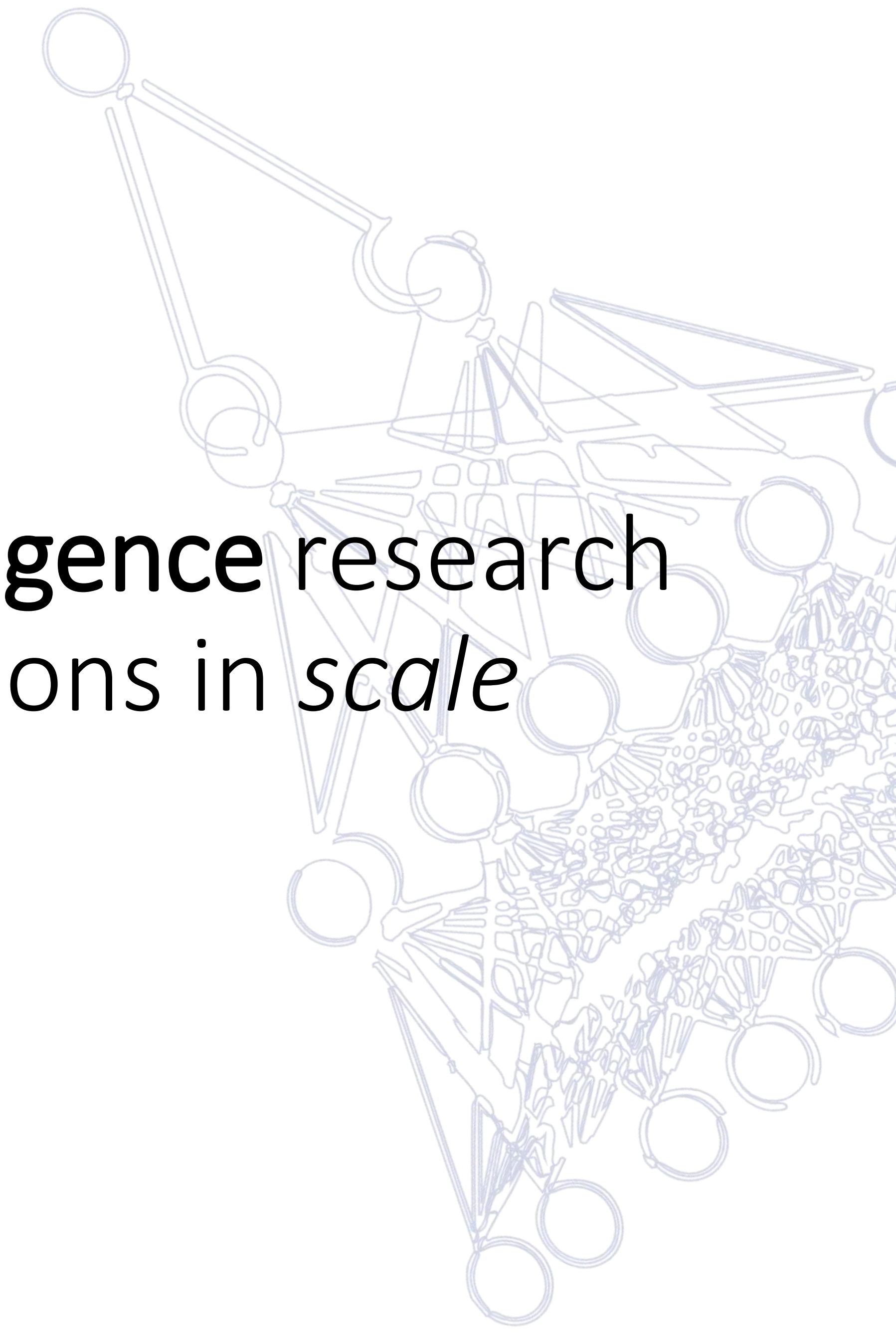
NeuroAI Lab

Martin Schrimpf

martin.schrimpf@epfl.ch

EPFL

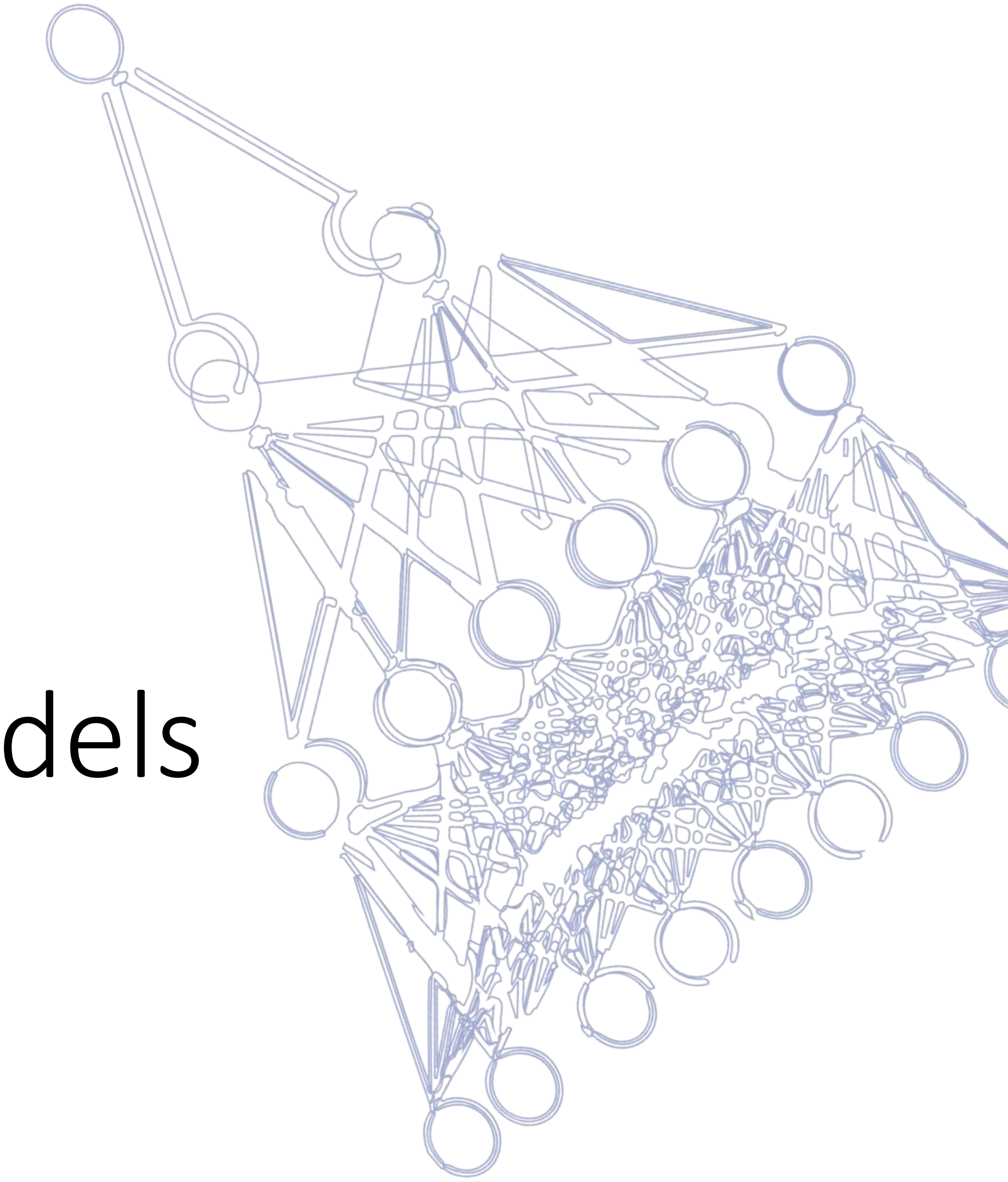
Neuro X Institute



Neuroscience and **artificial intelligence** research are both undergoing revolutions in *scale*

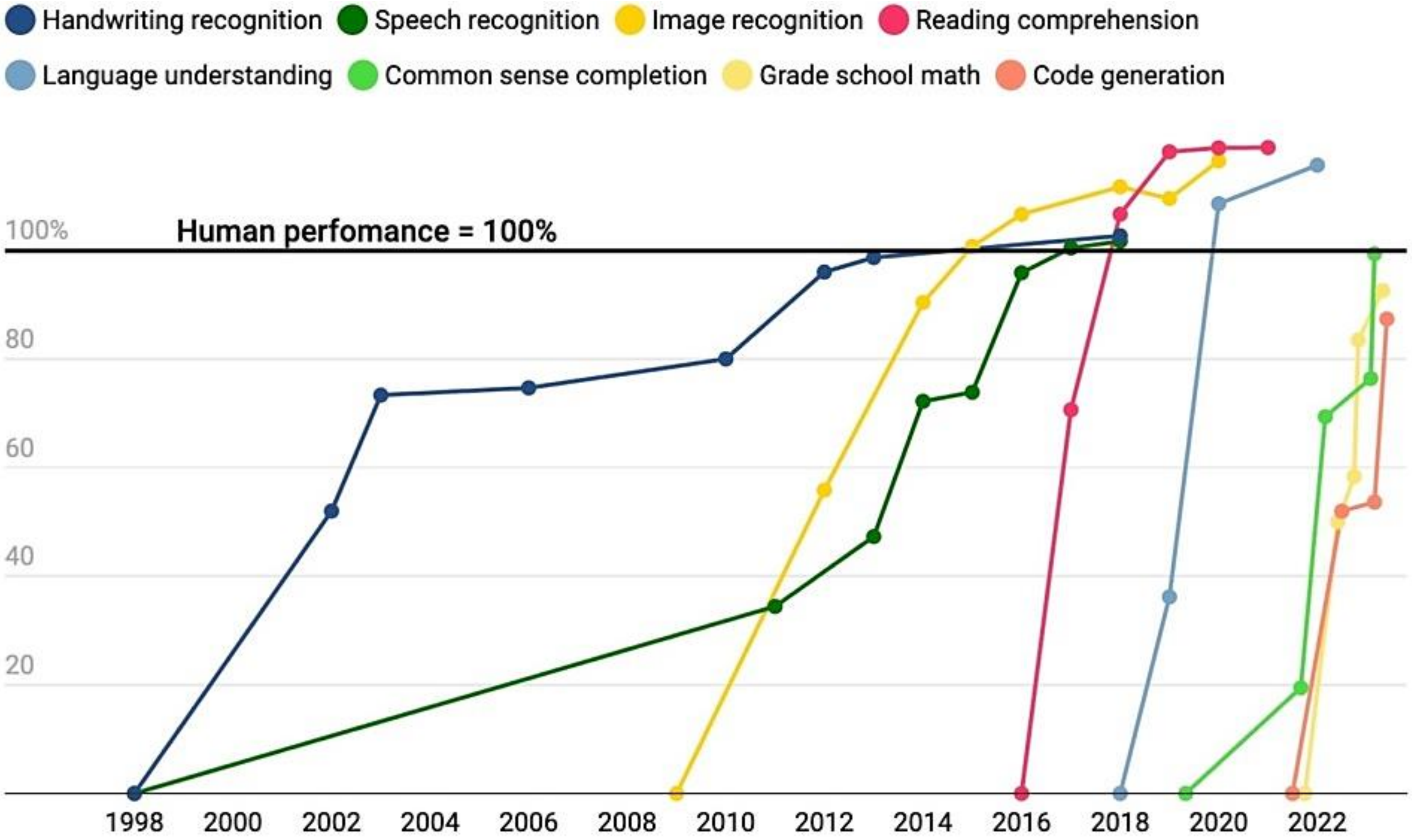
AI Revolution:

High-Performance Machine Learning Models



Neural network models achieve human performance in a range of tasks (not all)

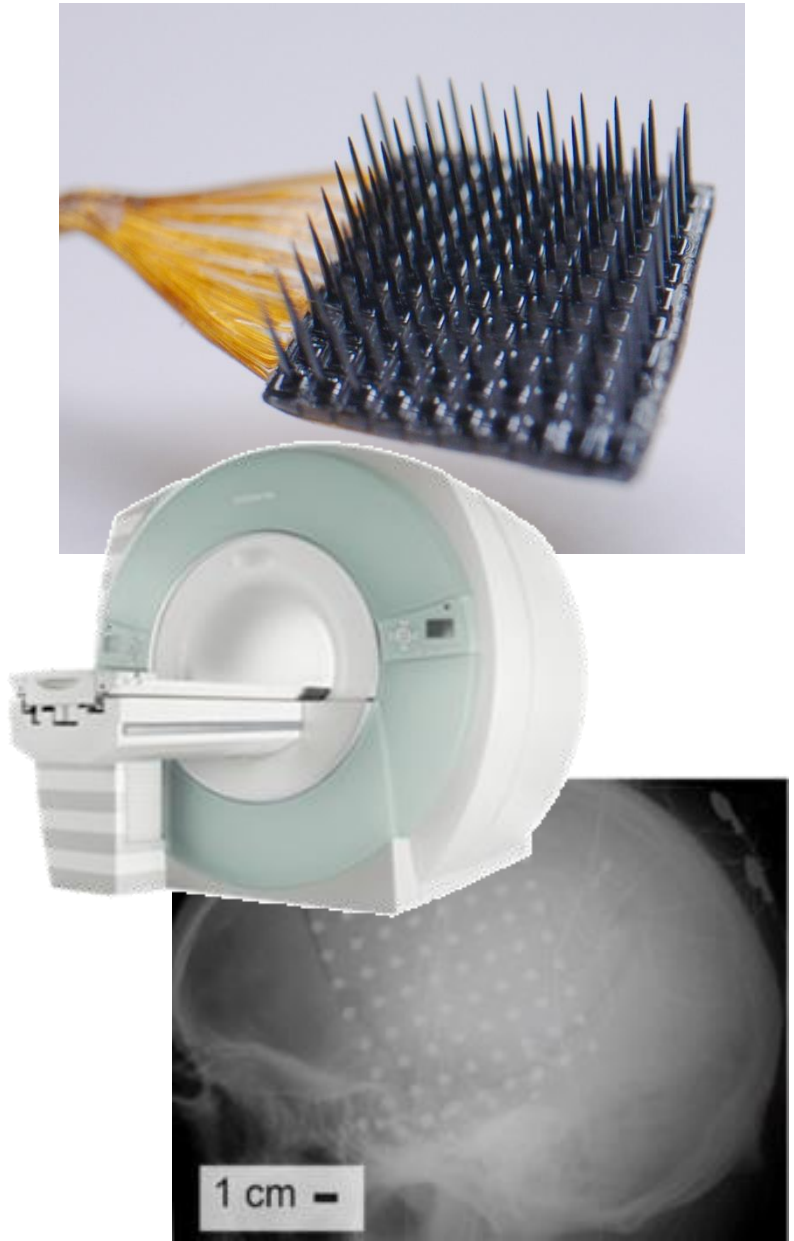
Model performance relative to humans

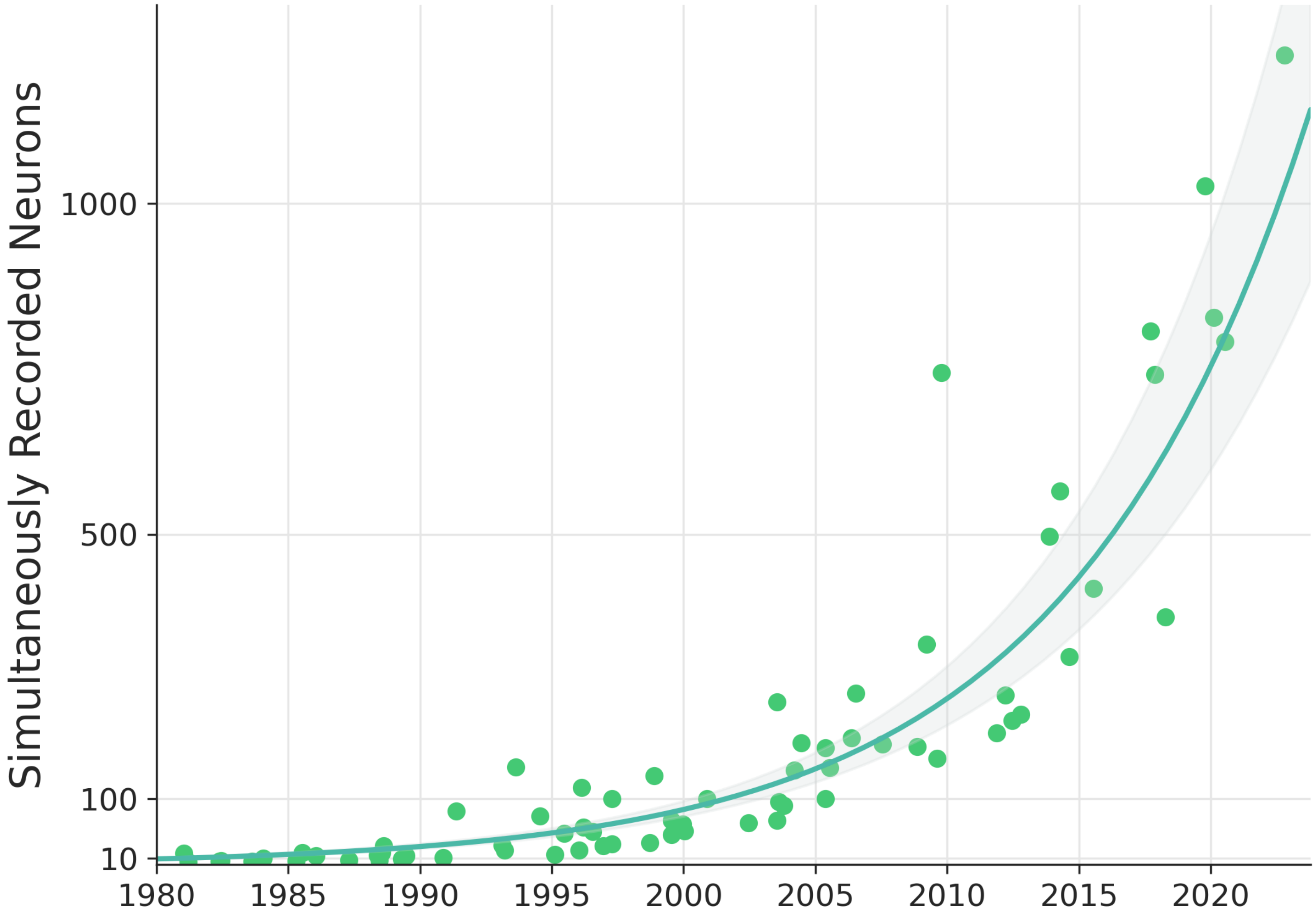


Kiela et al. 2023

Neuroscience Revolution: Increasing Availability of Large-Scale Brain Data

Access to neural data is increasing exponentially



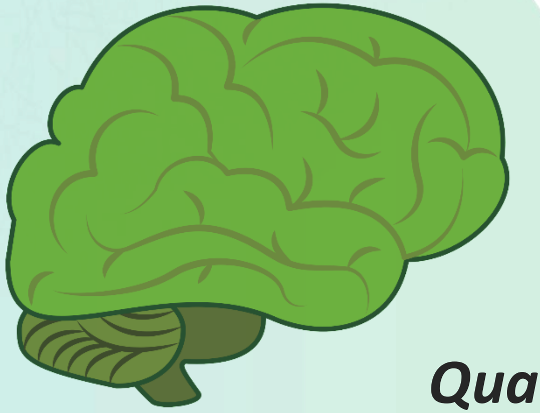


https://stevenson.lab.uconn.edu/scaling

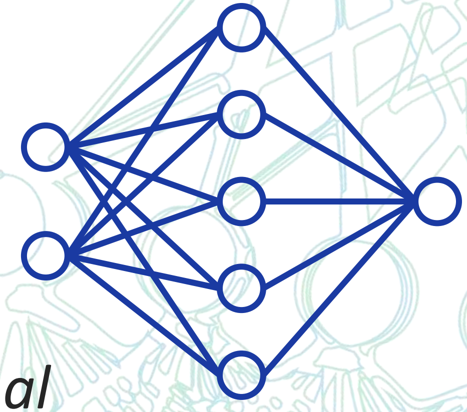
Synergy between science + engineering

Brain and Cognitive Sciences

AI/ML Engineering



System Models of **Natural Intelligence**

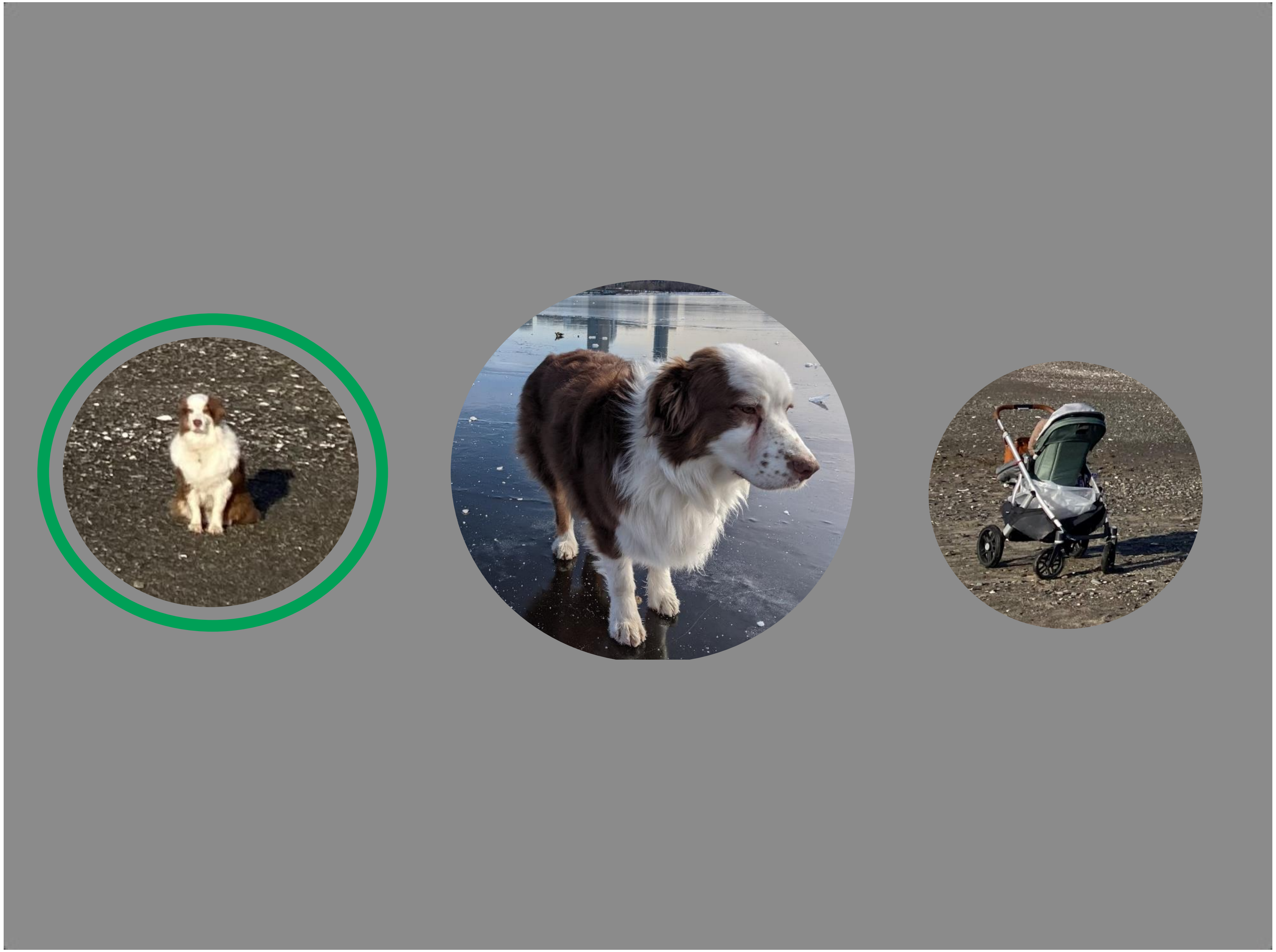


Quantitative *Measurements & Discoveries*

Computational ***Hypotheses*** *& Fine-grain* ***Predictions***

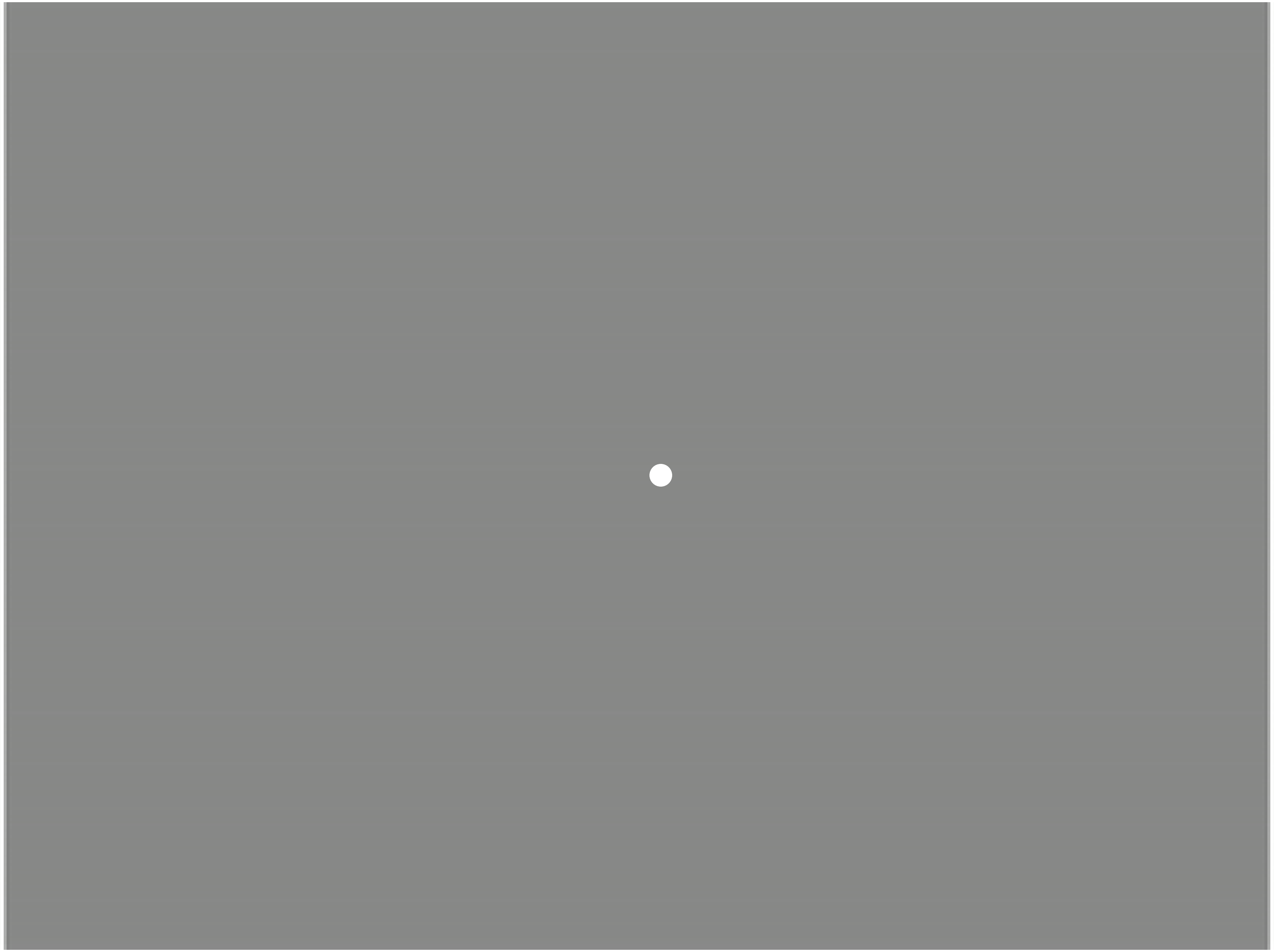
Key activity: build **models** that are **aligned to behavioral and neural data**

Behavioral experiment



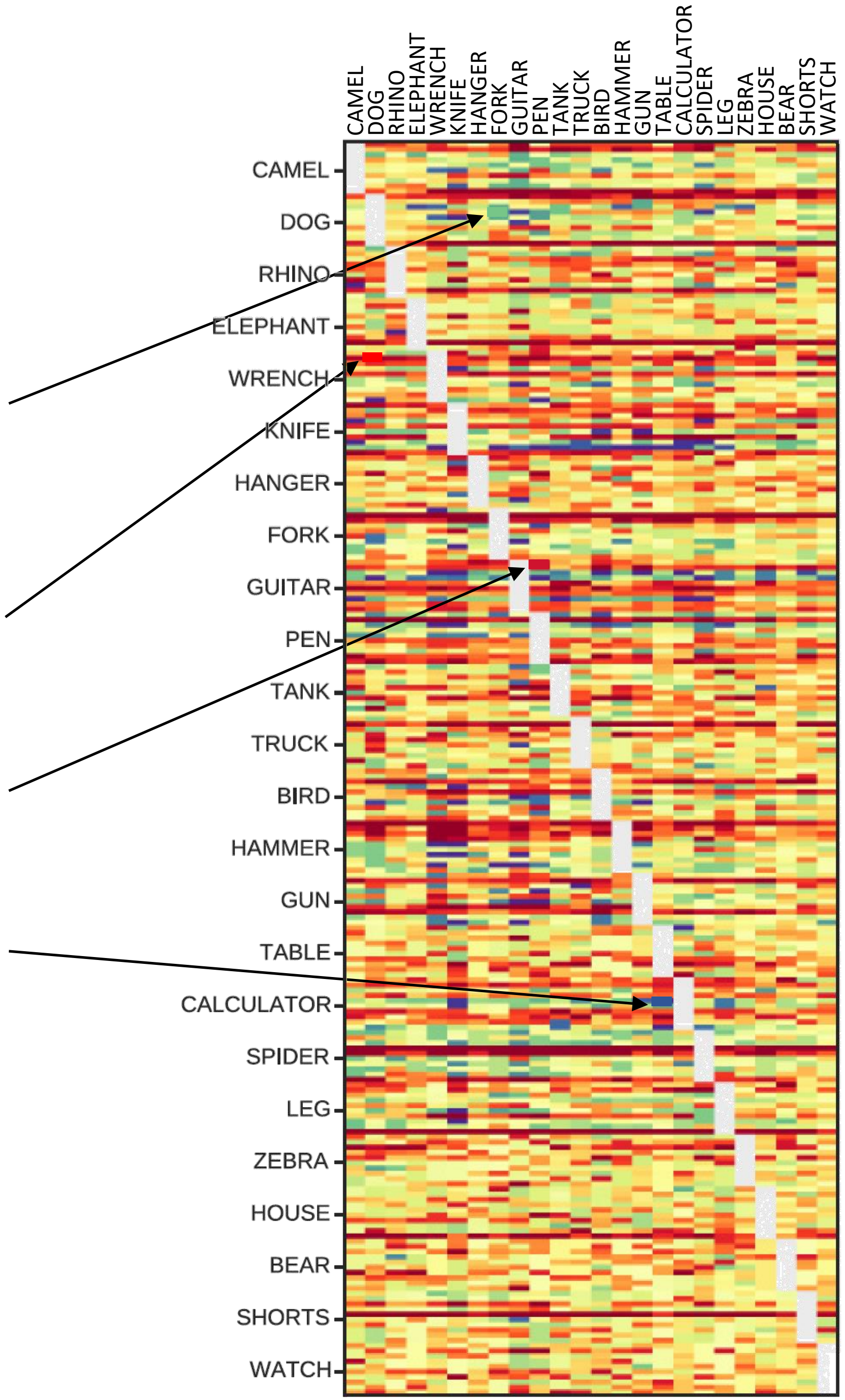
Rajalingham, Issa*, et al. (JNeuro 2018)*

Behavioral experiment

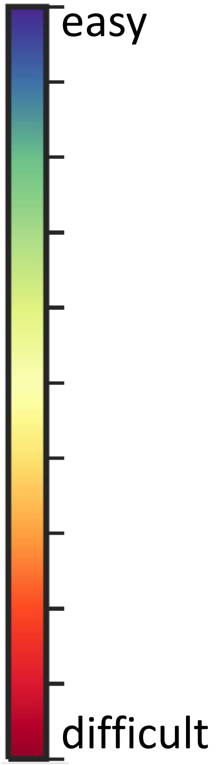


Rajalingham, Issa*, et al. (JNeuro 2018)*

Behavioral data

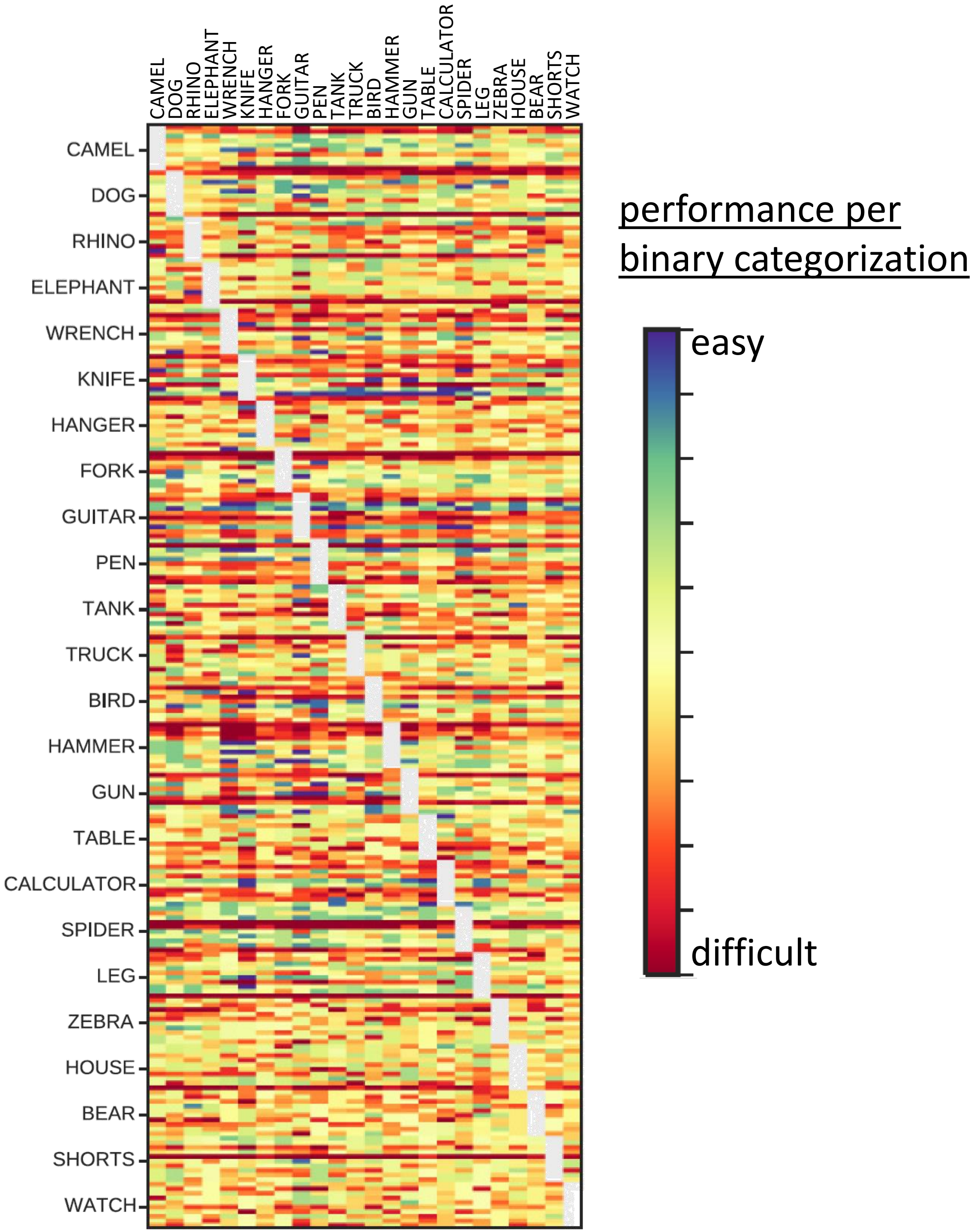


performance per binary categorization



Rajalingham, Issa*, et al. (JNeuro 2018)*

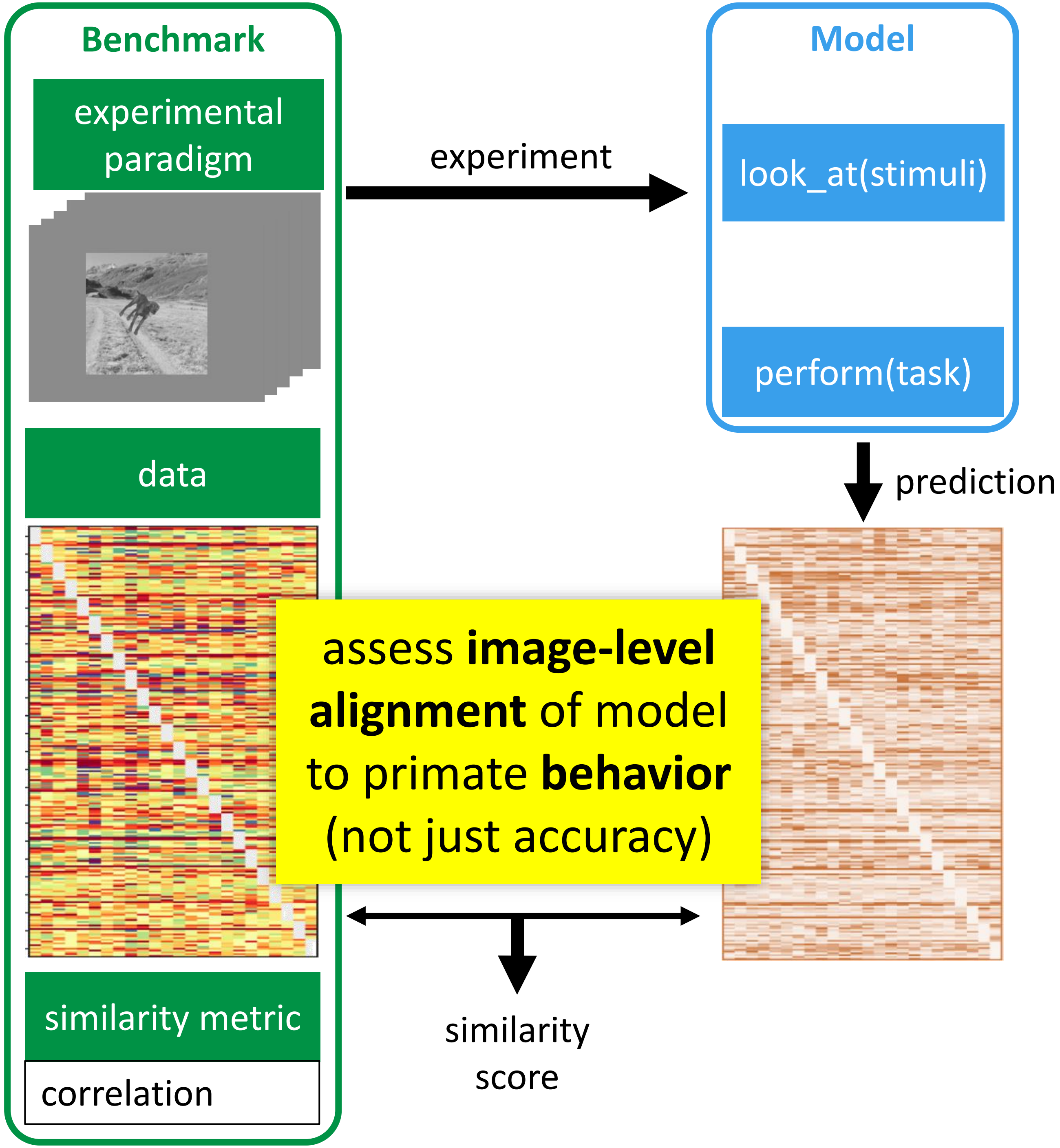
Behavioral data



performance per binary categorization

Rajalingham, Issa*, et al. (JNeuro 2018)*

Behavioral benchmark



Benchmark

experimental paradigm

data

similarity metric

correlation

experiment

Model

look_at(stimuli)

perform(task)

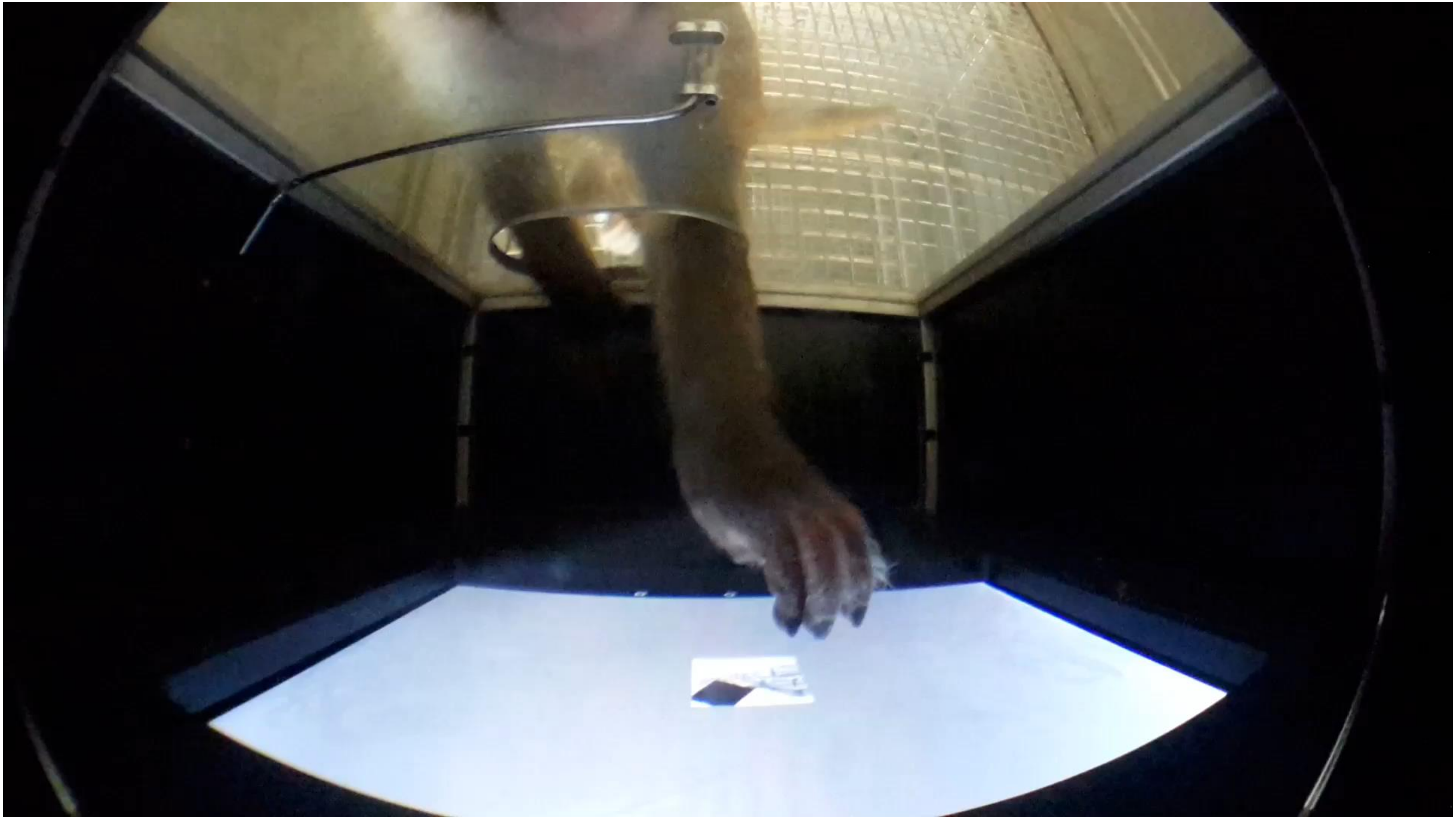
prediction

similarity score

assess **image-level alignment** of model to primate **behavior** (not just accuracy)

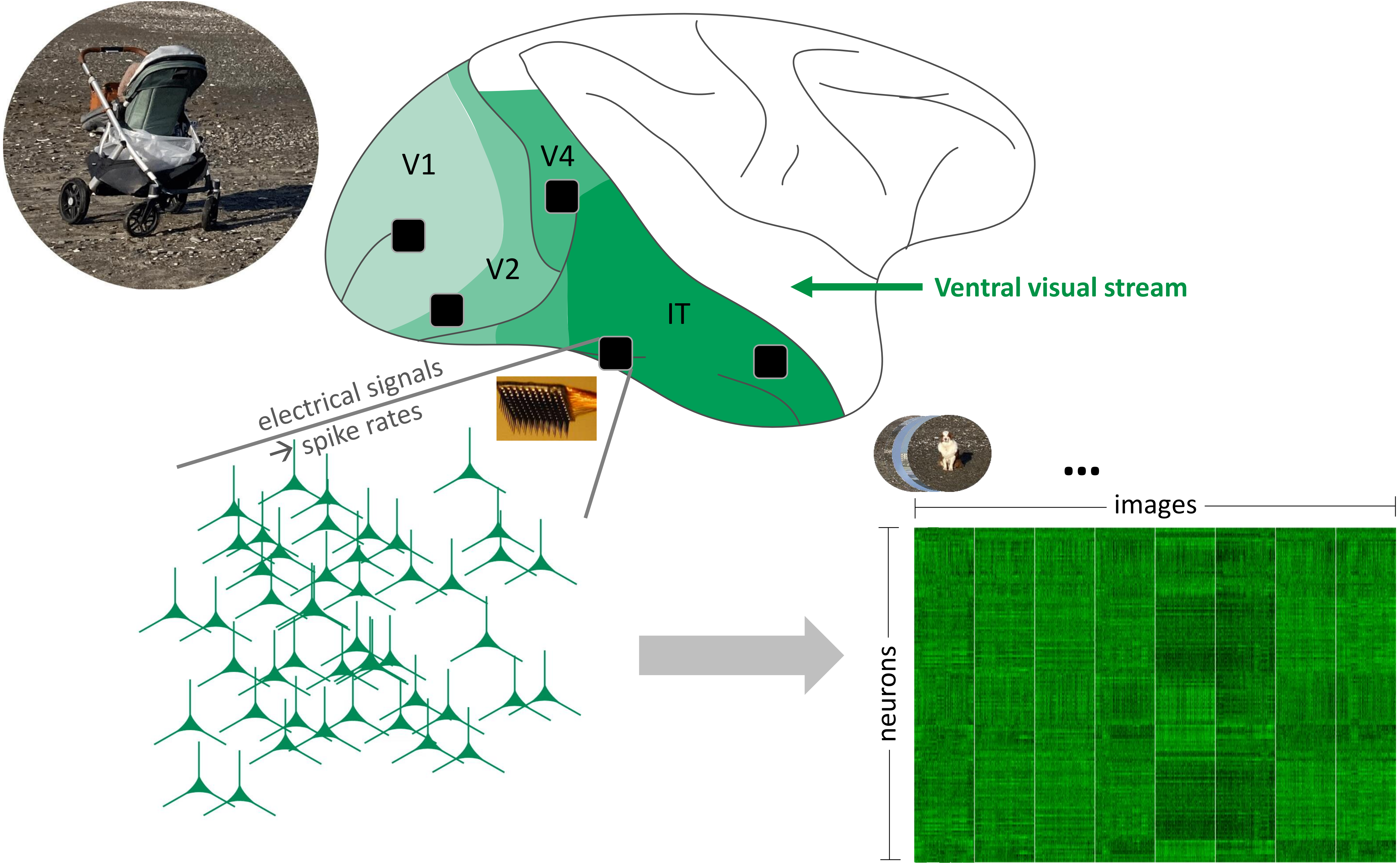
Neural data

Neural data

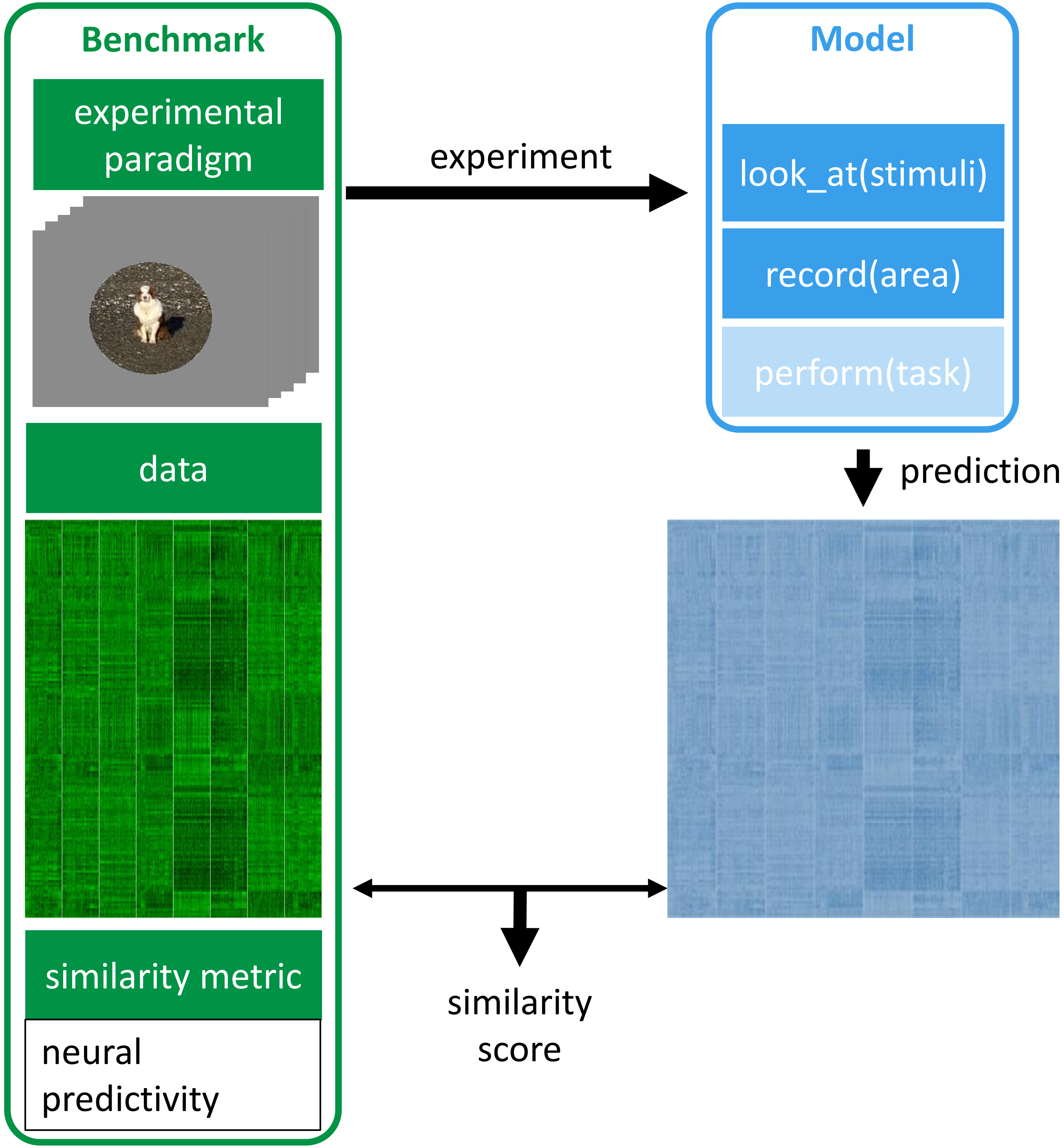


video courtesy of Kailyn Schmidt

Neural data


V1
V4
V2
IT
Ventral visual stream
electrical signals
→ spike rates
...
images
neurons

Neural benchmark



Benchmark

experimental paradigm

data

similarity metric

neural predictivity

experiment

Model

look_at(stimuli)

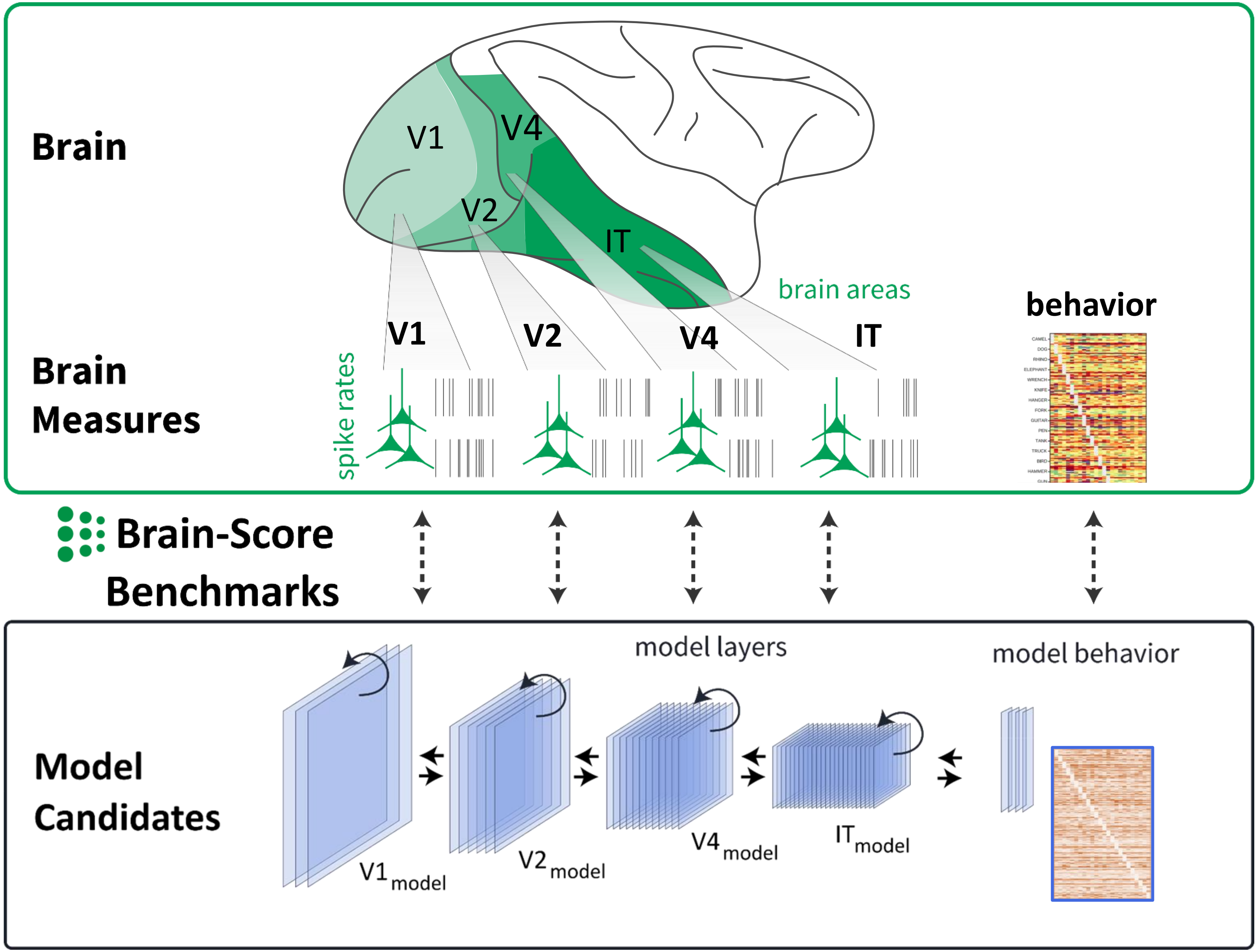
record(area)

perform(task)

prediction

similarity score

Particular models are aligned to vision in the brain

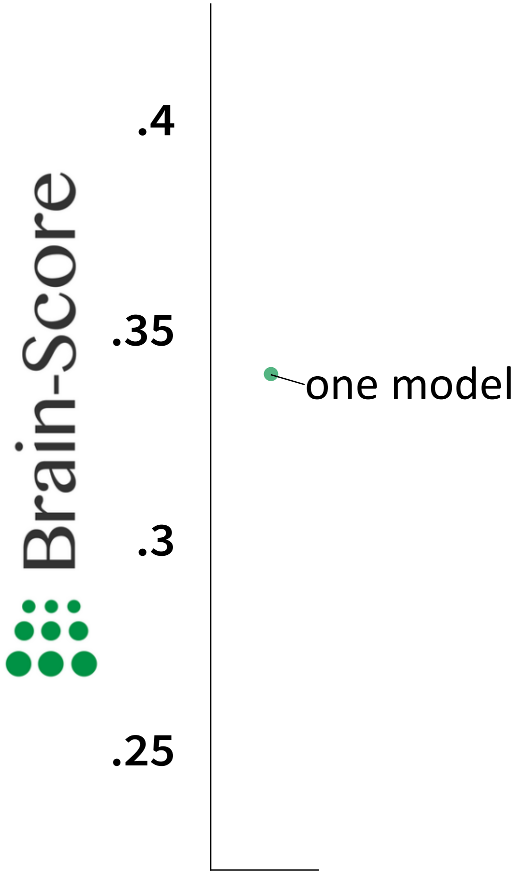


Artificial neural network models

- Trained for computational task, weights optimized via backprop

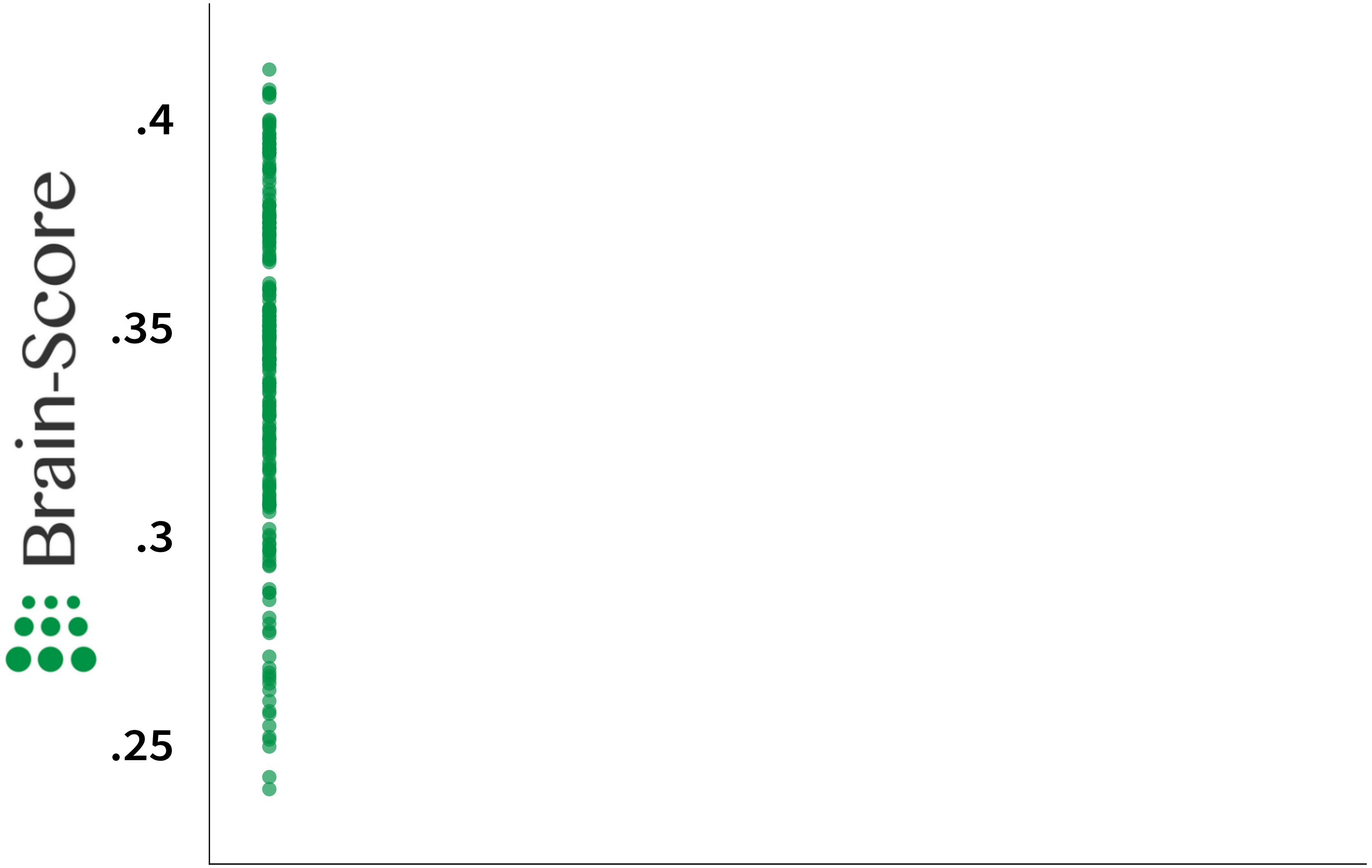
- Internal processing stages (hidden layers, “deep” learning)
- Accept any new input (pixels)

Particular models are aligned to vision in the brain



Schrimpf, Kubilius*, et al. 2018*

What explains the model differences?



Normative variable

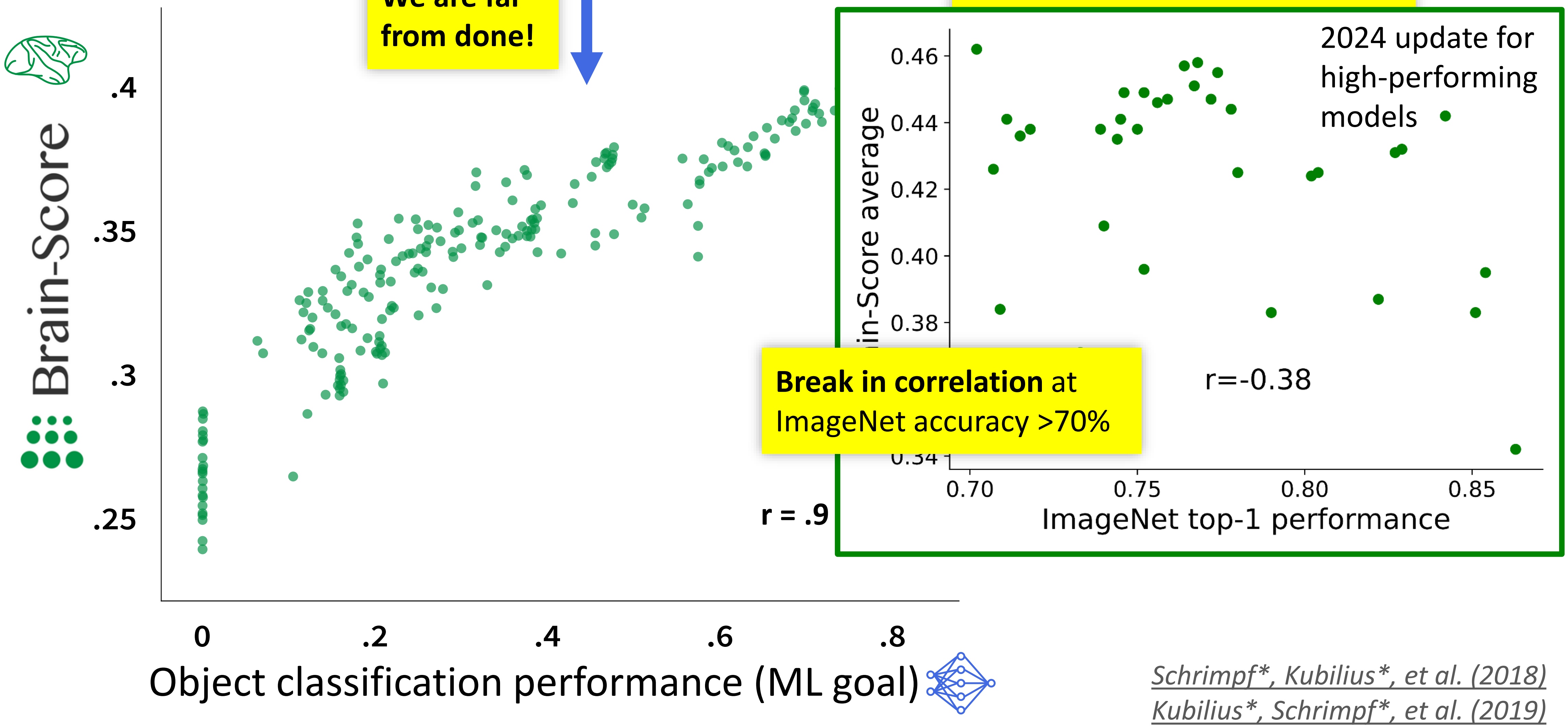
cf. Yamins, Hong*, et al. 2014*

Schrimpf, Kubilius*, et al. 2018*

What explains the model differences?

Not the small-scale circuits

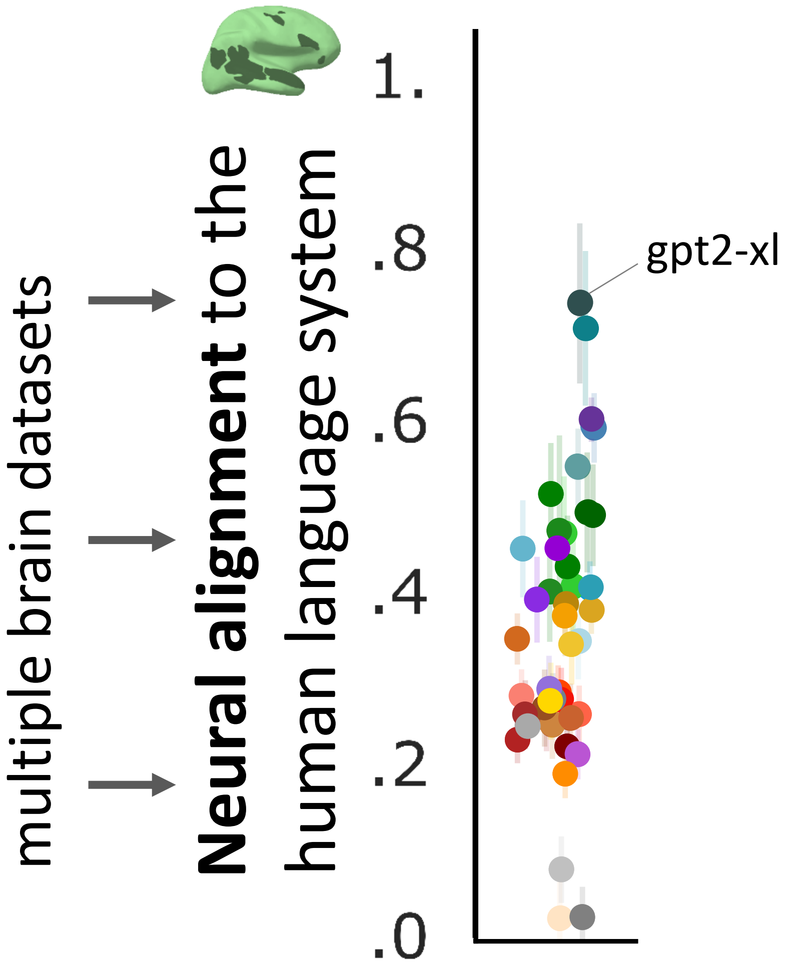
We are far from done!



Break in correlation at ImageNet accuracy >70%

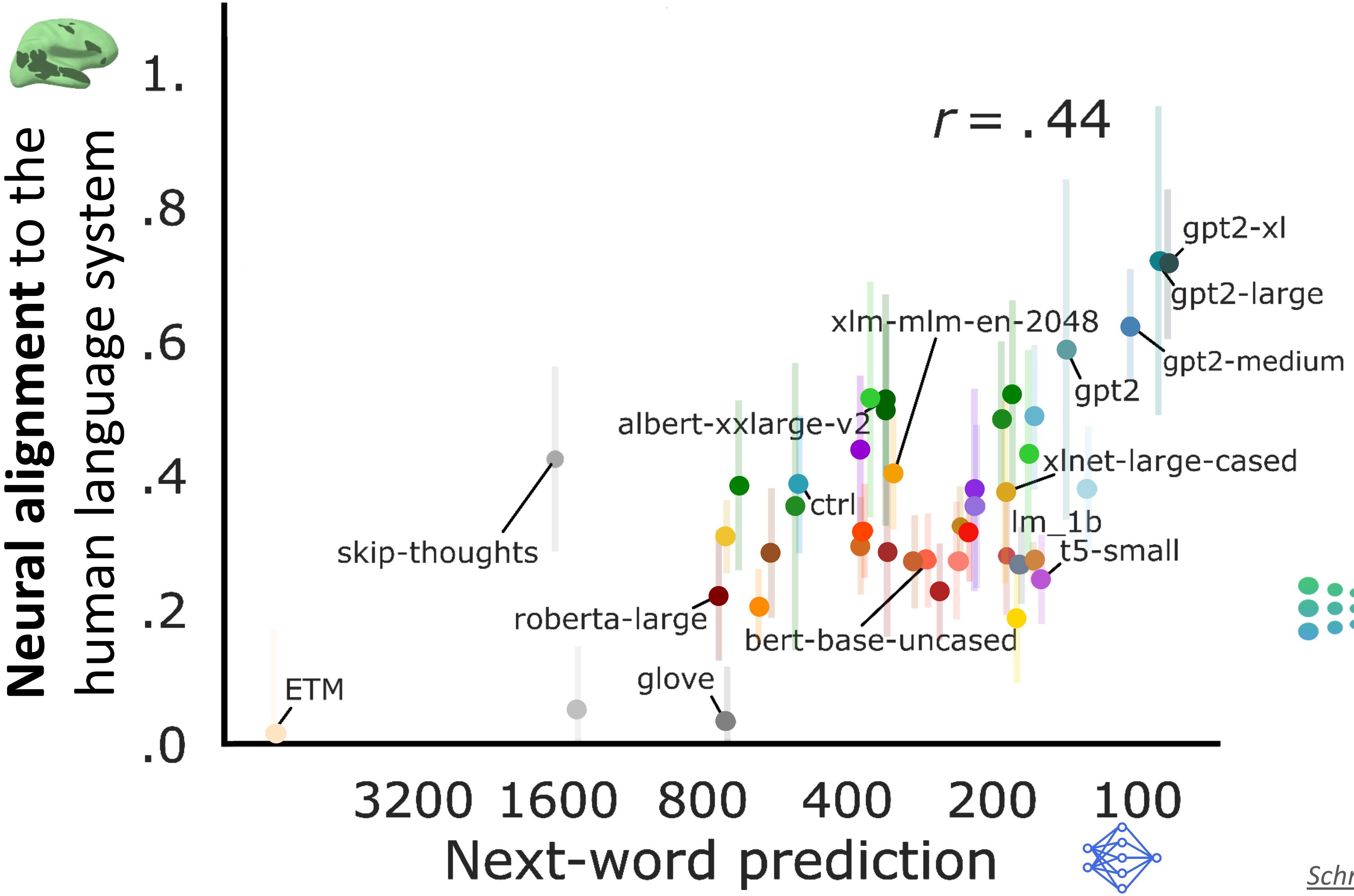
Schrimpf, Kubilius*, et al. (2018)*
Kubilius, Schrimpf*, et al. (2019)*

Particular ML language models predict the human language system



Schrimpf et al. (PNAS 2021)

The better models can predict the next word, the more brain-like they are

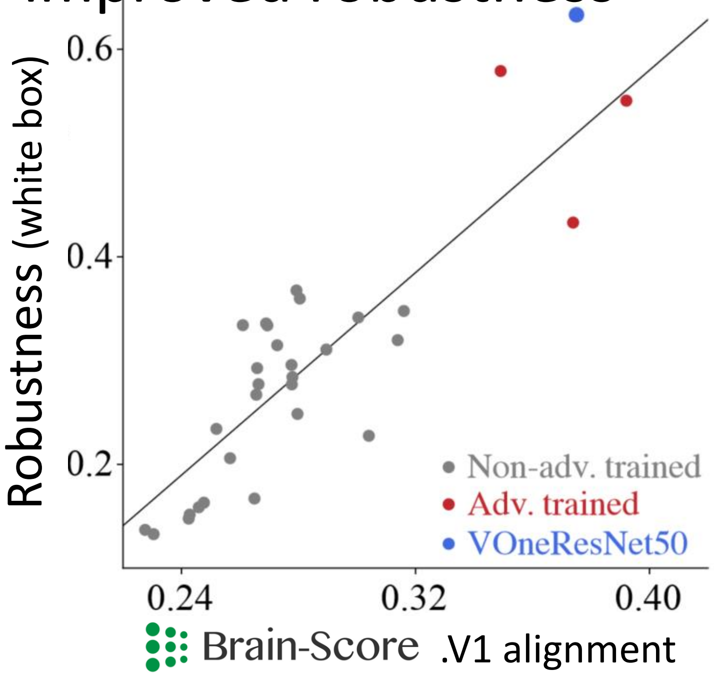


Brain-Score .org/language

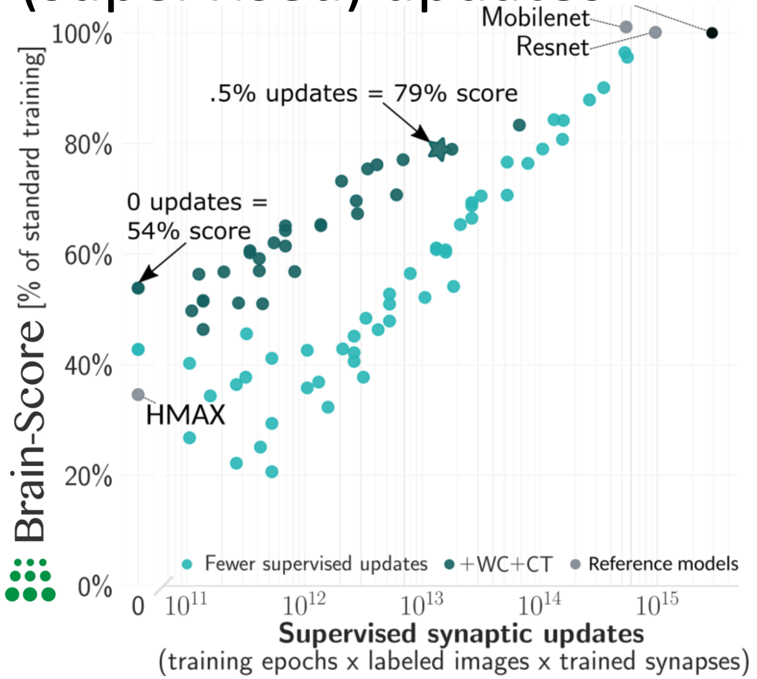
Schrimpf et al. (PNAS 2021)

Today’s models are not perfect! But we can make them better

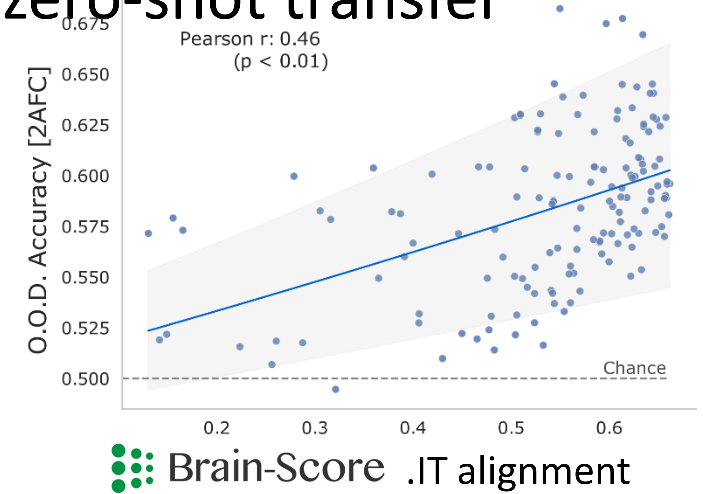
VOneNet: more brain-like → improved robustness



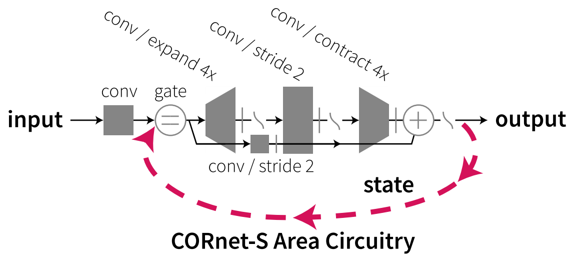
Wiring Up Vision: reduce (supervised) updates



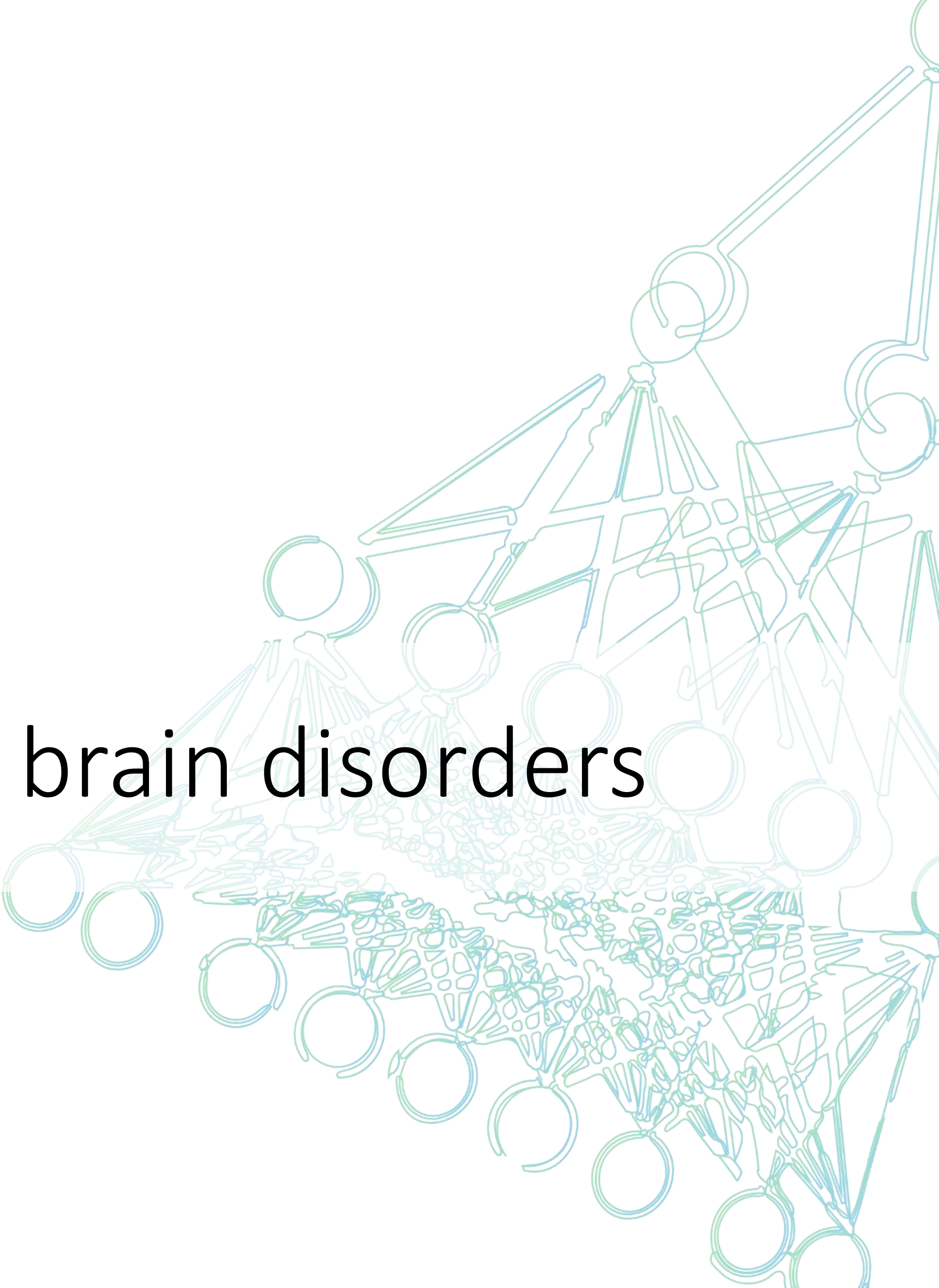
Generalization: more IT-like → zero-shot transfer



CORnet: shallow recurrent neuroanatomy → predict temporal dynamics



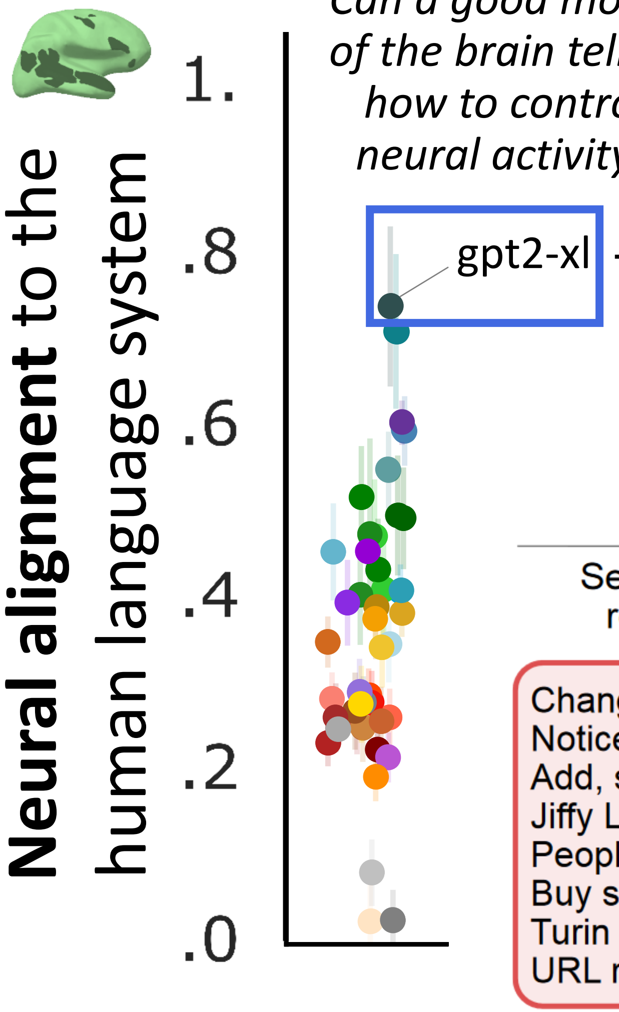
Dapello, Marques*, et al. (NeurIPS 2020 Spotlight)*
Dapello, Kar*, et al. (ICLR 2023 Notable Top-5%)*
Geiger, Schrimpf*, et al. (ICLR 2022 Spotlight)*
Zhuang et al. (PNAS 2021)
I Gusti Bagus et al.(SVRHM 2023)
Kubilius, Schrimpf*, et al. (NeurIPS 2019 Oral)*
…

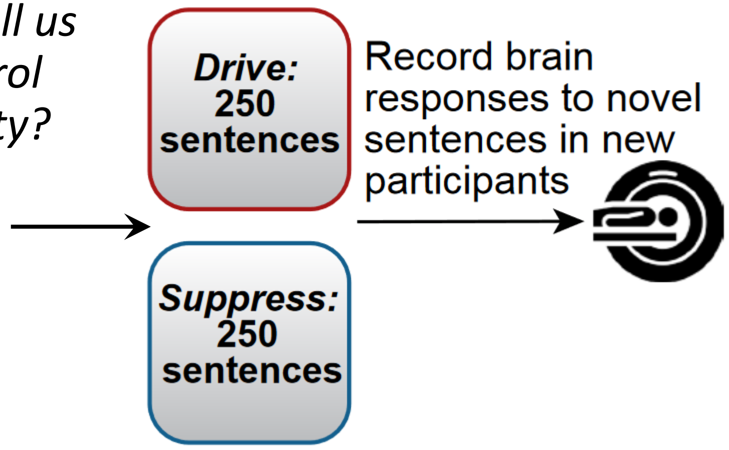


digital twins to help treat brain disorders

We can use brain-aligned LLMs to noninvasively control neural activity

Can a good model of the brain tell us how to control neural activity?





Record brain responses to novel sentences in new participants

Drive

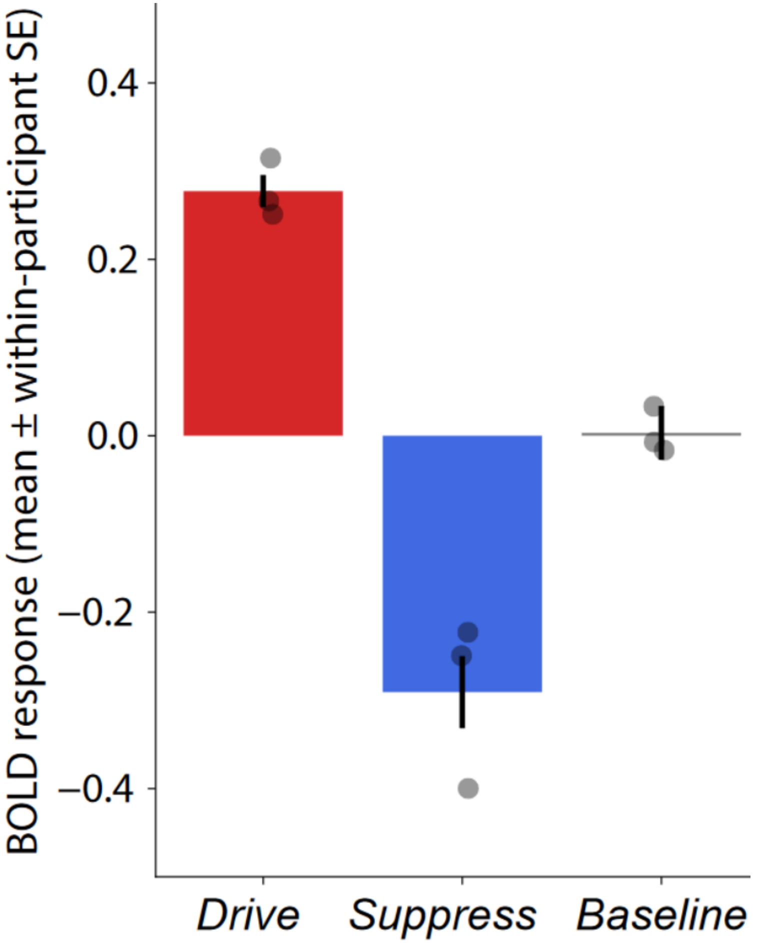
Sentences identified t
response in the lang

Changing PhD group: Yes
Notice how you reacted to
Add, some feminists are c
Jiffy Lube of -- of therapie
People on Insta Be Like, "
Buy sell signals remains a
Turin loves me not, nor wi
URL right, or report review

Suppress

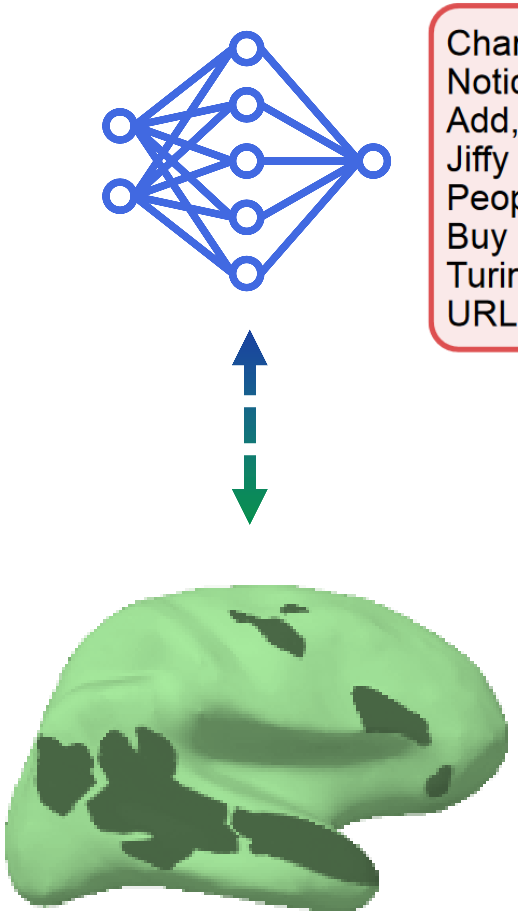
Sentences identified to elicit minimal response in the language network

We were sitting on the couch.
That is such a beautiful picture!
They stood there for a moment.
They went up the stairs together.
Inside was a tiny silver sculpture.
They walked out onto the balcony.
Cas gazed up at the sky.
What else is there to do?



Tuckute et al. (Nature Human Behavior 2023)
See also Bashivan*, Kar*, et al. (Science 2019)

NeuroAI models can control brain activity



Drive

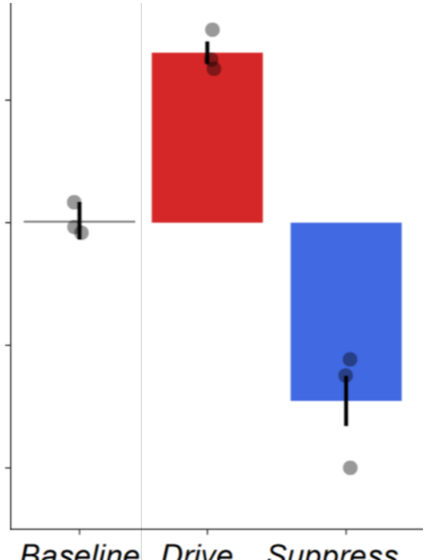
Sent
resp

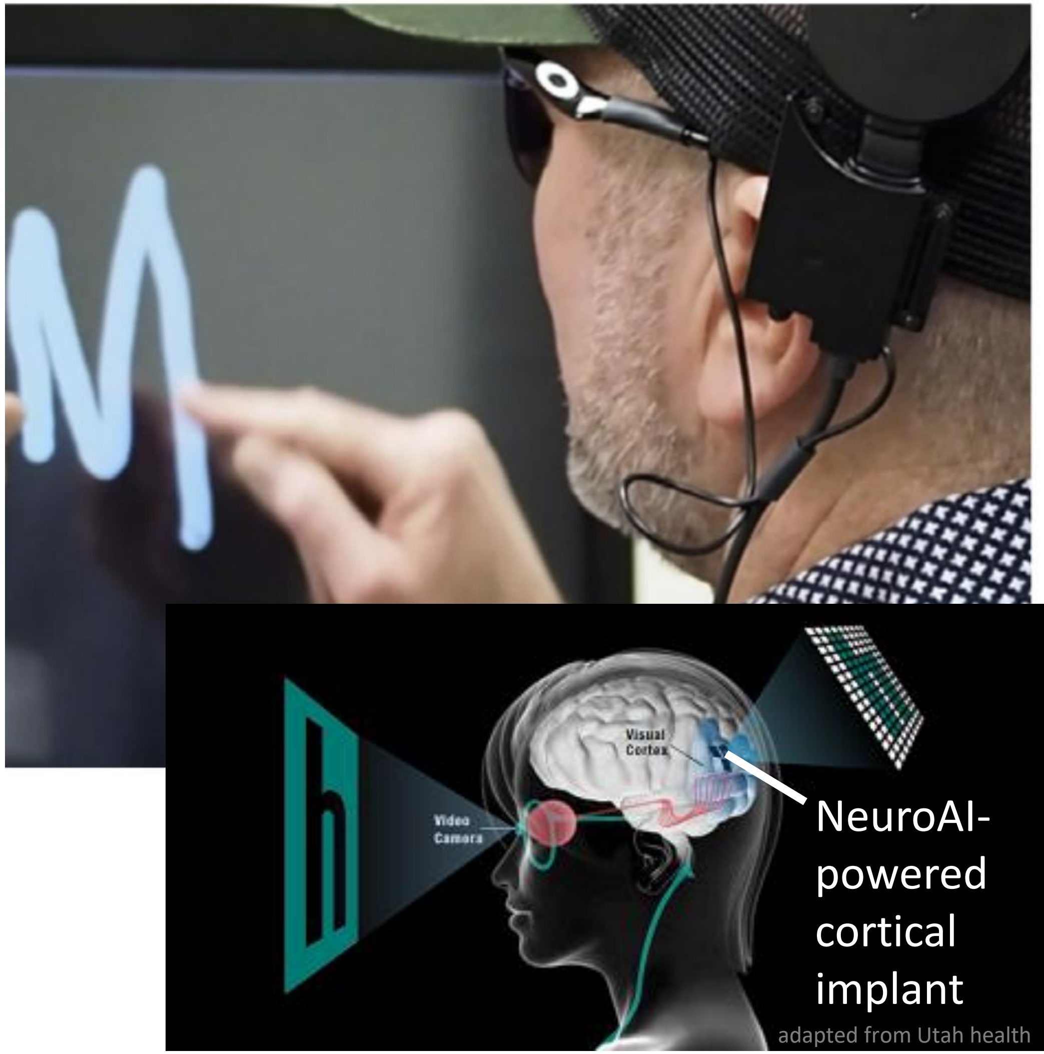
Changin
Notice h
Add, sor
Jiffy Lub
People c
Buy sell
Turin lov
URL righ

Suppress

Sentences identified to elicit minimal response in the language network

We were sitting on the couch.
That is such a beautiful picture!
They stood there for a moment.
They went up the stairs together.
Inside was a tiny silver sculpture.
They walked out onto the balcony.
Cas gazed up at the sky.
What else is there to do?



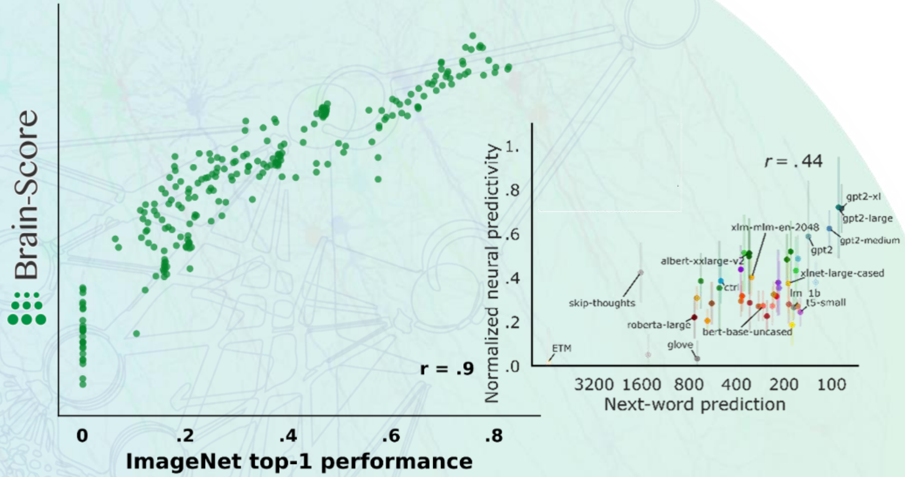


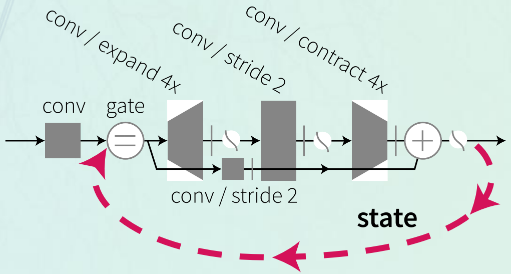
NeuroAI-powered cortical implant

adapted from Utah health

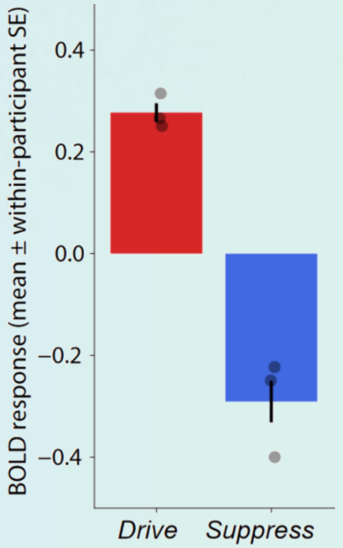
Tuckute et al. (2023)

Beauchamp et al. (2020) Chen et al. (2021) Schrimpf et al. (2024)





CORnet-S Area Circuitry



NeuroAI Lab

1. **Understand natural intelligence** by discovering relationships between brain alignment and computational objectives.

Brain-Score

2. Build **better** deep network **models** of brain and behavior.

3. Use the best models to **drive new experiments**, invasive and non-invasive, which might lead to future applications.