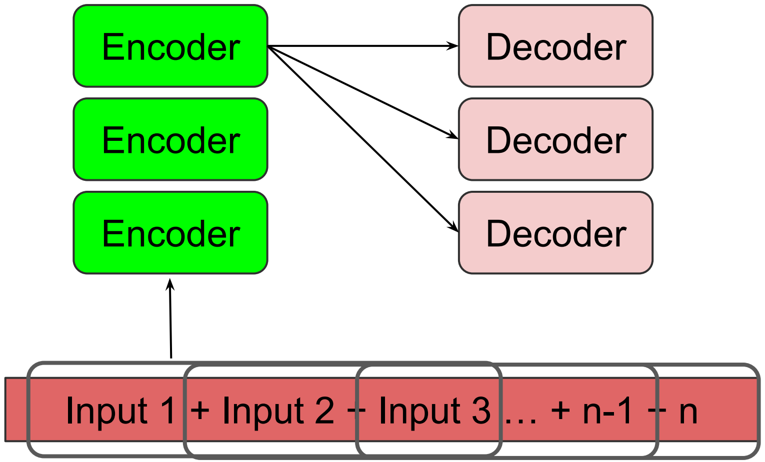
- Trained to **predict the next word** - *across the whole internet.*



Dive into Deeplearning, 2023

Recurrent Networks vs. Transformer Networks





RNN Tries to ***squeeze* all previously seen information** inside a hidden state.

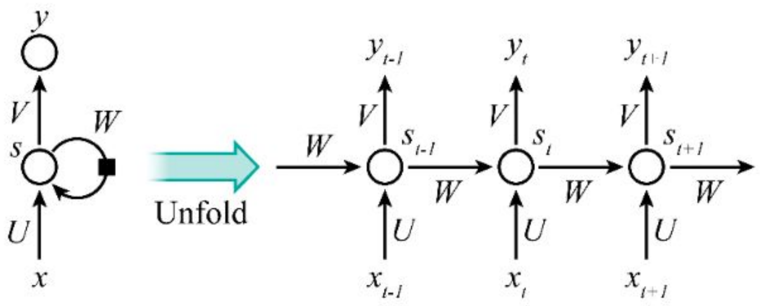
Autoregressive Transformer ***sees* all previous information** at each time step. (within a sliding window)

Solves: Sparseness in *Time*

Recurrent Networks vs. Transformer Networks

Actually - **it's mostly about scale.**

- You cannot stack enough layers of RNNs to reach the size of GPTs.
- Hidden state is an information bottleneck.
- Causes exploding/vanishing gradients which make training unstable.



W is *large* = ***exploding*** *gradients*

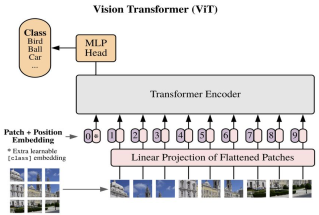
W is *small* = ***vanishing*** *gradients*

RNNs are typically stable at **3-5 layers**. GPT-3 has **96 layers**.

Transformers and Multi-Modality

Unlike Models such as CNNs, the transformer makes little assumptions about the input modality (**weak *inductive bias***).

- Proved effective at also processing *images, sound* and more

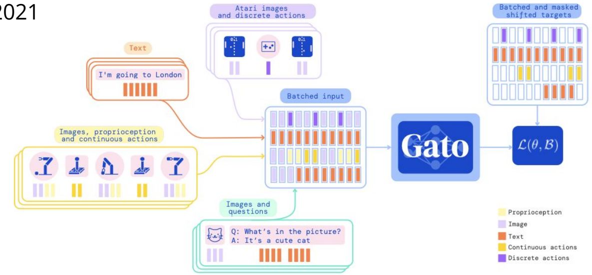


Dosovitskiy et al., 2021

- Led to performant models which can combine multiple modalities

Lee et al., 2021

- Models trained on multiple modalities (**multimodal**), and multiple tasks (**multitask**)



Reed, Zolna, Perisotto et al., 2022

Building a multimodal model of the brain

Goal: Build a **data-driven**, **flexible** framework to inspire new, large-**scale**, **multimodal** Neuroscience research

Brain	**Transformers**
Large number of **neurons**	Can learn representations for **many tokens**
Sparse	Can learn relationships across **large context** windows
Diverse number of **inputs**	Modality-**Agnostic**
Constrained by architecture of brain and **connectivity pattern**	**Unsupervised** Learners

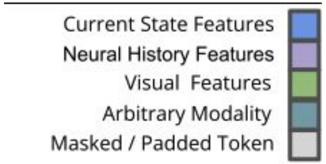
Solution: Build a Large **multimodal** Neuroscience Model! (LNM?!)

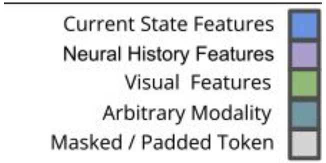
Neuroformer

(architecture)

Cosyne 23, ICLR 24

Neuroformer *(0 - preliminaries)*





Sequence of **Neuron firings** at **current** time window

Sequence of **Neuron firings** within **previous** time window

Full FOV video **stimulus** presented to mice within current time window

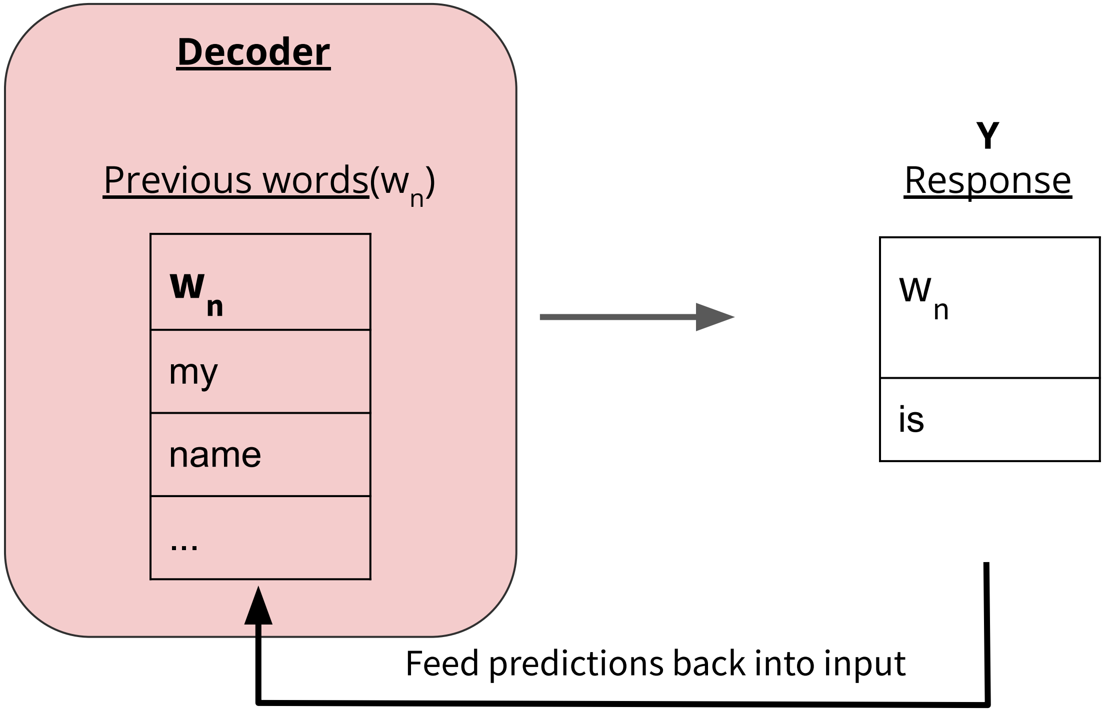
Any other **modality** (we want to use pose, eye movement etc)

Attention cannot attend to these positions

Pretraining Objective - *Language*

- Our goal is to predict the next most likely **word** (**w_n**) given all **previous words**

$$P(w_n \mid w_{n-1} \ldots)$$

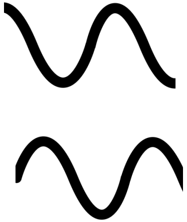


Pretraining Objective - *Language*

- Our goal is to predict the next most likely **word** (**w_n**) given all **previous words**

$$P(w_n \mid w_{n-1} \ldots)$$

Attention operation is permutation equivariant. Need to bias logits



+

+

$$PE_{(pos,2i)} = sin(pos/10000^{2i/d_{\text{model}}})$$
$$PE_{(pos,2i+1)} = cos(pos/10000^{2i/d_{\text{model}}})$$

Decoder

Previous words(w_n)

w_n
my
name
...

→

Y
Response

w_n
is

Feed predictions back into input

Pretraining Objective - *Neuroformer*

- Our goal is to predict the next most likely neuron **(id_n)** that will fire *and* ***when*** **(t_n)** given previous spikes

$$P(\,id_n, t_n \mid id_{n-1}, t_{n-1} \ldots)$$

Decoder

Population History (id_n)

id_n	T
55	0.01
112	0.02
...	...

Y

Response

ID	T
25	0.03

Feed predictions back into input

Pretraining Objective - *Neuroformer* *(multimodal)*

- Our goal is to predict the next most likely neuron **(id)** that will fire *and* ***when (dt)*** given previous spikes *and* ***modalities***

$$P(id_n, t_n \mid id_{n-1}, t_{n-1}, s_{n-1} \ldots)$$

Stimulus (S)

Stimulus	ΔT (s)
Frame 1	0.0 - 0.2
Frame 2	0.1 - 0.3
Frame 3	0.2 - 0.4

\+

Decoder

Population History (id_n)

id_n	T (s)
55	0.01
112	0.02
...	...

Y

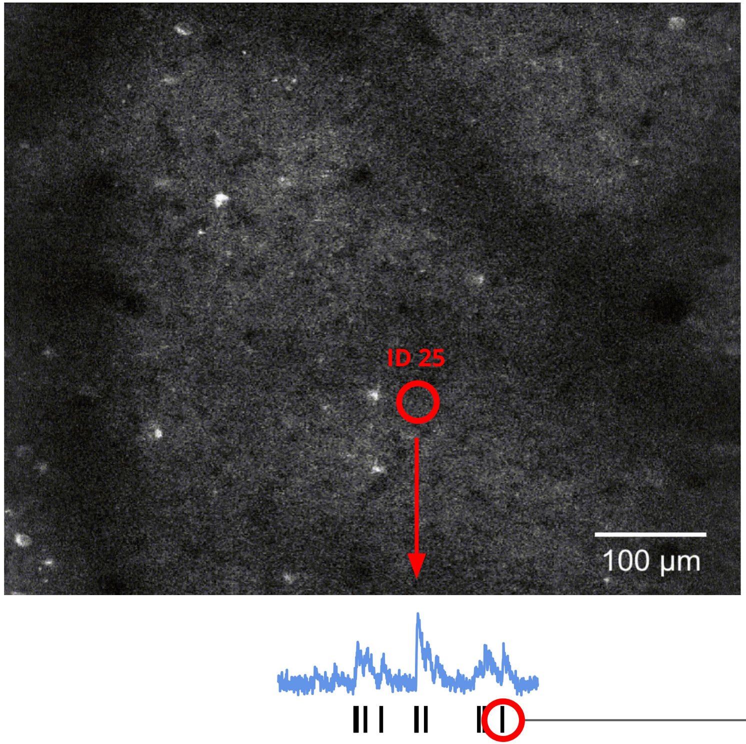
Response

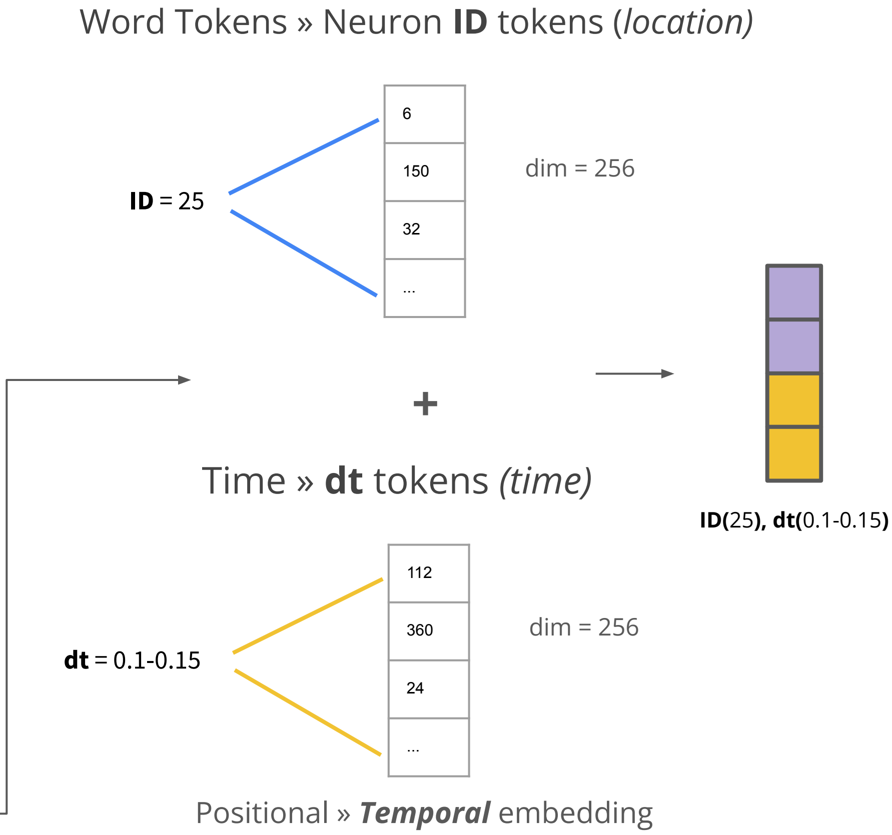
id_n	T
25	0.03

Feed predictions back into input

Spike (ID + dt) Tokenization

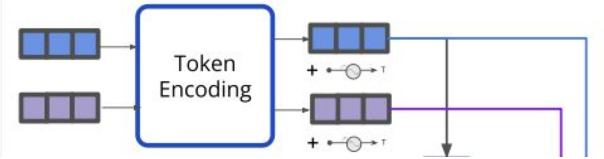
Dr. Che-Hang Yu, Smith Lab (UCSB)



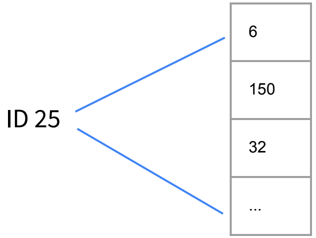


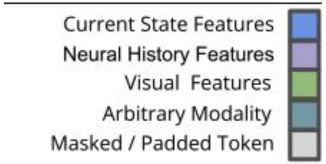
Feature Backbone

Neural

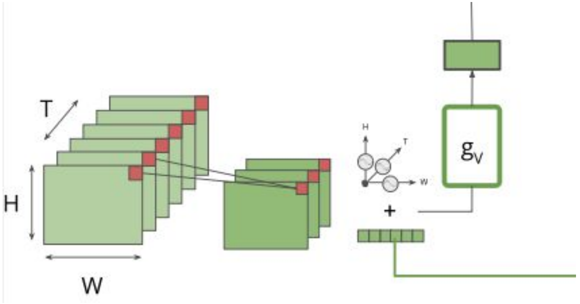


- Key-Value codebook, where keys are neuron IDs





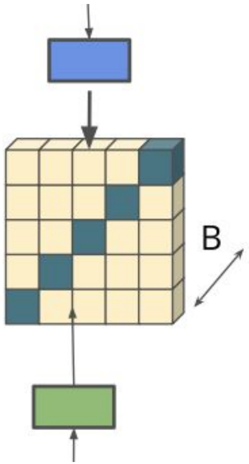
Visual



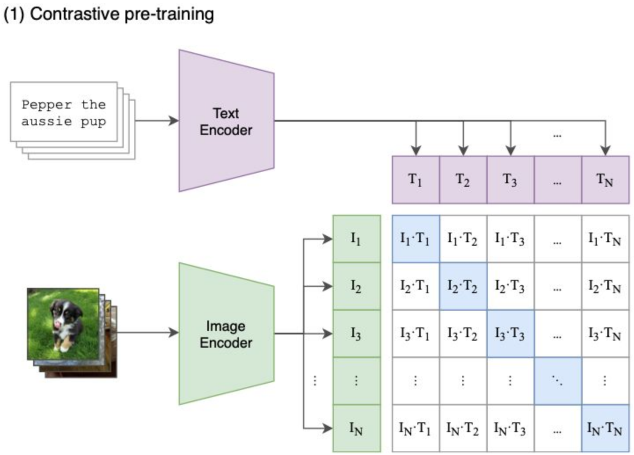
- Randomly Initialized Convolutions
- Pre-trained ResNet backbone Convolutions
- Raw Frames

Neuroformer (*2 - feature alignment)*

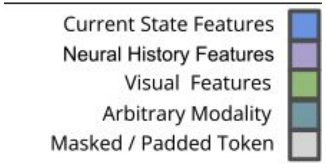
Use cosine similarity to maximize resemblance of coinciding visual and neural features



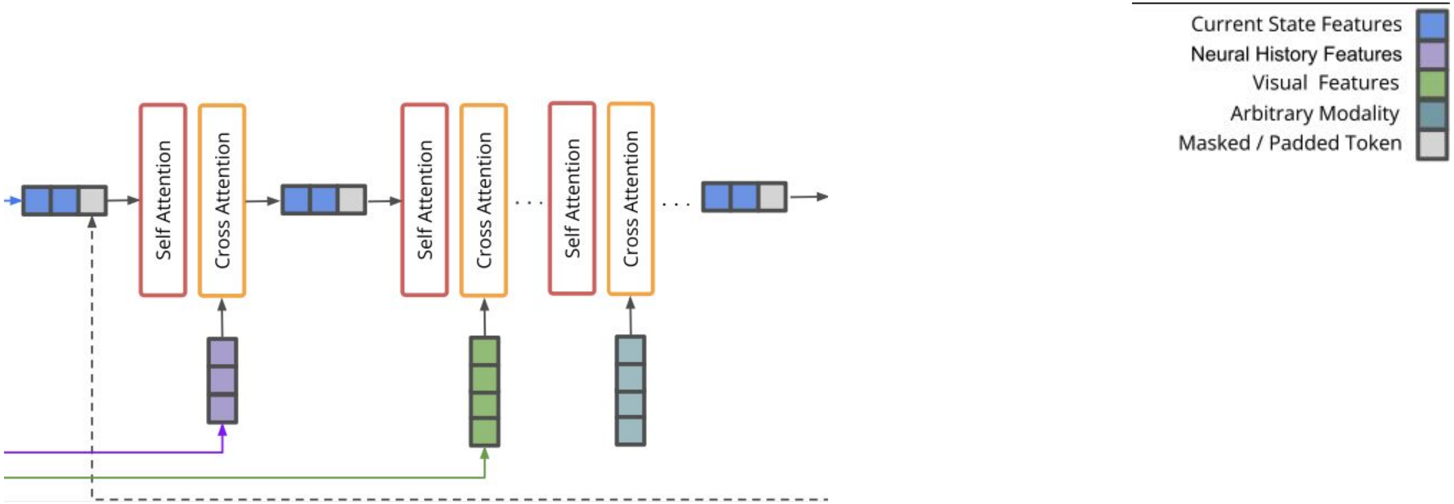
Using Negative examples to *contrast* our positive pair avoids collapse.



Radford et al., 2021

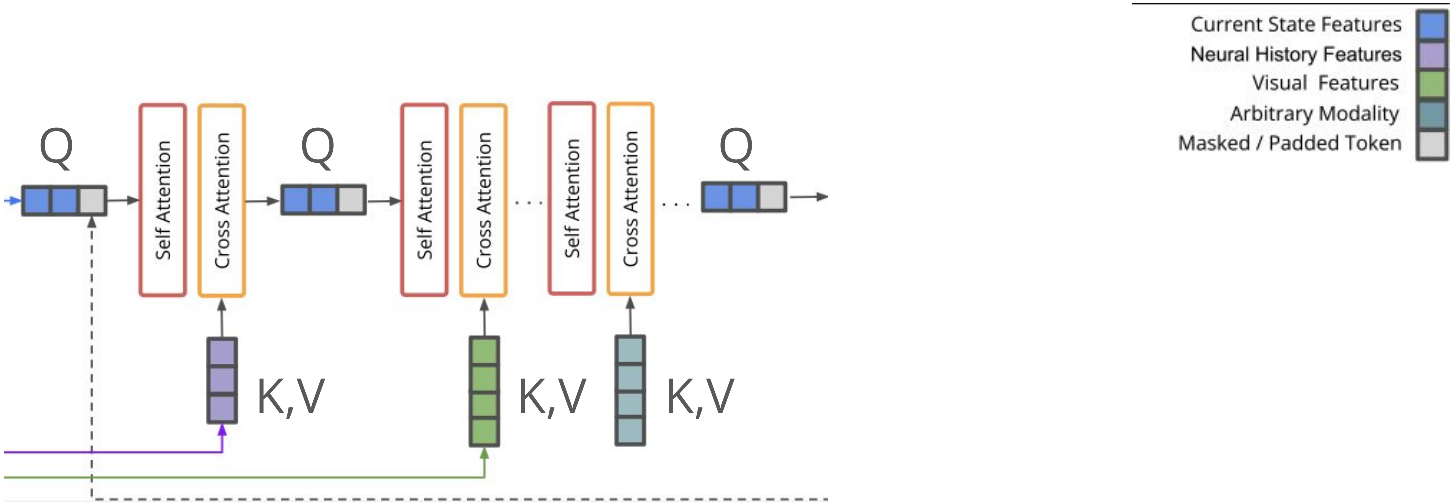


Neuroformer (*3 - feature fusion)*



Lee et al., 2021

Neuroformer (*3 - feature fusion)*

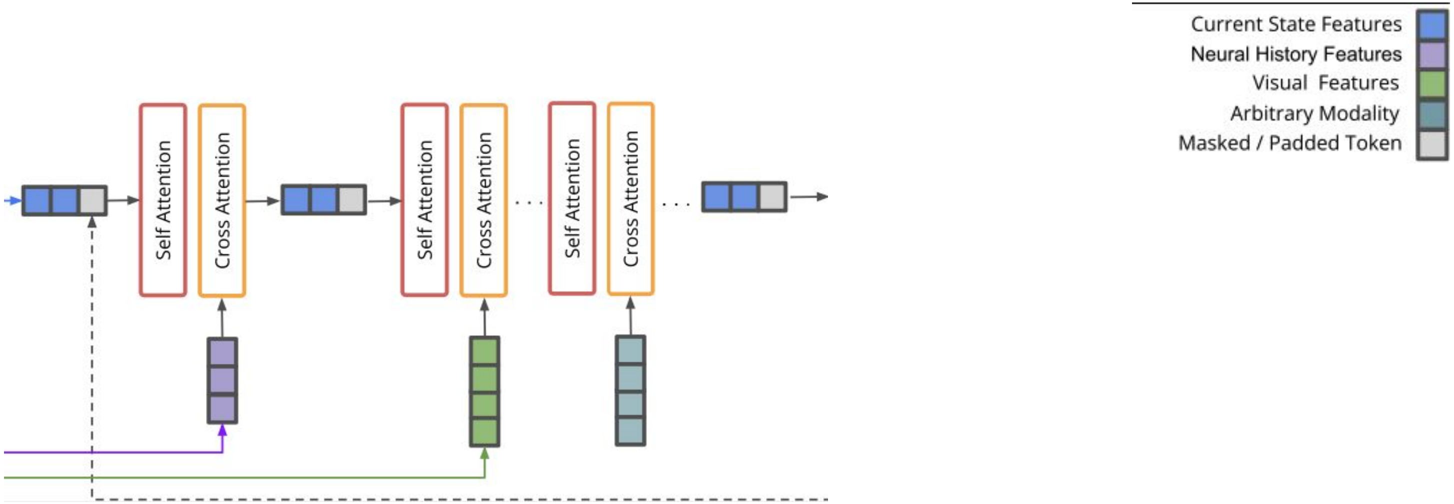


Cross Attention Maintains shape of Q.

$Q_{dim} = (B, T, E)$
$K_{dim}, V_{dim} = (B, S, E)$

$$\textbf{Attention} \sim (QK^T) \times V \sim \underset{Q}{(B, T, E)} \times \underset{K}{(B, E, S)} \times \underset{V}{(B, S, E)} = (B, T, E) = \mathbf{Q_{dim}}$$

Neuroformer (*3 - feature fusion)*

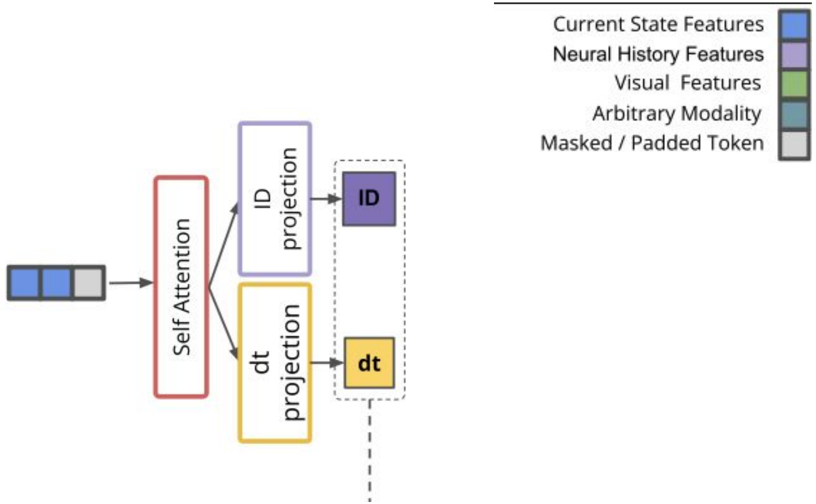


Matrix Sizes
Q = (m x 1), K = (n x 1), V = (k) = (n x 1)
where *m small, n large*

Attention formulation **(n x n)**
(QK^T) V = (m x 1) (1 x n) (v x n) = **(m x n)** (n x 1) = (m x 1)

Normally
$O(n^2)$ (*quadratic complexity)*
Ours
O(mn) (*linear complexity)*

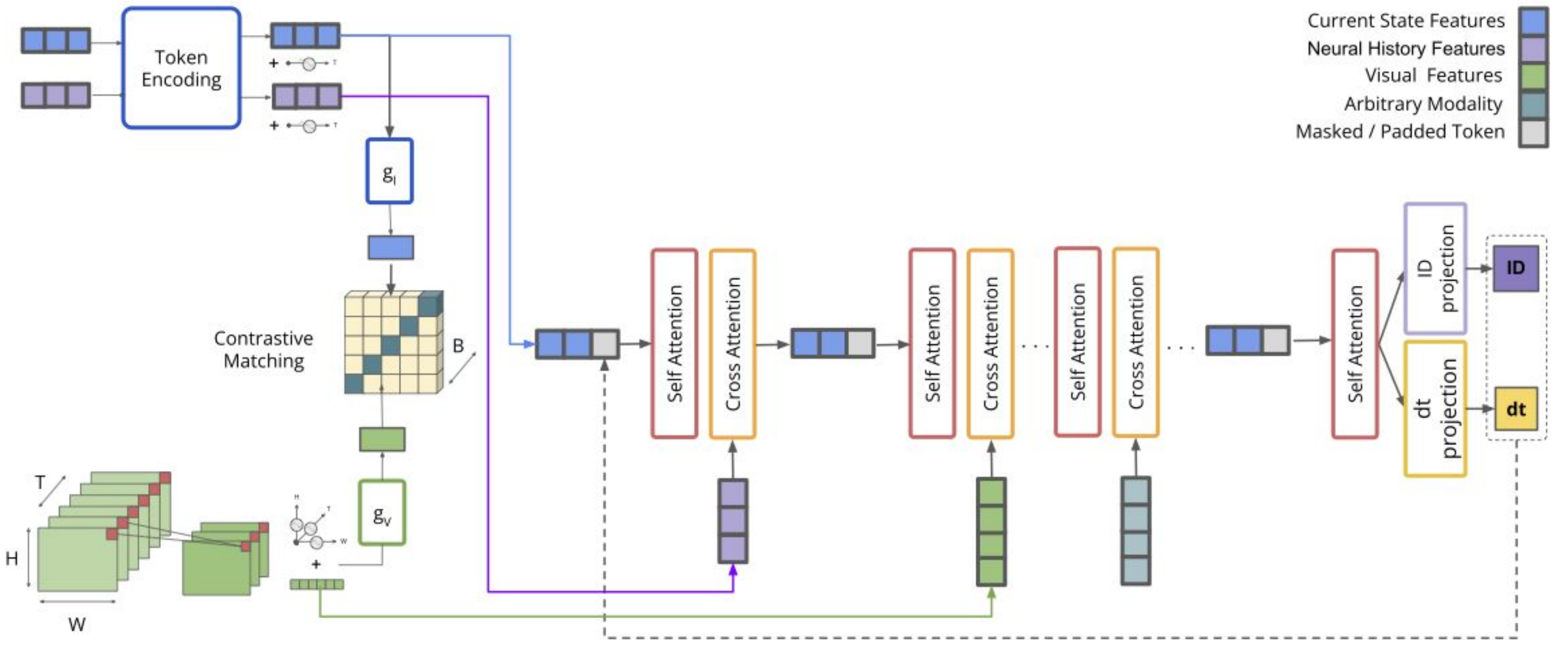
Neuroformer (*4 - decoding)*



$$P(\, id_n, t_n \mid id_{n-1}, t_{n-1}, s_{n-1} \ldots)$$

Probability of each neuron in population to fire at current time given ***past spikes, stimulus*** *(and/or* ***other modalities****)*

Neuroformer



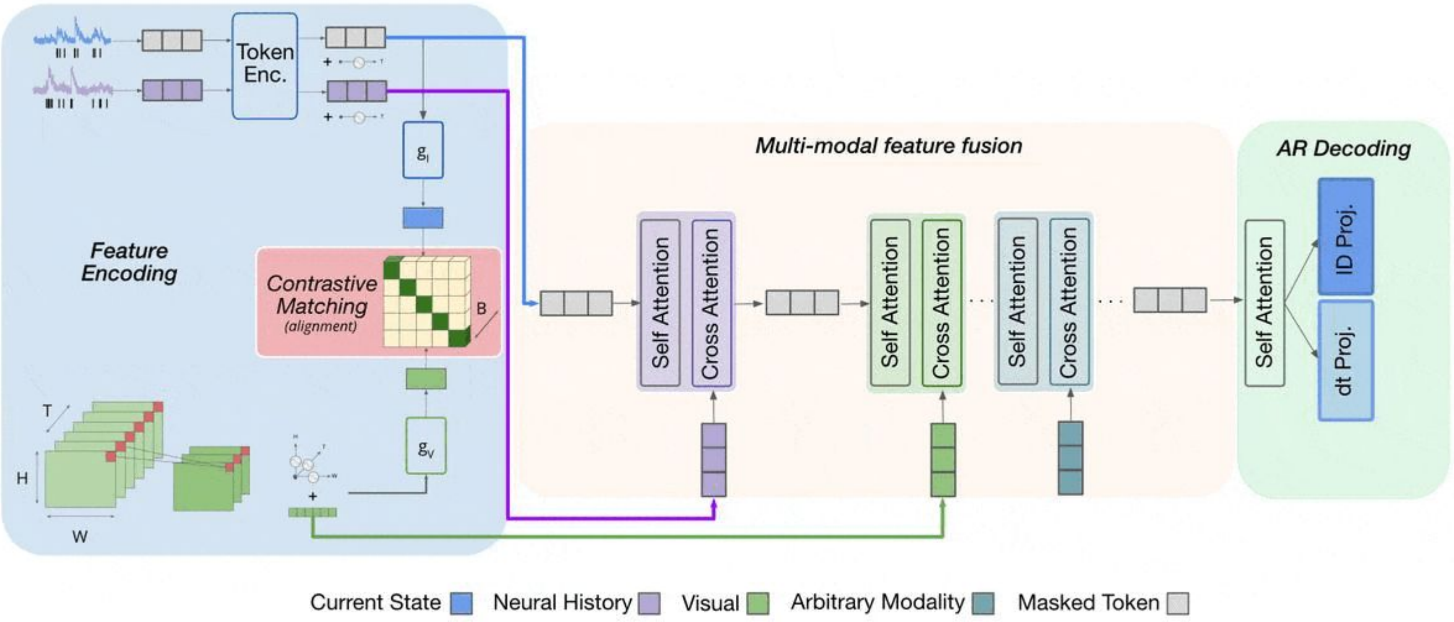
Contrastive Learning between neural/visual features

$$L_{vnc} = \frac{1}{2}\mathbb{E}_{(F,I)\in d}[H(\boldsymbol{y}^{fi}(F), \boldsymbol{p}^{fi}(F)) + \boldsymbol{y}^{if}(I), \boldsymbol{p}^{if}(I))]$$

Maximum Likelihood on id, dt:

$$L_{ce(I)} = \frac{1}{2}\mathbb{E}_{(I)\sim d} H(\boldsymbol{y}_I, \boldsymbol{p}_I) \qquad L_{ce(dt)} = \frac{1}{2}\mathbb{E}_{(dt)\sim d} H(\boldsymbol{y}_{dt}, \boldsymbol{p}_{dt})$$

Neuroformer - *inference*



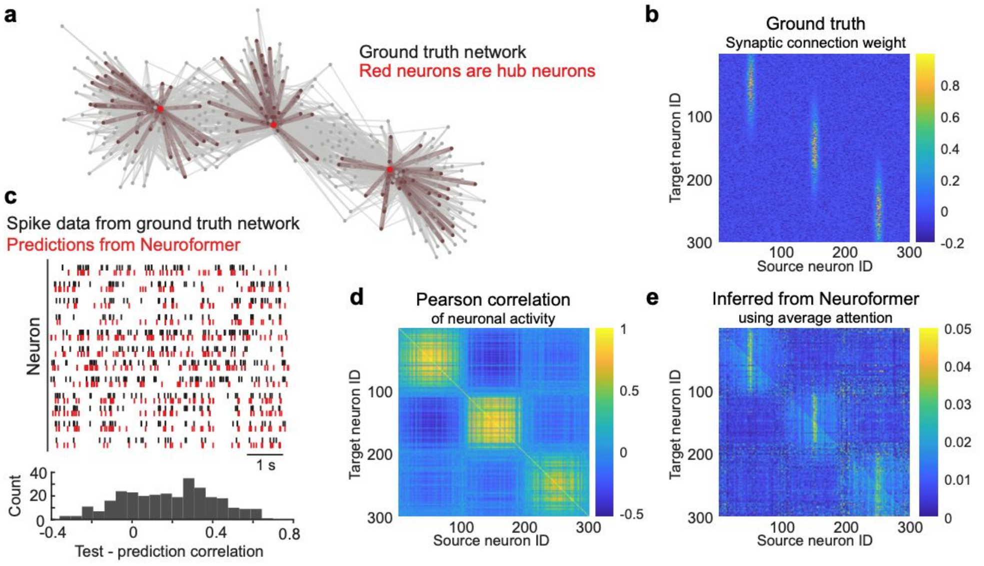
Both **spatial (ID)** and **temporal (dt)** dimensions need to be predicted

Neuroformer

(results)

Simulated Data - Uncovering *neuron connectivity*

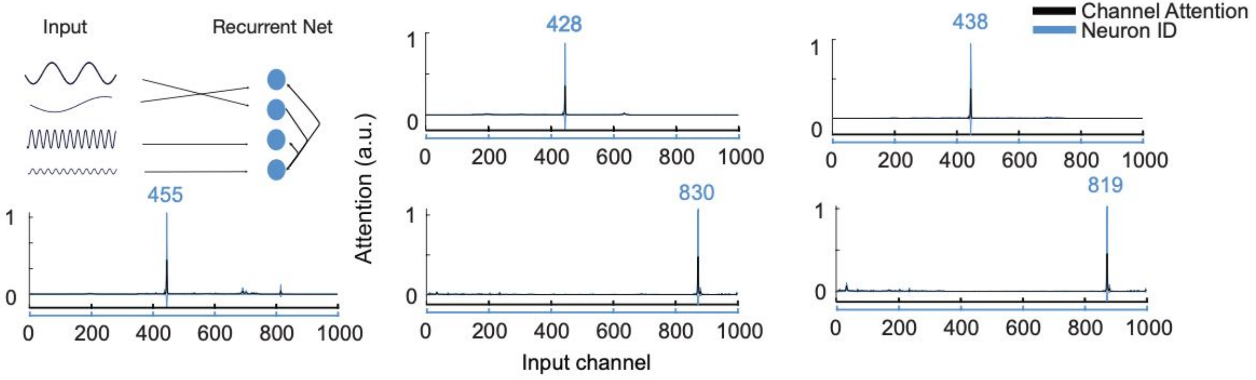
- **Simulated** dataset of **hub neurons**.
- **Self-attention** uncovers ground-truth **connectivity**.
- Explains **20% variability** (compared to 13% for Pearson correlation)



Hub Simulation experiment. Dr. Yiyi Yu

Simulated Data - *input-neuron connectivity*

- **Simulated** dataset of **1000 neurons**, each receiving **one of 1000** different **periodic inputs**.
- Cross-attention reveals the **1-to-1 input mapping**.

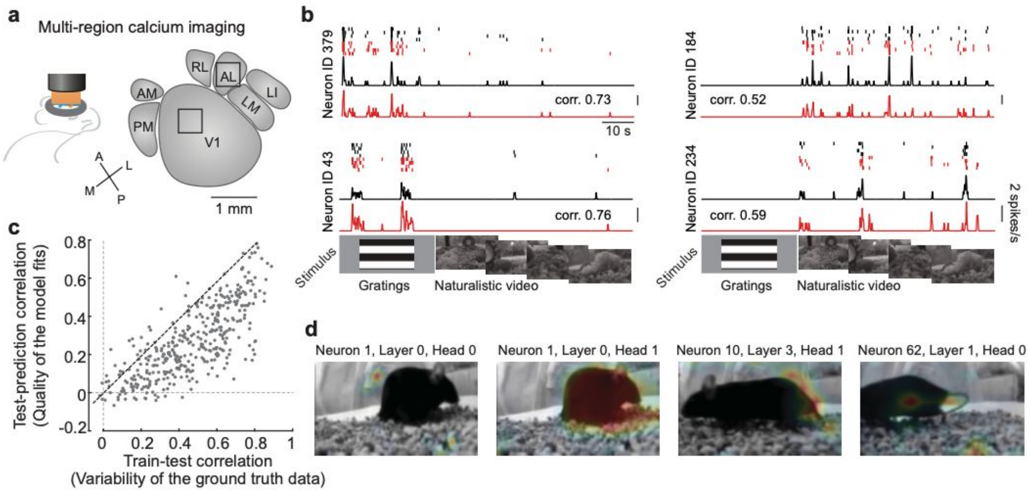


Multi-region recordings of mouse cortex

- **Mouse** watching a **naturalistic video**.
- Neuroformer can generate **high-precision simulations** of ground-truth trials ***over 32 seconds*** *auto-regressively*.
- Cross-attention between neurons and video stimulus reveal salient features.

2p calcium imaging
386 Neurons

V1 + AL



Simulations of full trials on real data **(V1 + AL)**. (Data: *Yu et al., 2022*)

Multi-region recordings of mouse cortex

- **Mouse** watching a **naturalistic video**.
- Neuroformer can generate **high-precision simulations** of ground-truth trials ***over 32 seconds*** *auto-regressively*.
- Cross-attention between neurons and video stimulus reveal salient features.

Neuron-Video Cross-Attention as computed by a trained model

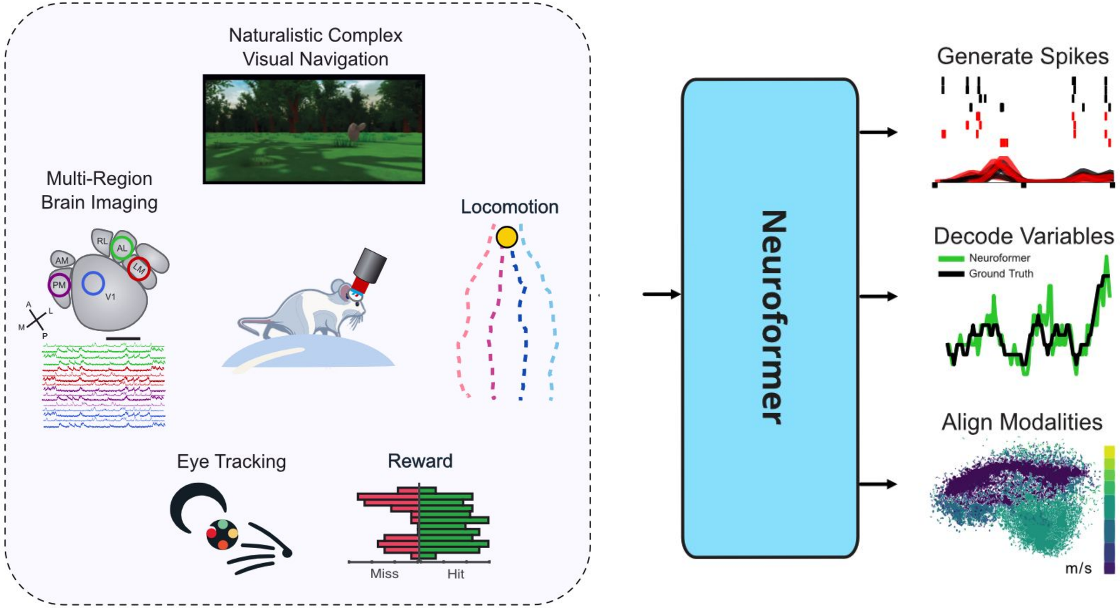
Multi-task Training and Decoding

- Neuroscientists are in possession of some of the most unique and valuable data out there.
- What can a task-driven model do with them?

2p imaging,
2000 Neurons

Lateral visual
cortex

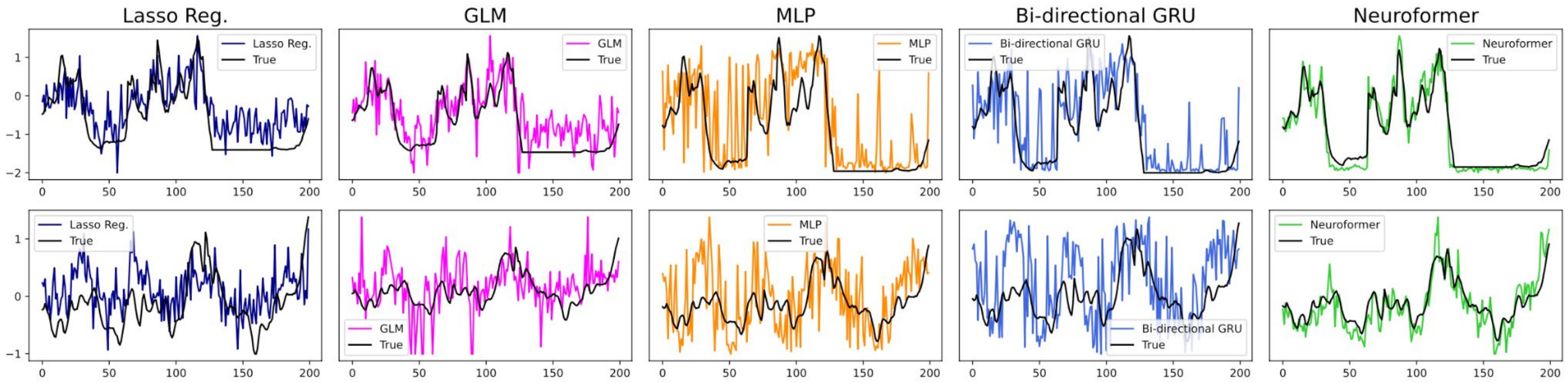
L2/L3 **Medial,**
V1, higher visual
areas



Data: Joseph Canzano, Smith LAB (UCSB)

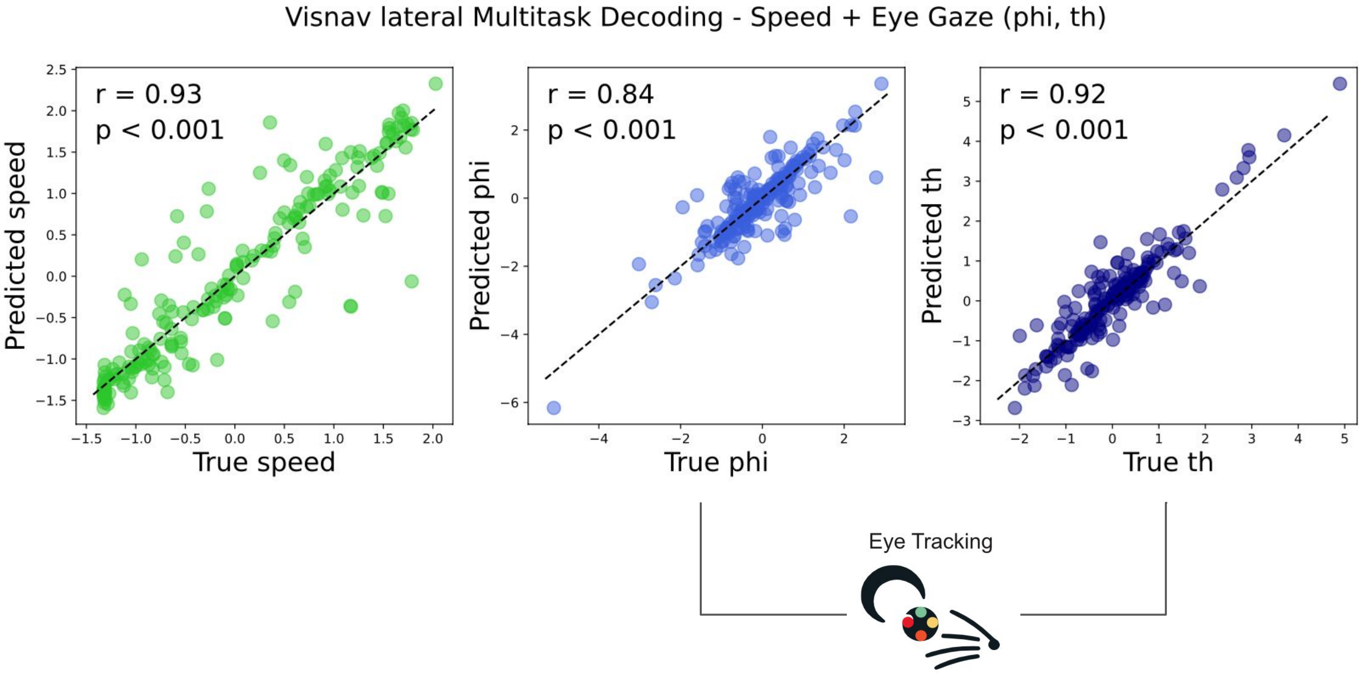
Speed Decoding

Model	Pearson Corr (r) Lateral	Medial
Lasso Regression	0.62	0.73
GLM	0.69	0.81
MLP	0.83	0.85
Bidirectional GRU	0.83	0.88
Neuroformer	0.95	0.97



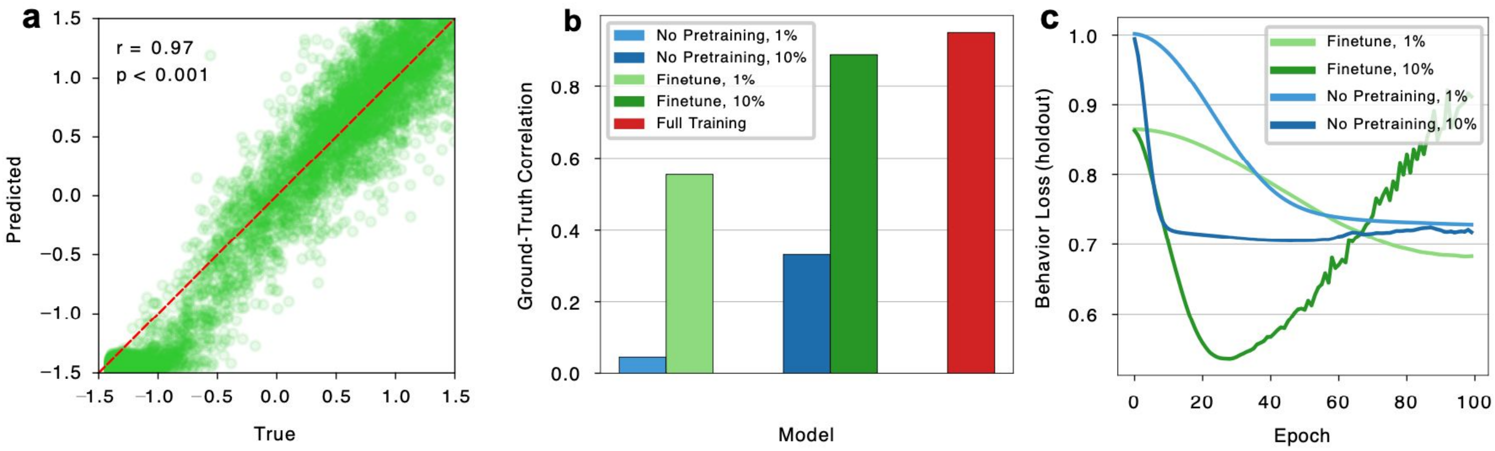
Speed *decoding, benchmarks.* ***Medial*** *(top),* ***Lateral*** *(bottom)*

Joint Training + Decoding on Many Objectives



Unsupervised Learning and Finetuning of Speed

- How much of the **speed** information is directly learnt from the spike prediction **pretraining**?



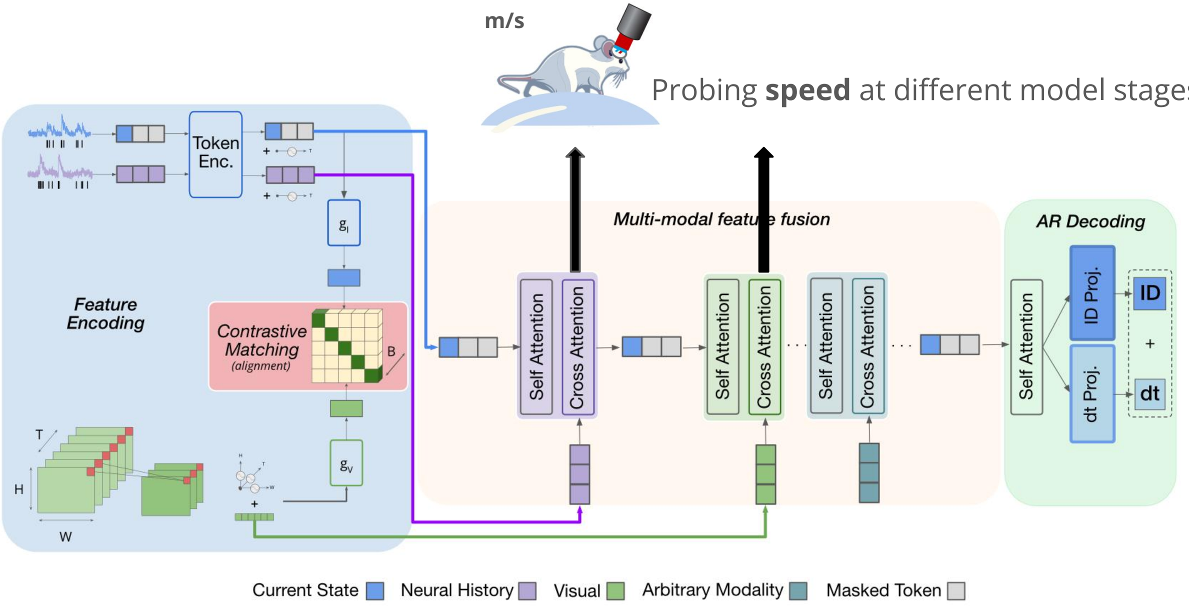
Few-shot speed accuracy for **pretrained** vs. randomly initialized models

Linear Probing Across Modalities, Tasks and Time

Feature	R^2
Random	0.0521
Past State	0.2299
Past State + Stimulus	0.5341
Past State + Stimulus + Current State	0.6357

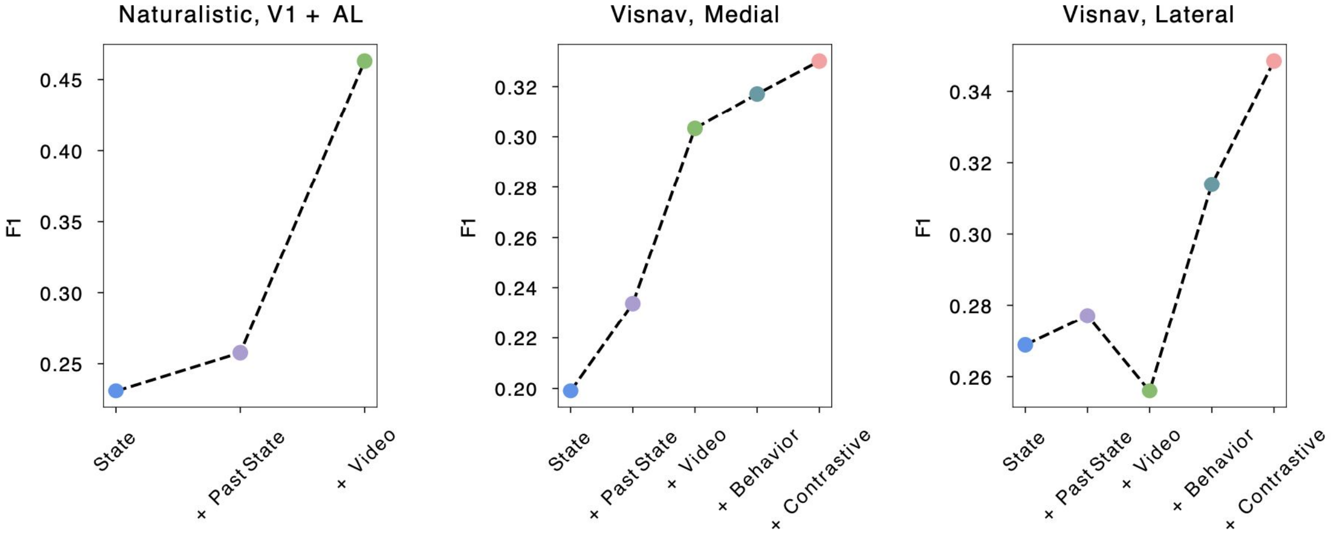
R^2 values for **linear probing** of **speed** using different representations

Probing **speed** at different model stages



Effect of Model Components

- Each subsequent **component** improves response **prediction** performance.



Response Prediction **F1** Score with each added **component**

Code

- New modalities and tasks can be easily incorporated: www.github.com/a-antoniades/Neuroformer

```
{
    'data': {
        'spikes': (N_neurons, N_timesteps),  # np.ndarray, required key
        'frames': (N_frames, N_timesteps),   # np.ndarray, optional key
        'behavior variables': (N_timepoints,),  # np.ndarray,
    'intervals': (N_timepoints,),  # np.ndarray, Denoting all intervals/time bins of th
    'train_intervals': (N_timepoints,) ,  # np.ndarray, The corresponding train interva
    'test_intervals': (N_timepoints,) ,   # np.ndarray, The corresponding test interval
    'finetune_intervals': (N_timepoints,) ,  # np.ndarray, The corresponding finetune i
    'callback': callback() # function
}
```

Neuroformer **data config** file

```
Modalities: Any additional modalities other than spikes and frames.

Modality Type: The name of the modality type. (for example behavior)

Variables: The name of the modality.

  Data: The data of the modality in shape (n_samples, n_features).
  dt: The time resolution of the modality, used to index n_samples.
  Predict: Whether to predict this modality or not. If you set predict to false, then i
  Objective: Choose between regression or classification. If classification is chosen,
```

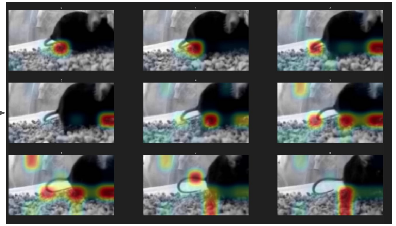
Neuroformer **model config** file

Future Directions

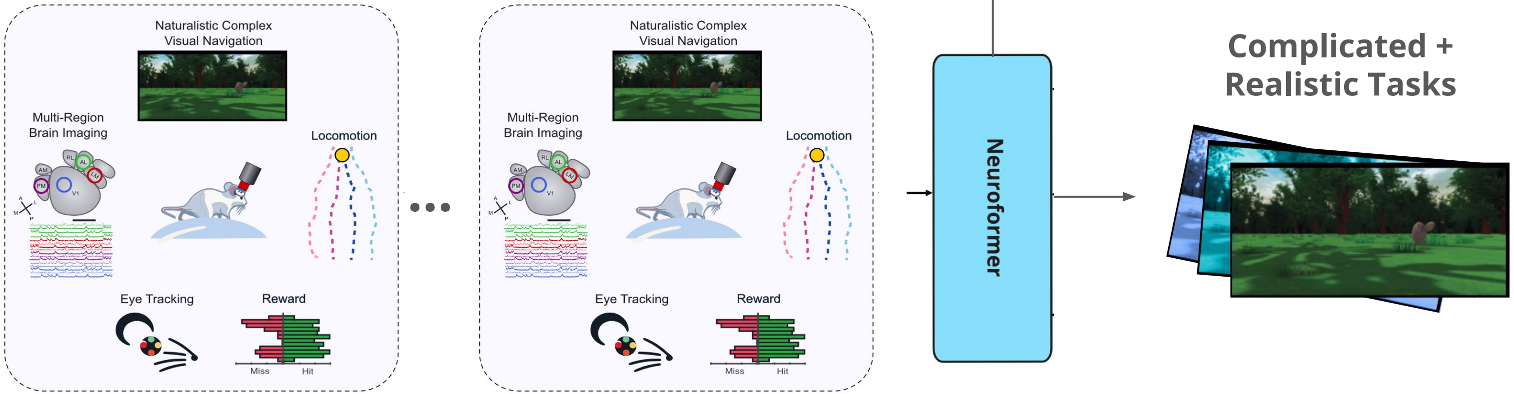
AI » Neuroscience

- Neuroformer proposes a new approach to Neuroscience analysis.
    - Large scale data
    - Diverse data (**modalities / tasks)**
    - Realistic Experimental Setups
    - Black-box analysis

Deeper Representational Analysis



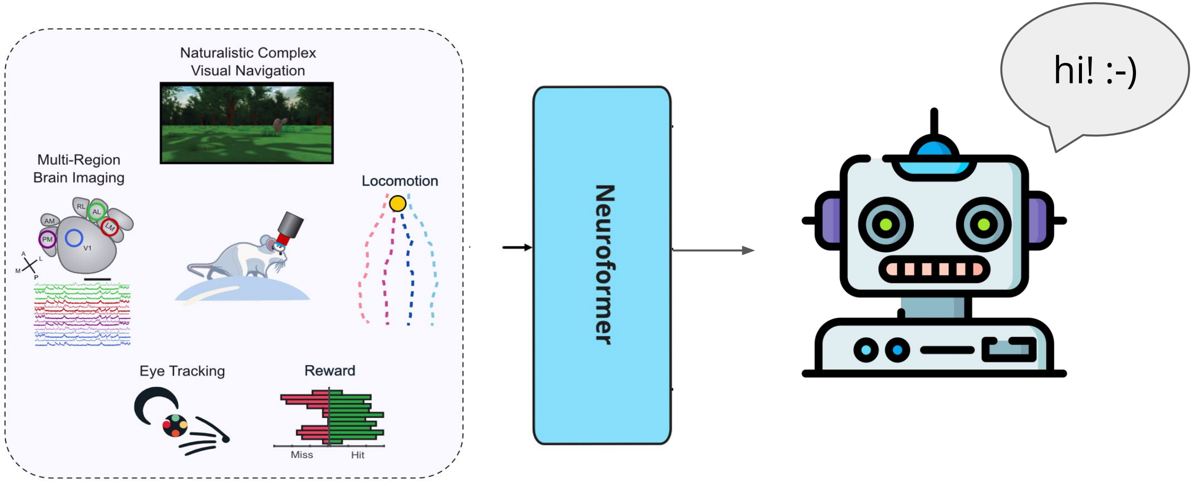
Complicated + Realistic Tasks



More experiments + Data + Compute

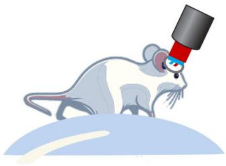
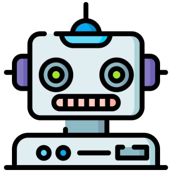
Neuroscience » AI

- Neuroscientists are in possession of some of the most unique and valuable data out there.
- What can a task-driven model do with them?



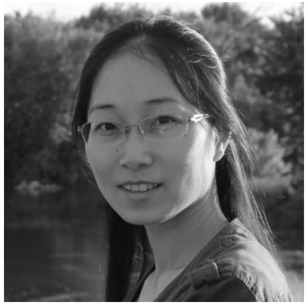
Conclusion AI » Neuroscience » AI » Neuroscience ...

- The bitter lesson taught us that in order to learn complex relationships, we need a vast amount of data.
- I think this applies to both **machine learning models**, but also complex **neuroscience theories**.
- By integrating the power of valuable **neuroscience datasets**, we can make many **new scientific discoveries**.
- And perhaps even models with **new/better abilities**...

With the capabilities we have today to harness the power of large-scale data, the field of Neuroscience + AI is **wide open** (imo).

Thanks!



Dr. Yiyi Yu

Joseph Canzano

Dr. Che-Hang Yu

Dr. Spencer LaVere Smith

Dr. William Wang

@anton_iades www.github.com/a-antoniades/Neuroformer