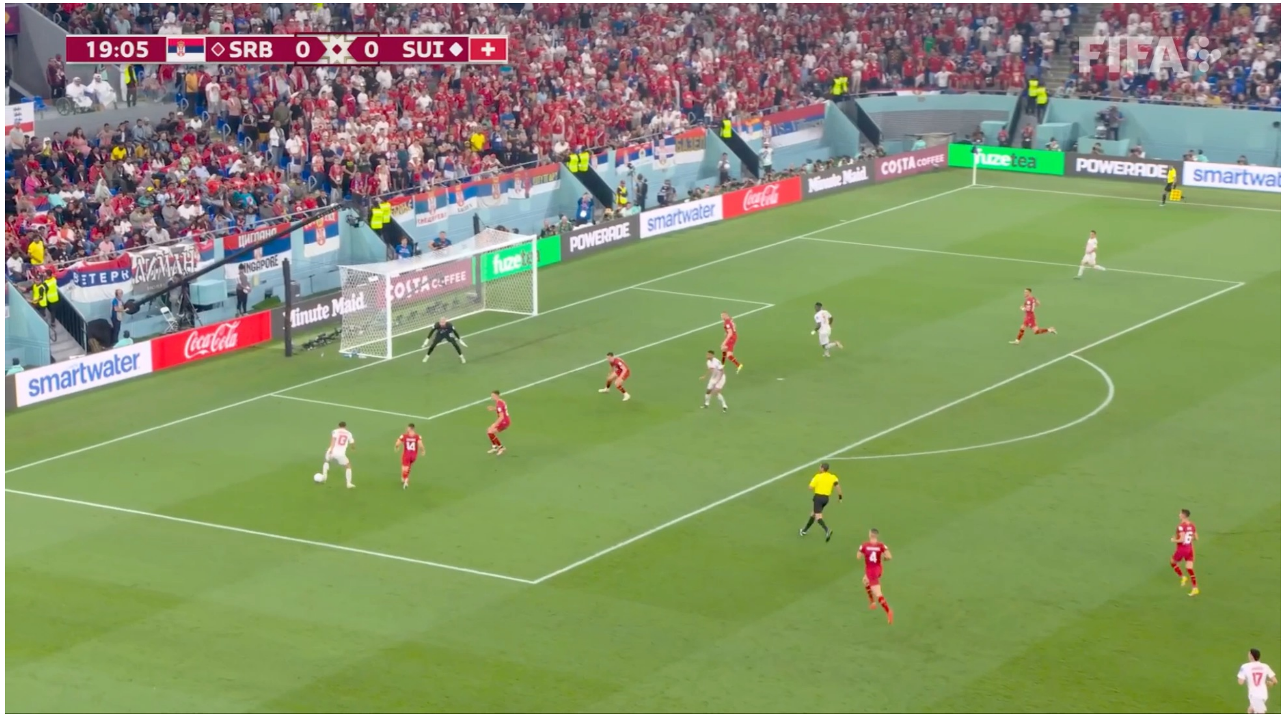
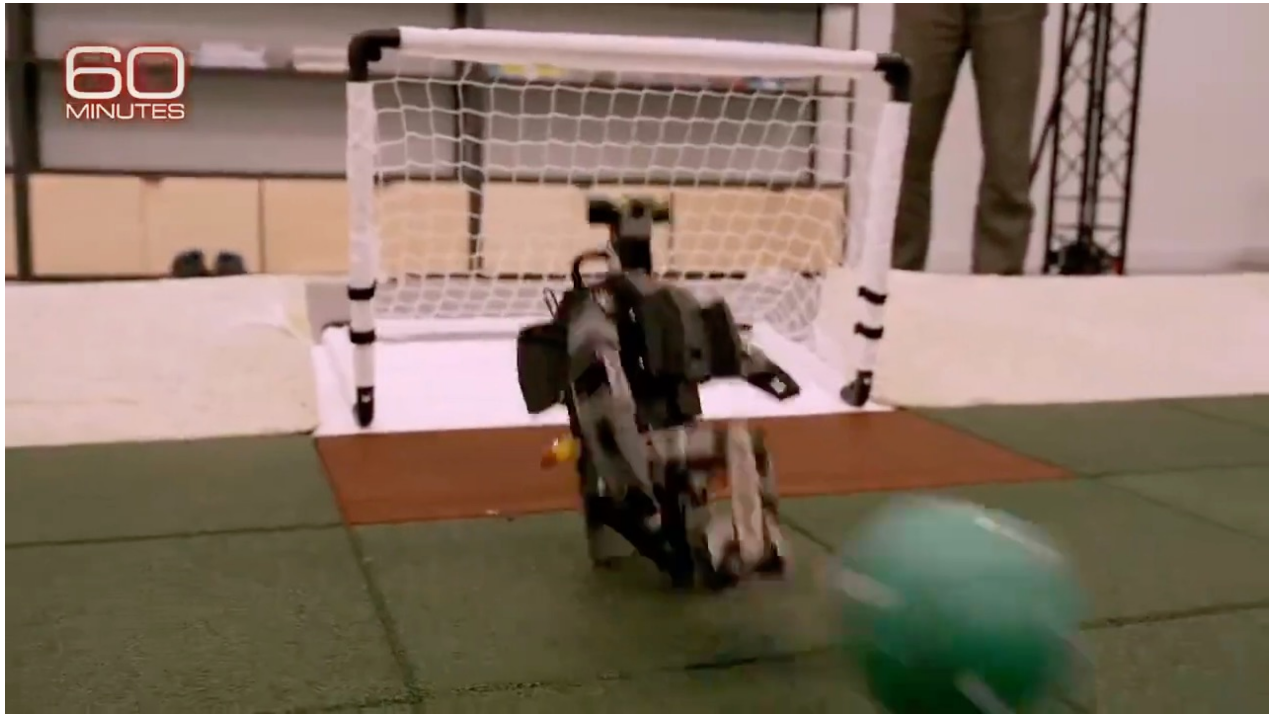
NX-435 NeuroAI

Prof. Mackenzie Mathis, PhD
Spring 2024

Biological Intelligence ⟷ Artificial Intelligence



World Cup 2022



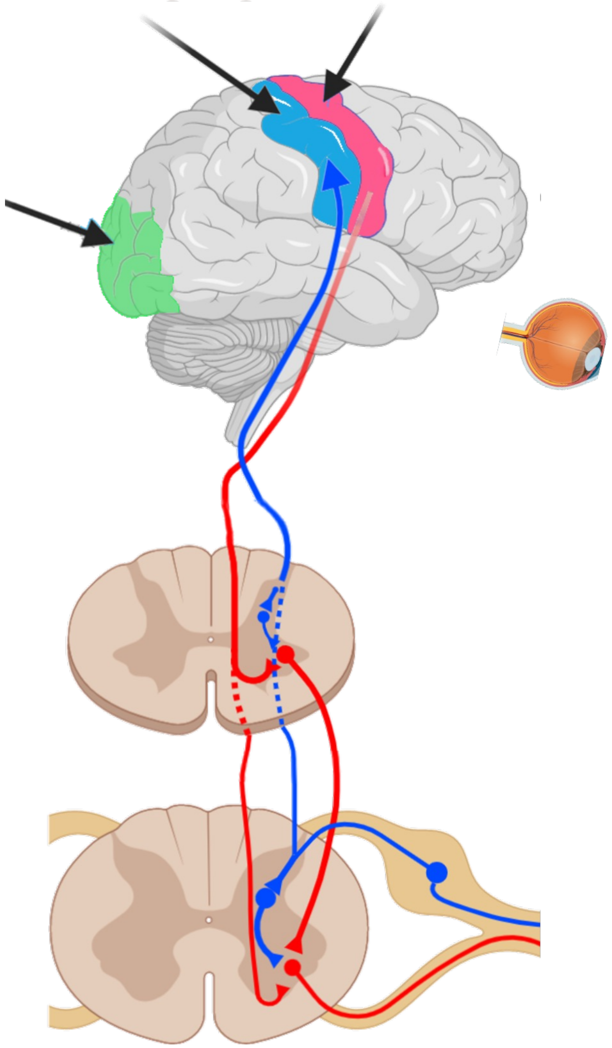
DeepMind 2023

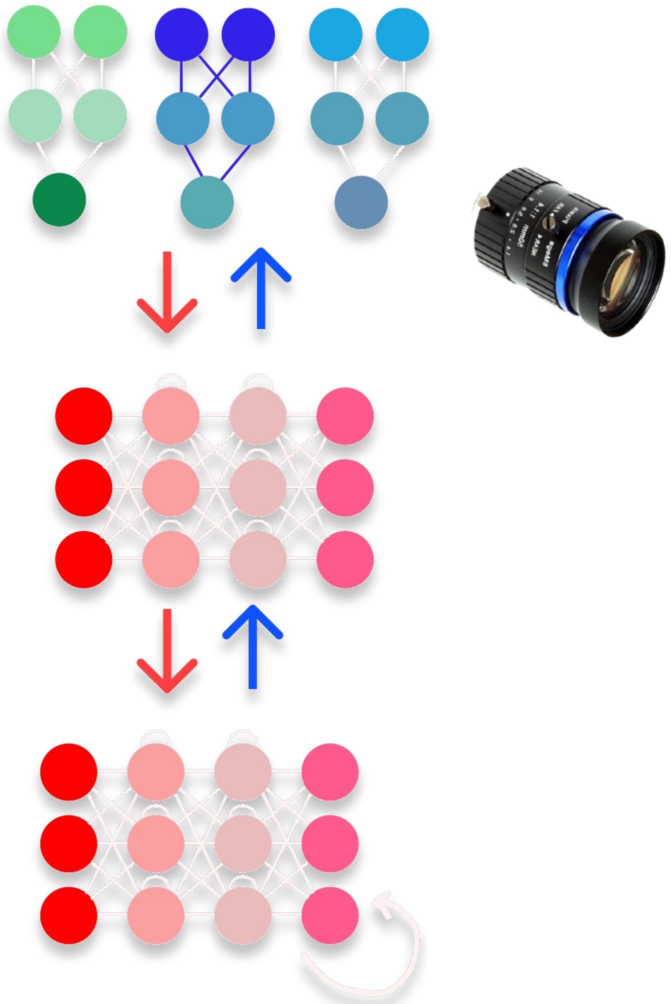
This is still a large gap in our ability to build adaptive embodied AI

Of course, artificial Intelligence has made huge leaps with LLMs…

tell me a joke about Swiss soccer players vs. DeepMind robots playing soccer

Biological Intelligence ⟷ Artificial Intelligence



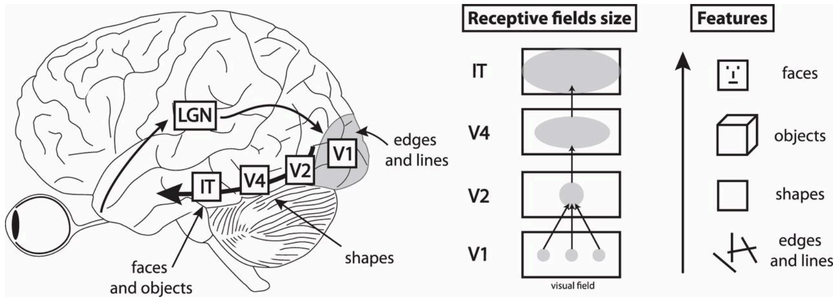


Hausmann*, Marin-Vargas*, Mathis**, Mathis**, Curr op. in Neuro., 2021

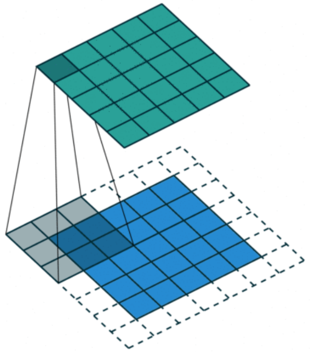
What is neuroAI?

Neuroscience

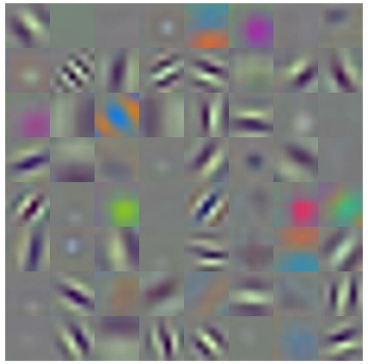
Artificial Intelligence (AI)



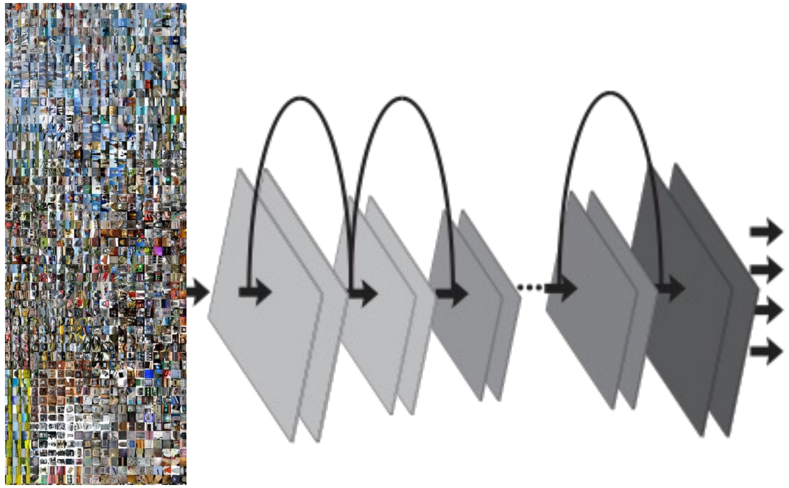
Hubel & Wiesel discoveries in cat V1 inspired convolutional neural networks



Convolutions (CNN)



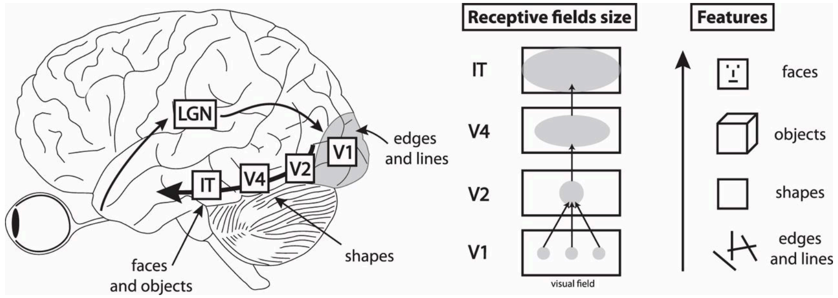
Representations in ImageNet trained (CNN)



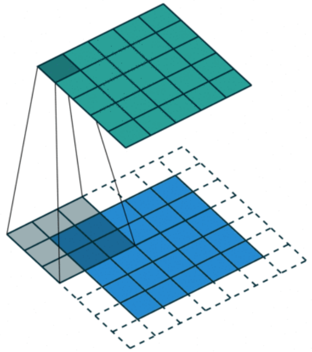
cat

Neuroscience

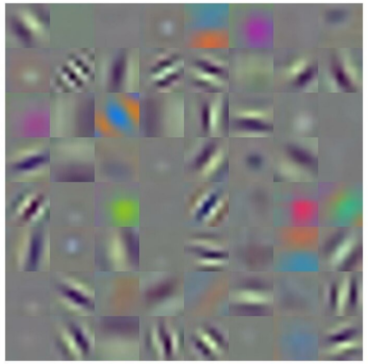
Artificial Intelligence (AI)



Hubel & Wiesel discoveries in cat V1 inspired convolutional neural networks



Convolutions (CNN)

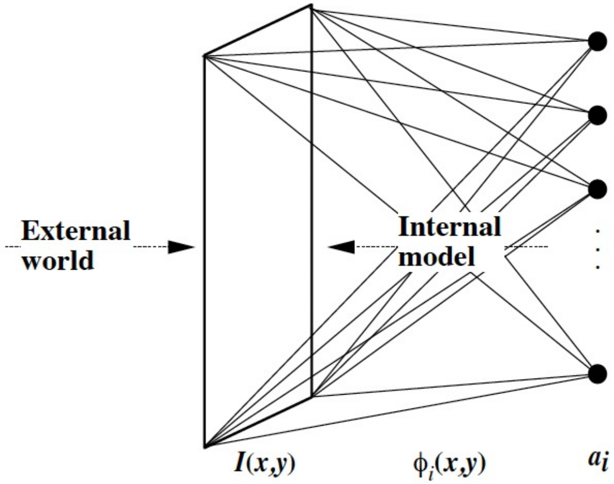


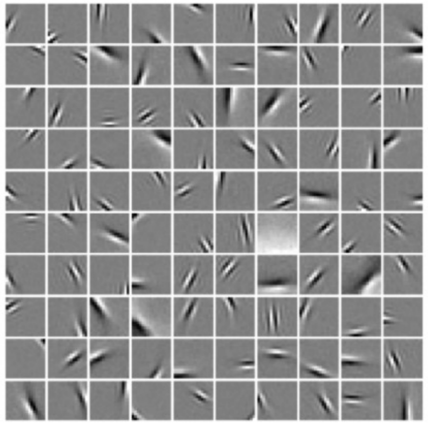
Representations in ImageNet trained (CNN)

edge detector neurons can be explained with sparse autoencoding

$$I(x,y) = \sum_i a_i \, \phi_i(x,y) + \epsilon(x,y)$$

Adapted from A. Mathis



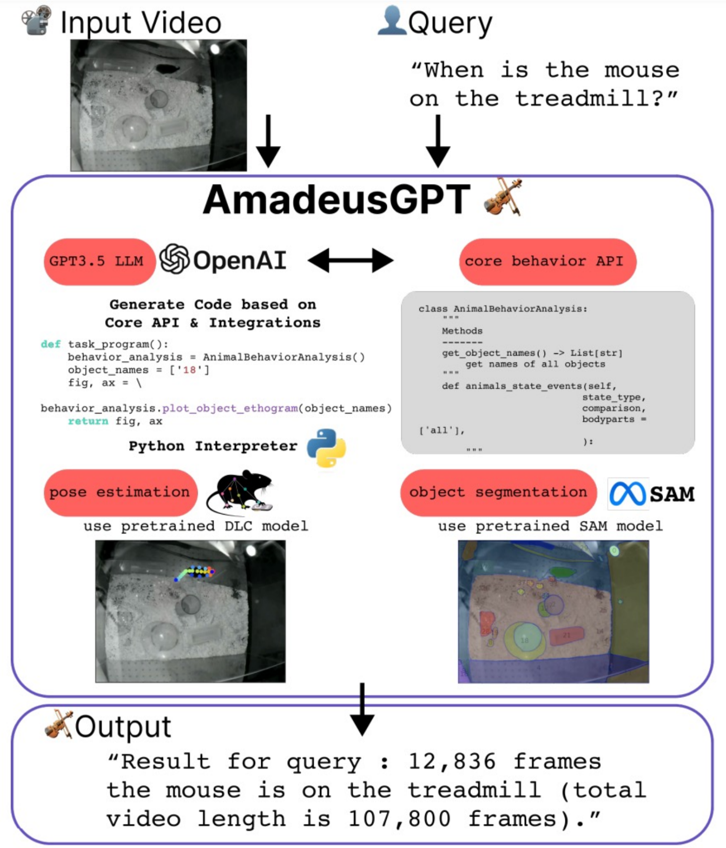


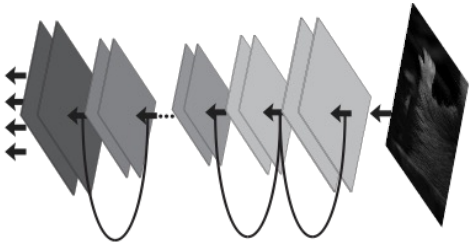
$\phi_i(x,y)$

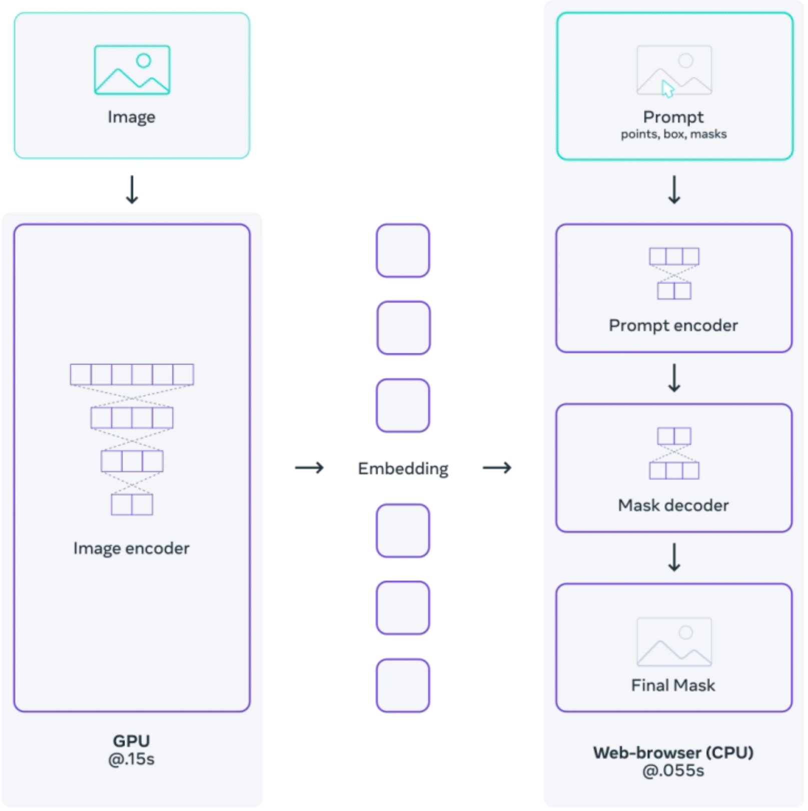
Olshausen & Field, 1996 Nature

What is neuroAI?

Neuroscience ⟷ Artificial Intelligence (AI)





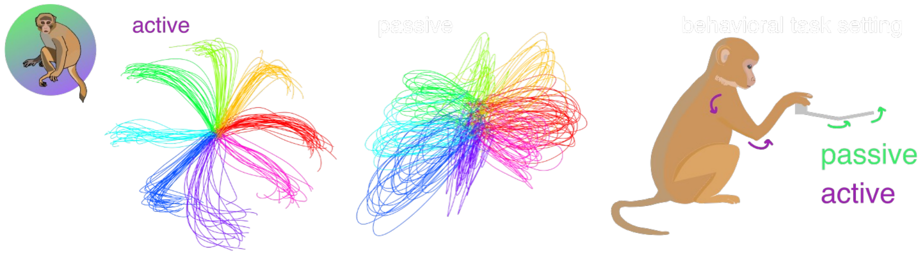


AmadeusGPT Ye et al. 2023

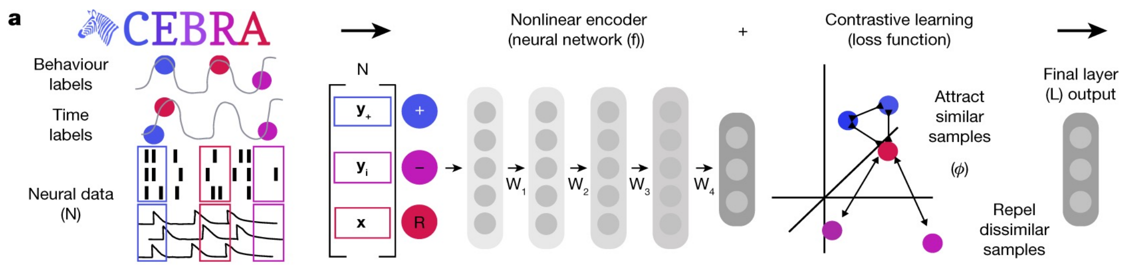
Segment Anything (Meta AI)

What is neuroAI?

Neuroscience ⟷ Artificial Intelligence (AI)



Data:Chowdhury et al 2020 eLife



Schneider, Lee, Mathis 2023 Nature

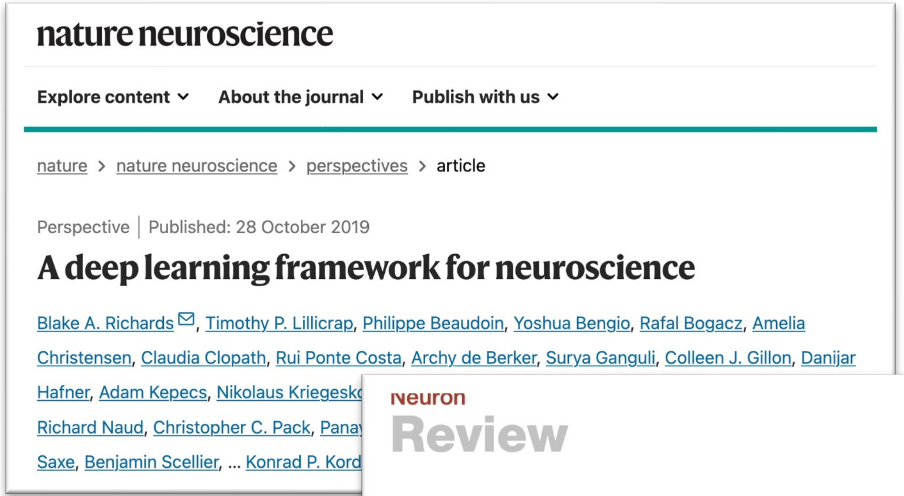
Which lead to better neural dynamical models of sensorimotor cortex (important for fundamental understanding and BCIs)

Advances in AI, such as **contrastive learning** in image processing could be expanded to neuro-specific domains

What is neuroAI:

- Many definitions, but widely accepted that it is the **new inter-disciplinary field of merging neuroscience and AI research (← →)**

- Others define is more narrowly as using neuroscience (→) to shape research in AI


nature neuroscience

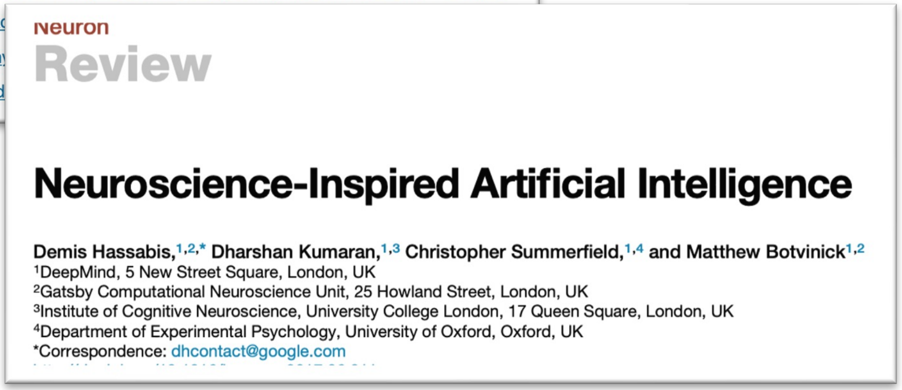
Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > nature neuroscience > perspectives > article

Perspective | Published: 28 October 2019

A deep learning framework for neuroscience

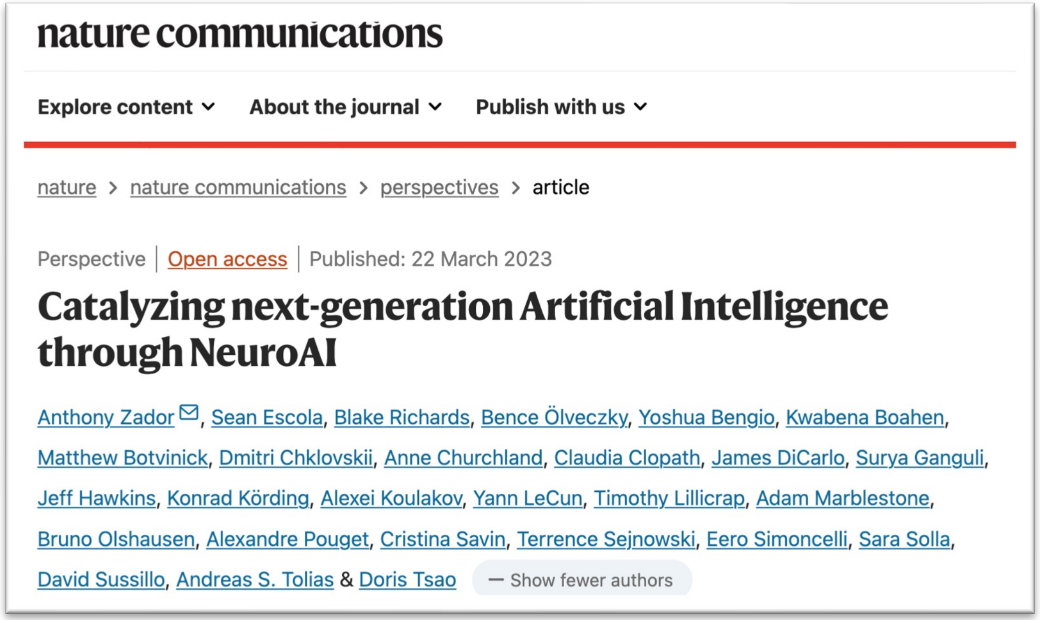
Blake A. Richards, Timothy P. Lillicrap, Philippe Beaudoin, Yoshua Bengio, Rafal Bogacz, Amelia Christensen, Claudia Clopath, Rui Ponte Costa, Archy de Berker, Surya Ganguli, Colleen J. Gillon, Danijar Hafner, Adam Kepecs, Nikolaus Kriegesk… Richard Naud, Christopher C. Pack, Pana… Saxe, Benjamin Scellier, ... Konrad P. Kord…


Neuron

Review

Neuroscience-Inspired Artificial Intelligence

Demis Hassabis,[1,2,*] Dharshan Kumaran,[1,3] Christopher Summerfield,[1,4] and Matthew Botvinick[1,2]
[1]DeepMind, 5 New Street Square, London, UK
[2]Gatsby Computational Neuroscience Unit, 25 Howland Street, London, UK
[3]Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London, UK
[4]Department of Experimental Psychology, University of Oxford, Oxford, UK
*Correspondence: dhcontact@google.com


nature communications

Explore content ⌄ About the journal ⌄ Publish with us ⌄

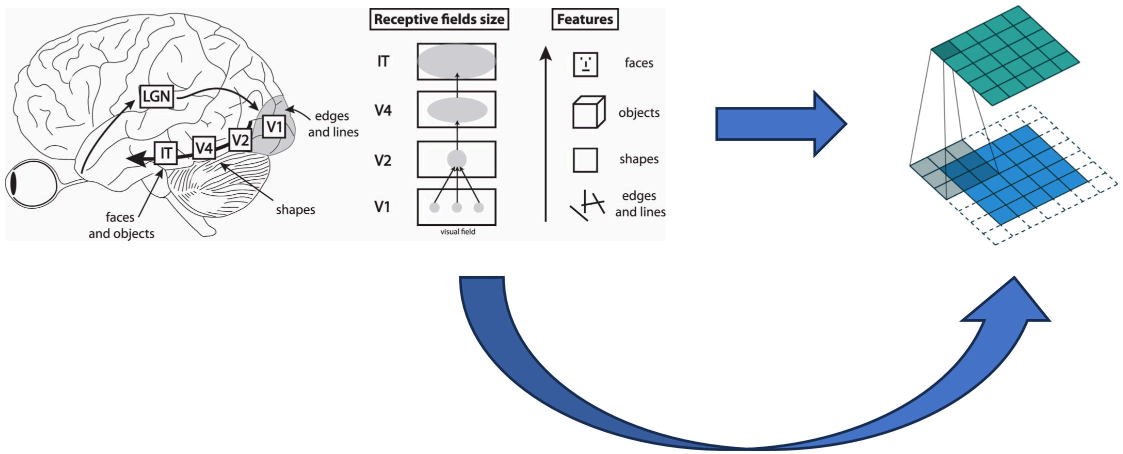
nature > nature communications > perspectives > article

Perspective | Open access | Published: 22 March 2023

Catalyzing next-generation Artificial Intelligence through NeuroAI

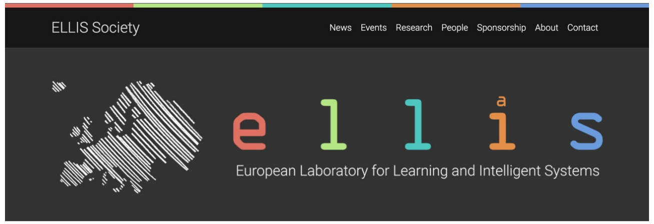
Anthony Zador, Sean Escola, Blake Richards, Bence Ölveczky, Yoshua Bengio, Kwabena Boahen, Matthew Botvinick, Dmitri Chklovskii, Anne Churchland, Claudia Clopath, James DiCarlo, Surya Ganguli, Jeff Hawkins, Konrad Körding, Alexei Koulakov, Yann LeCun, Timothy Lillicrap, Adam Marblestone, Bruno Olshausen, Alexandre Pouget, Cristina Savin, Terrence Sejnowski, Eero Simoncelli, Sara Solla, David Sussillo, Andreas S. Tolias & Doris Tsao — Show fewer authors

One rapidly growing area is trying to build better AI by better understanding Biological intelligence (BI) / Natural Intelligence (NI)



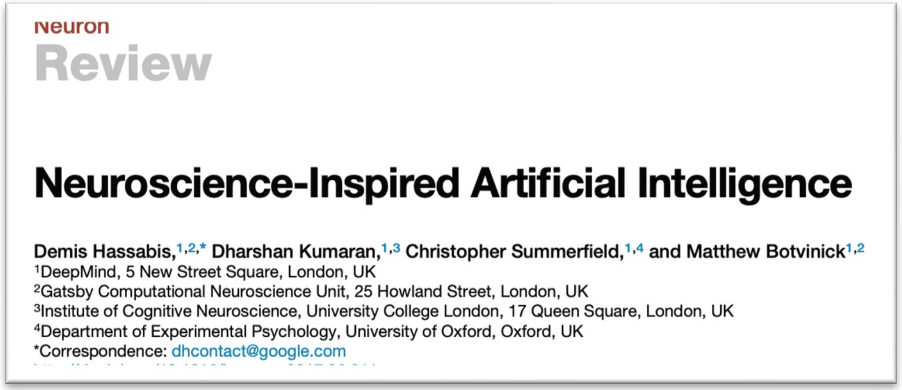
How can we build better models of neural systems, and what is the role of systems neuroscience?

https://ellis.eu/programs/natural-intelligence



note, ELLIS also has the top umbrella PhD program in Europe for ML/AI research

- “If a new facet of biological computation found to be critical to supporting a cognitive function, then we would consider it an excellent candidate for incorporation into artificial systems”


Neuron

Review

Neuroscience-Inspired Artificial Intelligence

Demis Hassabis,[1,2,*] Dharshan Kumaran,[1,3] Christopher Summerfield,[1,4] and Matthew Botvinick[1,2]

[1]DeepMind, 5 New Street Square, London, UK

[2]Gatsby Computational Neuroscience Unit, 25 Howland Street, London, UK

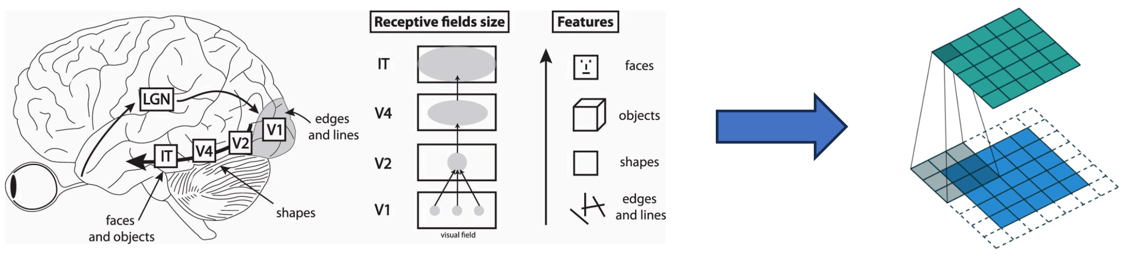
[3]Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London, UK

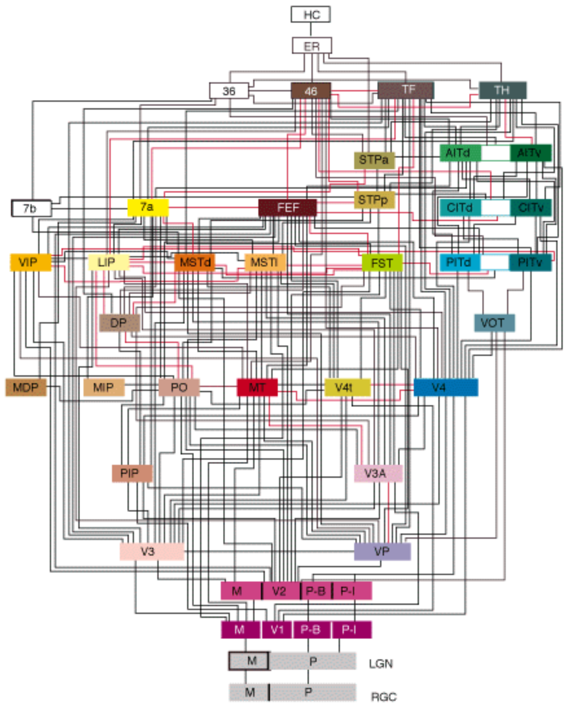
[4]Department of Experimental Psychology, University of Oxford, Oxford, UK

*Correspondence: dhcontact@google.com

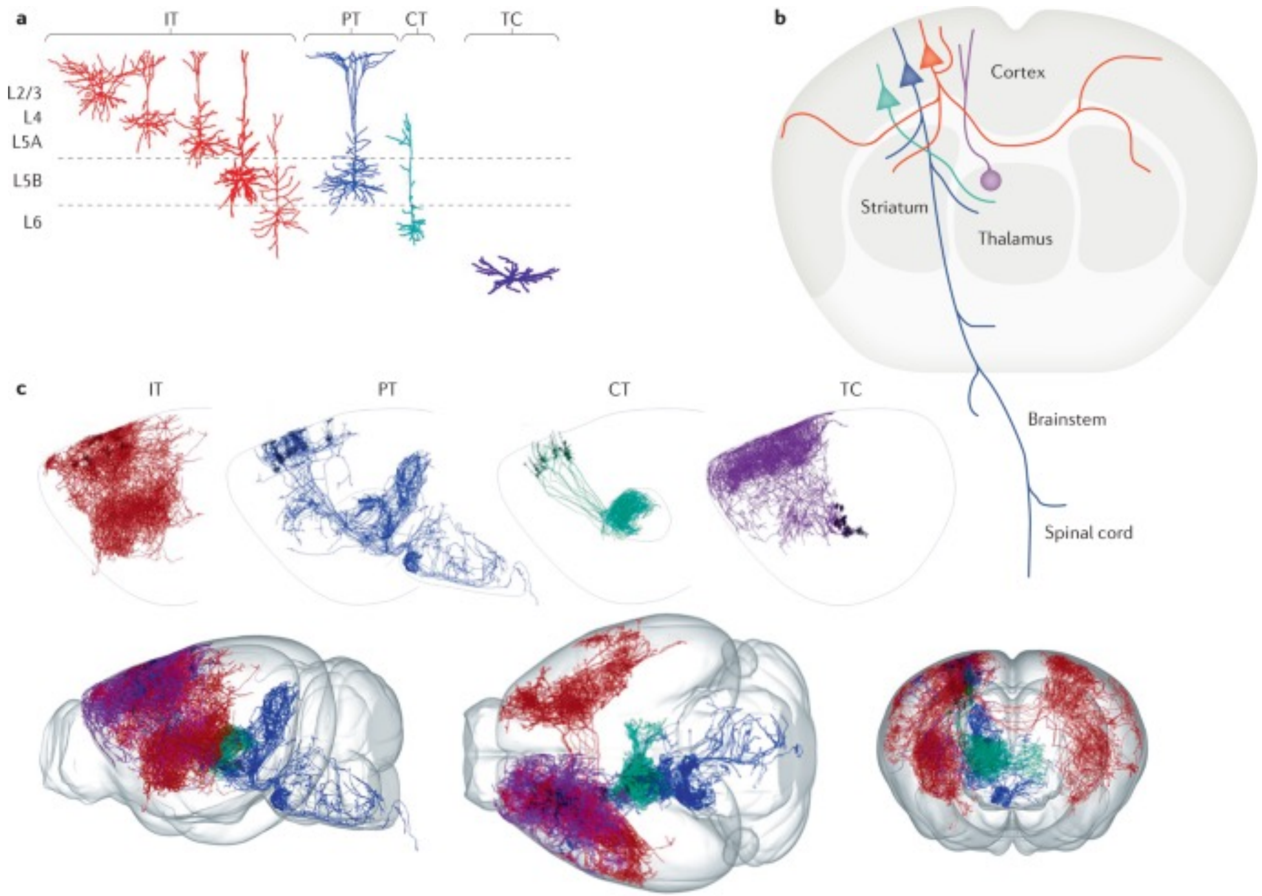
“neuroscience can provide validation of AI techniques that already exist. If a known algorithm is subsequently found to be implemented in the brain, then that is strong support for its plausibility as an integral component of an overall general intelligence system”

Neural circuits are much more complex than modern AI systems ...



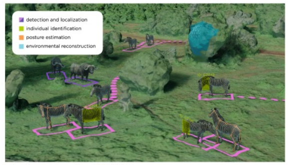


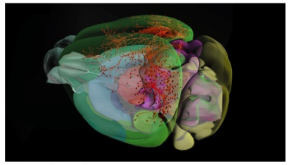
Felleman and Van Essen *Cerebral Cortex* 1991

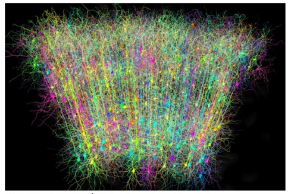


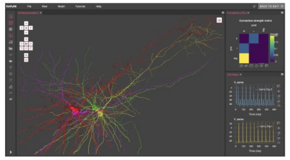
Shepherd and Yamawaki *Nature review neuroscience* 2021

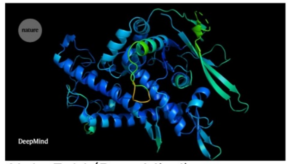
W 5 & 6 we covered (some) tools for measuring and modeling behavioral and neural data with data-driven approaches ...


Tuia et al. 2022


Allen Institute

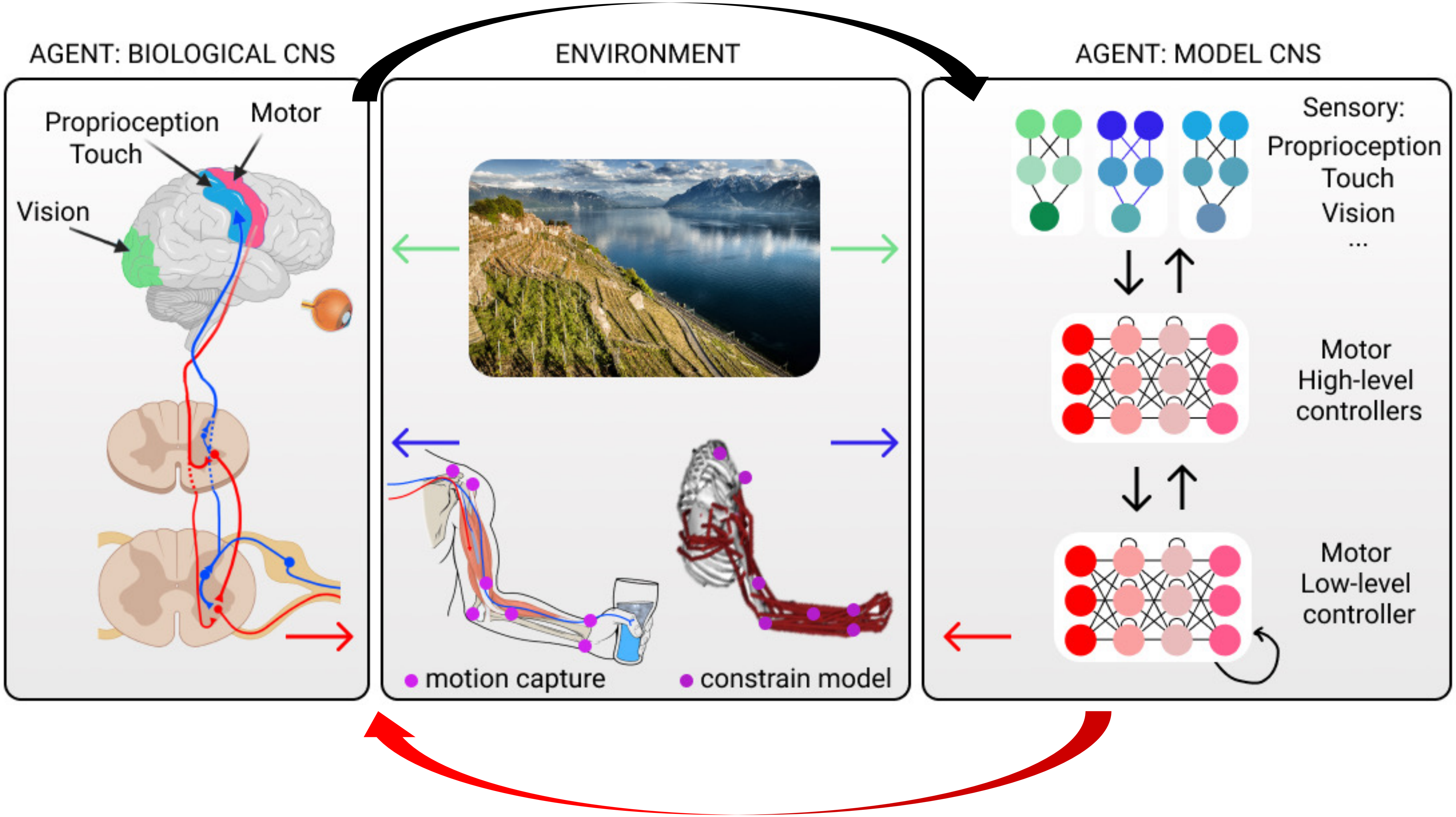

Brainbow (Litchman Lab)


NetPyNE


AlphaFold (DeepMind)

	Experimentally-derived data	Data-driven & Theory-driven models
Environment		multi-modal ML models
Behavior	video recordings, EMGs	computer vision & RL
Brain	functional imaging (2P, 1P, fMRI, EEG)	network neuroscience
system (circuits)		RNNs, transformers GLMs, latent variable models
Neurons (& glia)	pharmacology, patch clamping	Hodgkin-Huxley LIF (NEURON, nengo, etc)
Synapses		ODEs, LIF
Genes & molecules, proteins	RNAseq, proteomics	AlphaFold, gene-regulatory networks

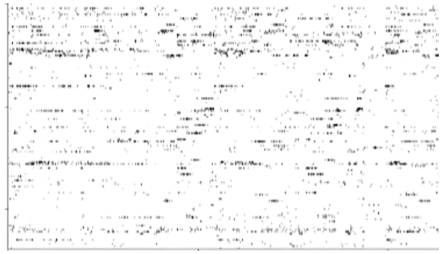
Data-driven and Task-driven modeling for understanding BI



Hausmann et al. Current Opinion in Neurobiology 2021

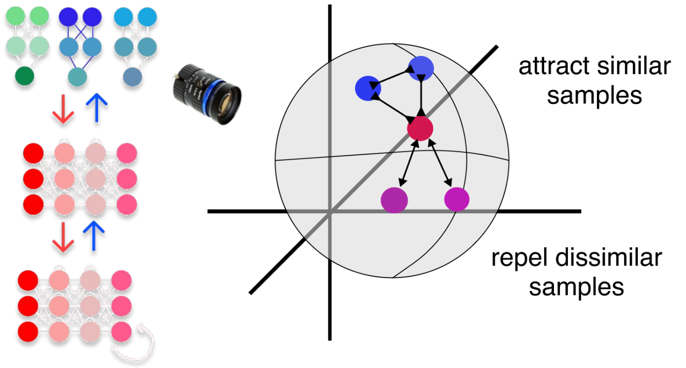
Data-driven modeling

GLMs, PCA, Sussillo et al. 2015 Nat Neuro
State-space models, …



Record from neural data during a behavioral task

GLMs
.
.
.
ANNs (CEBRA)

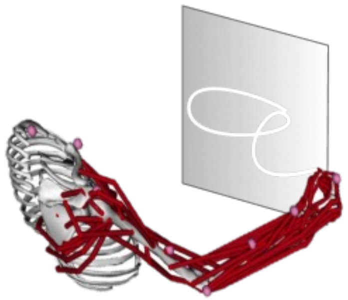


Schneider Lee Mathis 2023 Nature

Joint models that describe neural variance & representations

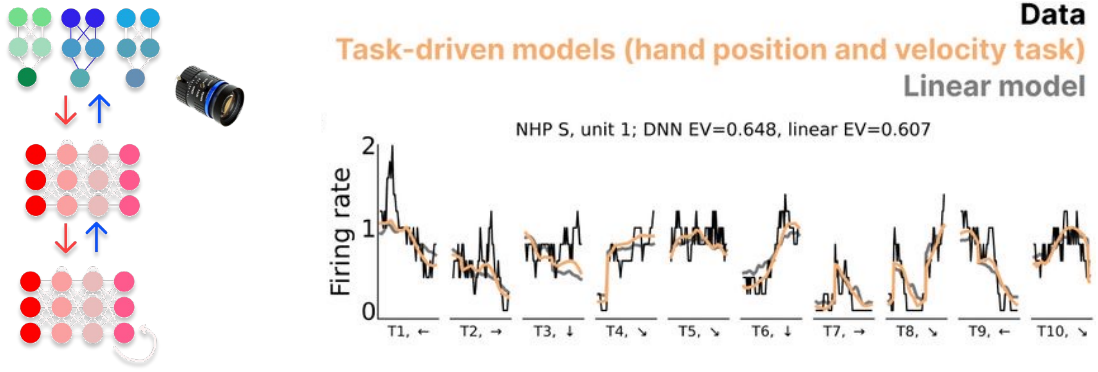
Task-driven modelling

Yamins et al. PNAS 2014, Kell et al. 2018 Neuron, Banino et al. 2018 Nature ….



Constrain ANN based on behavioral task to test hypotheses about a system

Sandbrink et al. 2023 eLife



Marin Vargas et al. 2024 Cell

NN models that describe neural variance & computationally constrain system

Data-driven modeling

"If a new facet of biological computation found to be critical to supporting a cognitive function, then we would consider it an excellent candidate for incorporation into artificial systems"

Task-driven modelling

"neuroscience can provide validation of AI techniques that already exist. If a known algorithm is subsequently found to be implemented in the brain, then that is strong support for its plausibility as an integral component of an overall general intelligence system"

Neuron
Review

Neuroscience-Inspired Artificial Intelligence

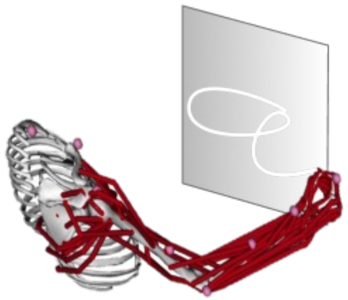
Demis Hassabis,[1,2,*] Dharshan Kumaran,[1,3] Christopher Summerfield,[1,4] and Matthew Botvinick[1,2]
[1]DeepMind, 5 New Street Square, London, UK
[2]Gatsby Computational Neuroscience Unit, 25 Howland Street, London, UK
[3]Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London, UK
[4]Department of Experimental Psychology, University of Oxford, Oxford, UK
*Correspondence: dhcontact@google.com

Task-driven modelling

Yamins et al. PNAS 2014, Sussillo et al. 2015 Nat Neuro, Kell et al. 2018 Neuron, Banino et al. 2018 Nature ….

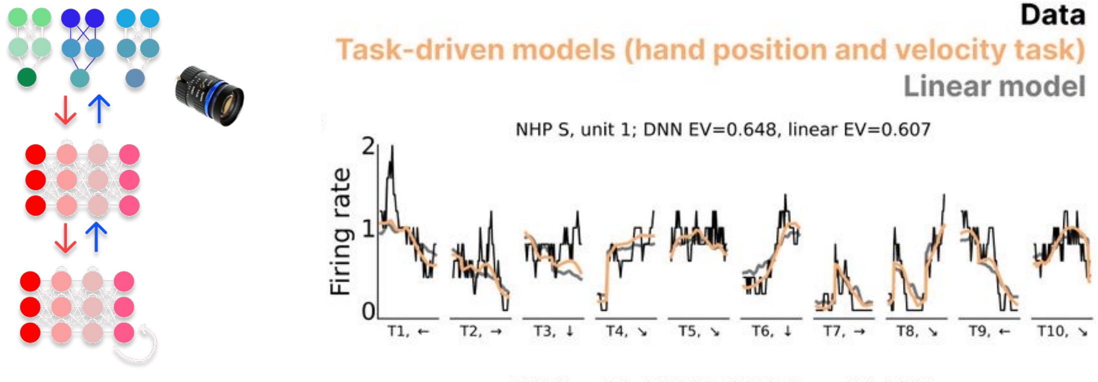
how task-driven models can inform us about the computational goals of the brain

(which can lead to validations and/or new ideas for AI)



Constrain ANN based on behavioral task to test hypotheses about a system

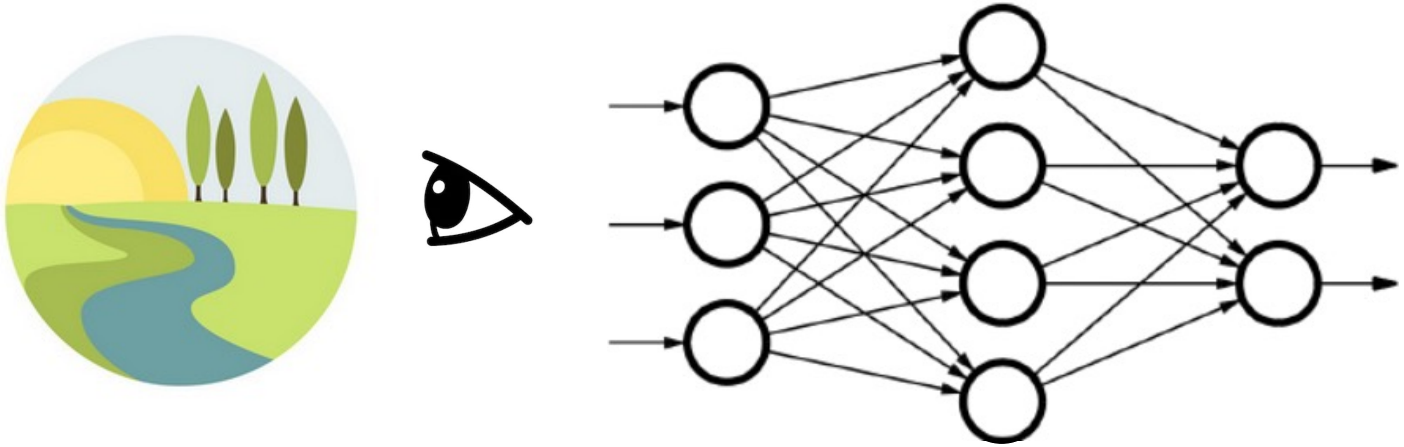
Sandbrink et al. 2023 eLife



Marin Vargas et al. 2024 Cell

NN models that describe neural variance & computationally constrain system

Using deep neural networks as task-driven models of a system



cat

Vision: Yamins et al. (2014)

Audition: Kell et al. (2018)- speech recognition, speaker identification, natural sound identification

Barrel Cortex: Zhuang et al. (2017)

Cognition: Mante et al. (2013)

Proprioception: Sandbrink et al. (2023), Marin Vargas*, Bisi* et al. (2024)

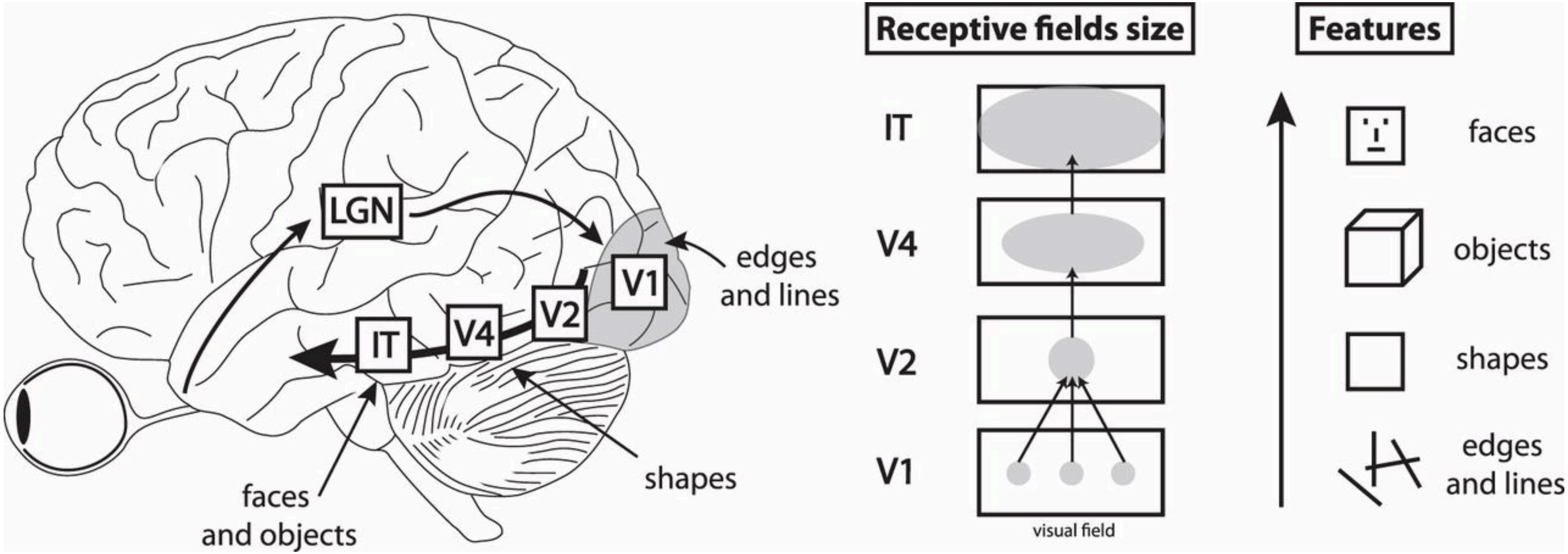
The goal: to build NN models that are constrained to goals of a neural system

Machine learning	Neuroscience
Architecture	Circuits
Task / objective	Ecological niche
Dataset	Environment
Optimization method (learning rule)	Natural selection + synaptic plasticity

NN model

Slide courtesy of A. Mathis

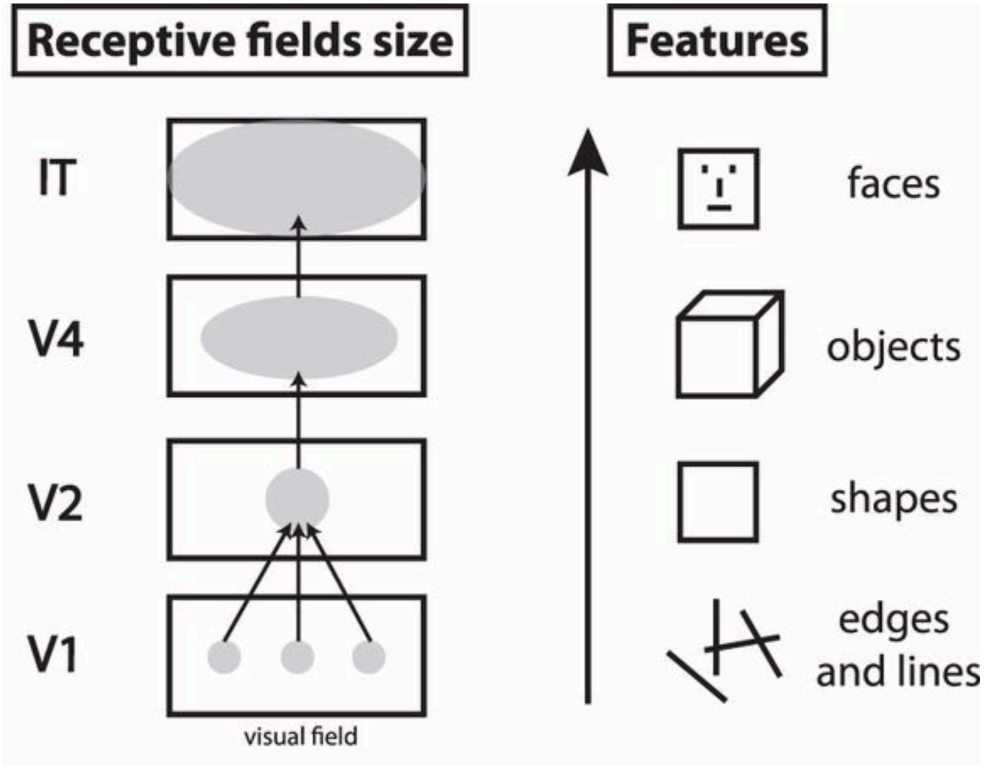
Hierarchical visual processing



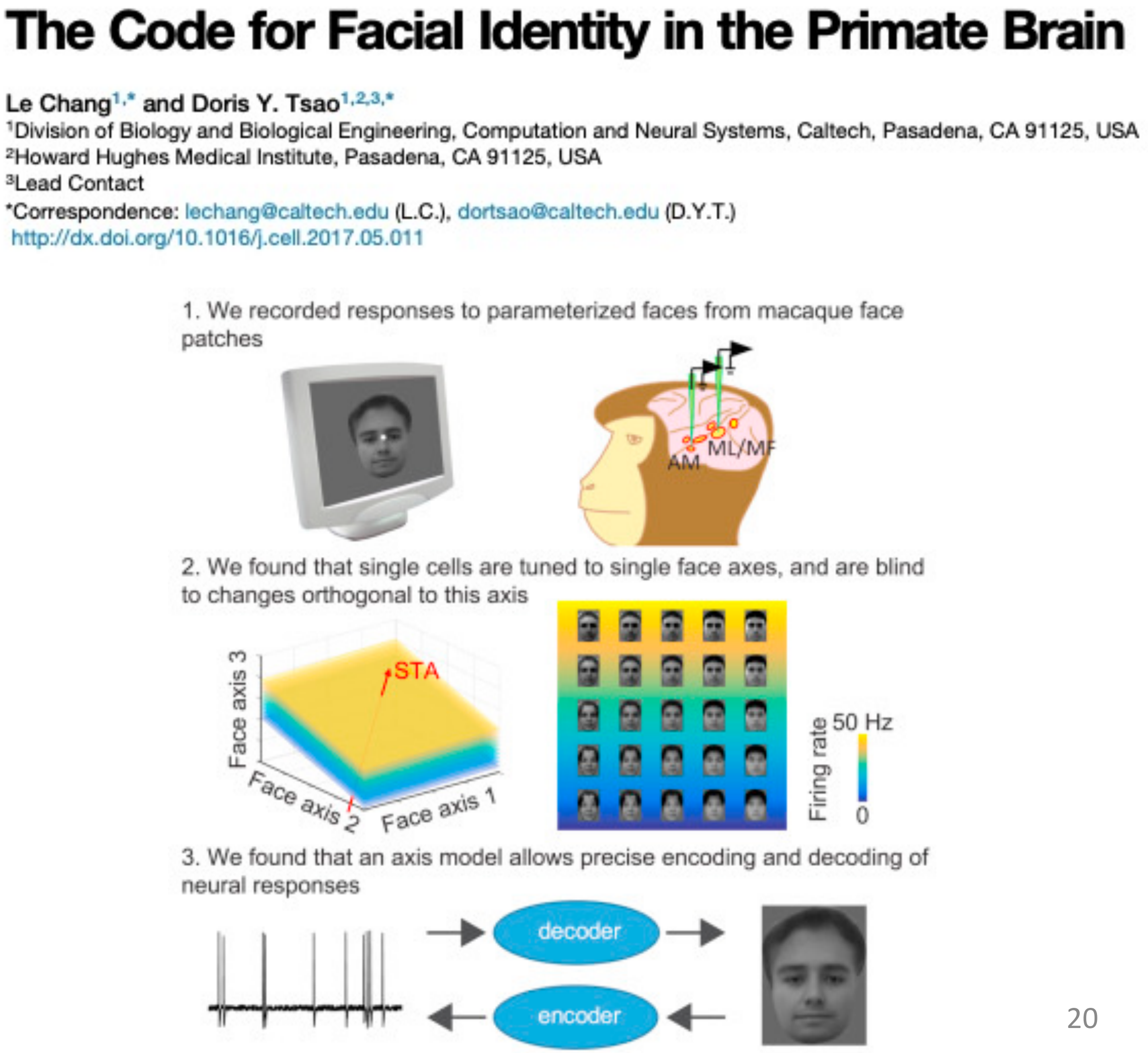
- V1 neurons are most sensitive to low-level features, such as edges and lines.
- In higher visual areas, like V4 and IT, receptive fields are larger, and neurons are sensitive to complex features, such as shapes and objects.
- Responses of high-level neurons are fully determined by the neural firing of lower-level neurons. For example, the neural firing to a square is determined by the neural firing for two vertical and two horizontal lines.

From: When crowding of crowding leads to uncrowding
Journal of Vision. 2013;13(13):10. doi:10.1167/13.13.10

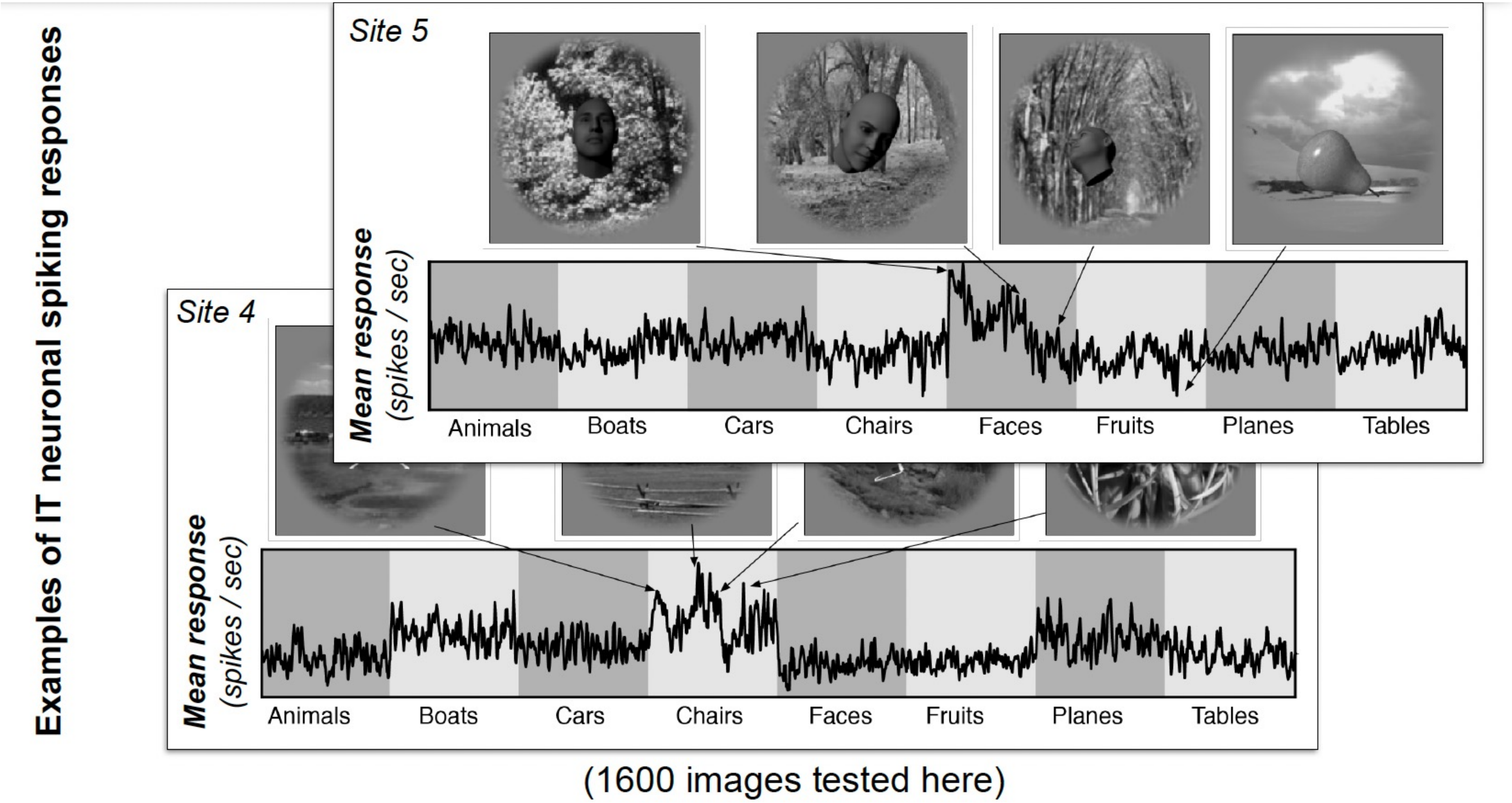
IT neurons are nonlinear



From: When crowding of crowding leads to uncrowding
Journal of Vision. 2013;13(13):10. doi:10.1167/13.13.10

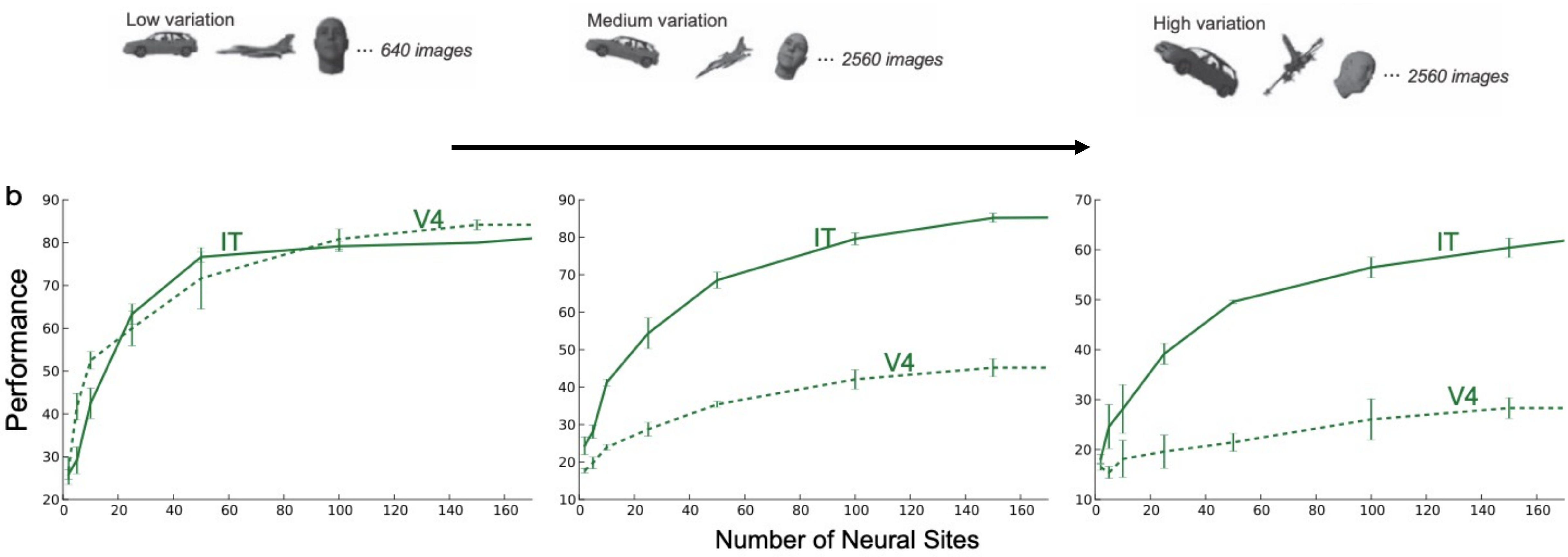


Example higher-order visual cortex responses



Slide from Jim DiCarlo, MIT

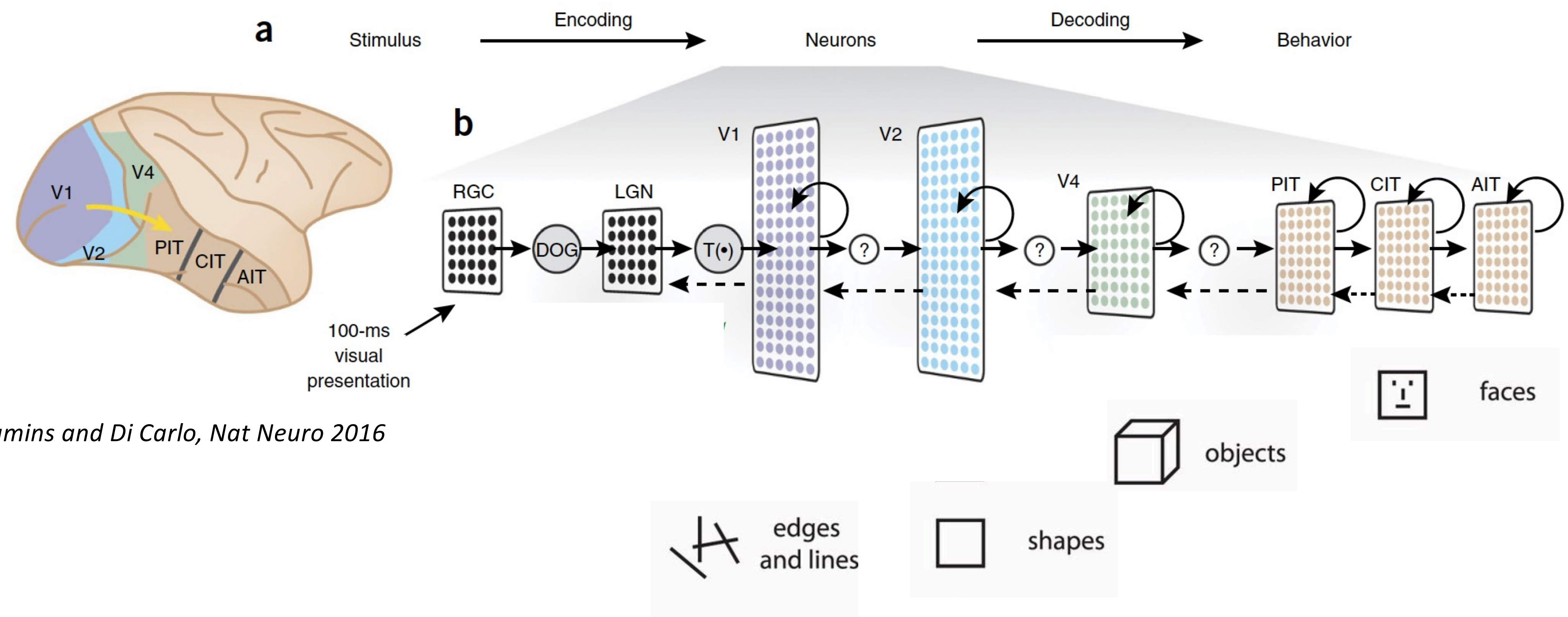
Decoding object identity from neural data



Slide courtesy of A. Mathis

Yamins et al., PNAS 2014

Core-object recognition and the visual pathway



Yamins and Di Carlo, Nat Neuro 2016

Building models of visual pathway: what is the computational goal?

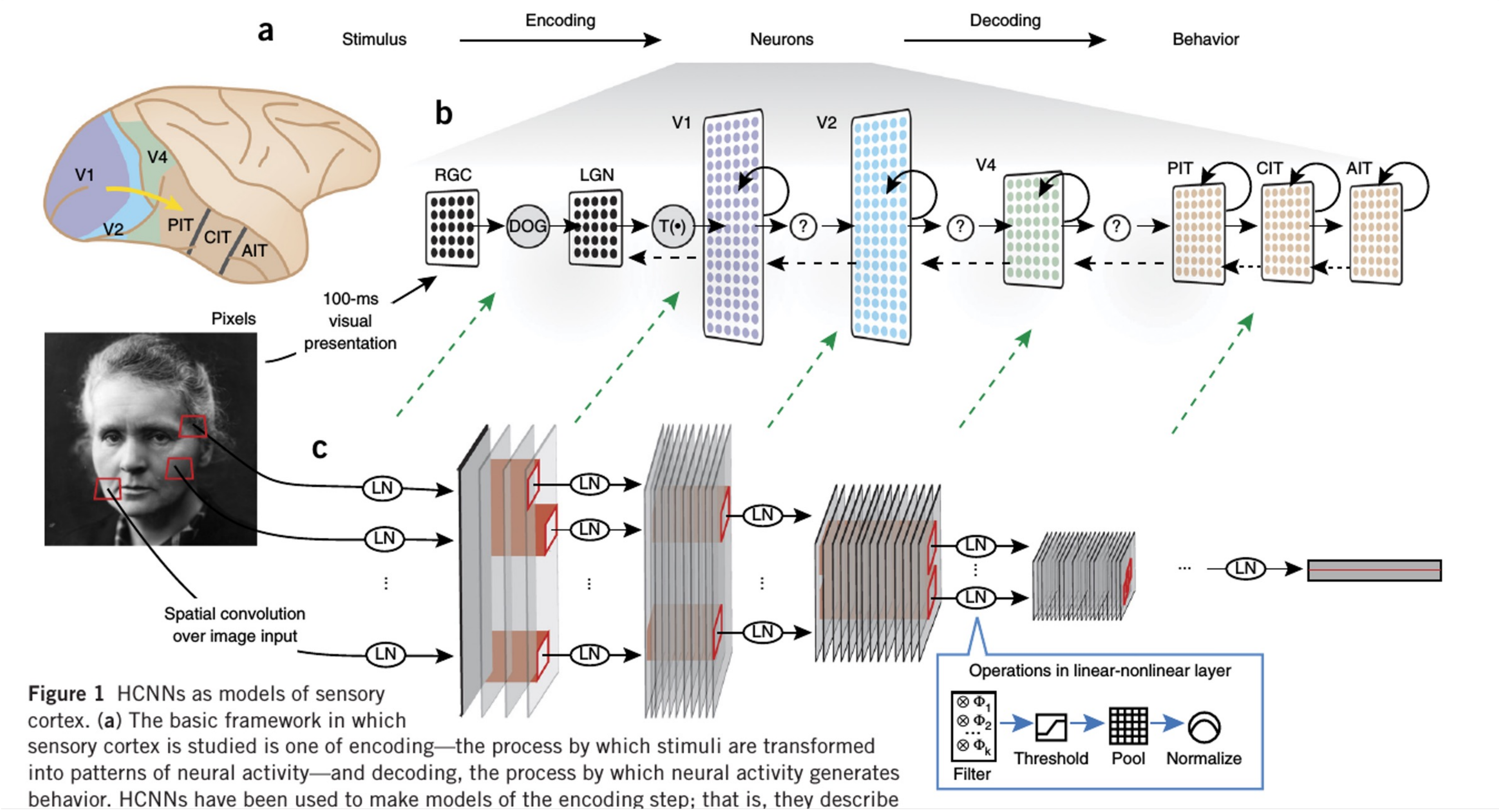


Figure 1 HCNNs as models of sensory cortex. (**a**) The basic framework in which sensory cortex is studied is one of encoding—the process by which stimuli are transformed into patterns of neural activity—and decoding, the process by which neural activity generates behavior. HCNNs have been used to make models of the encoding step; that is, they describe

Yamins & DiCarlo Nature Neuroscience 2016

Task information consistency

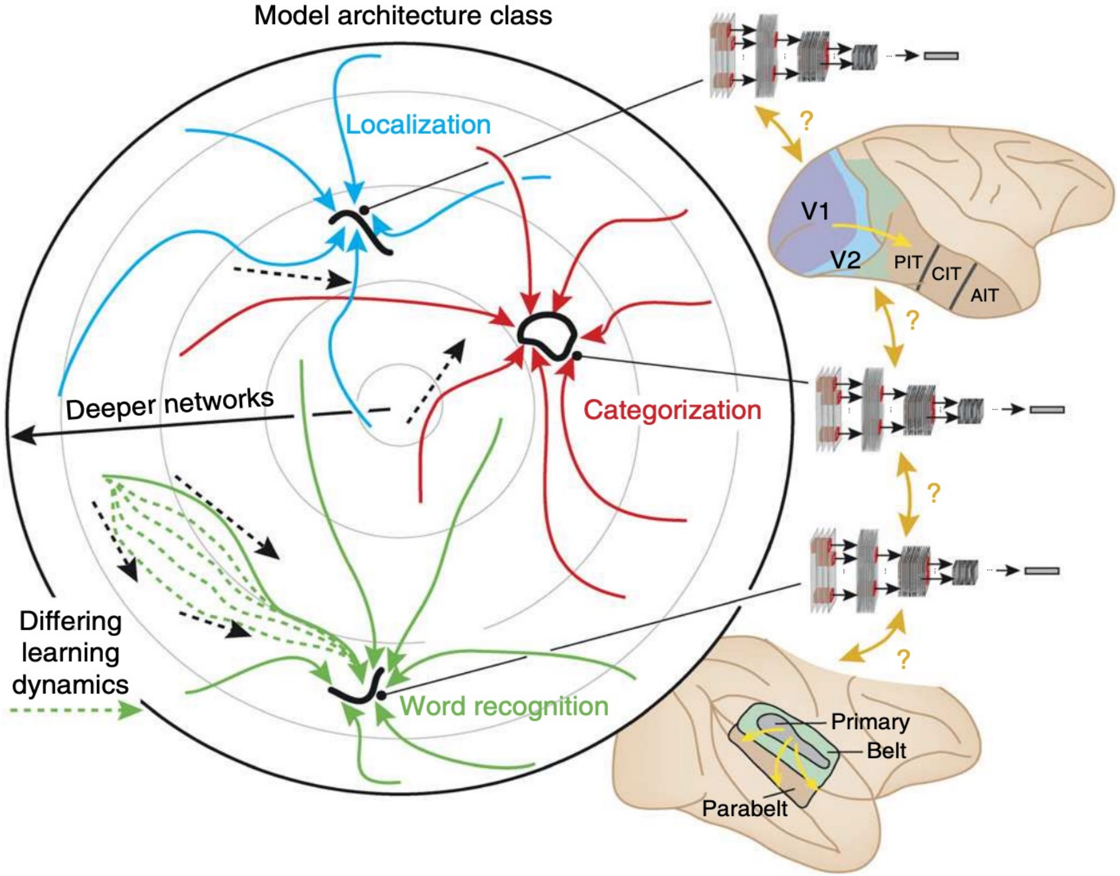
Single-unit response predictivity

Population representational similarity

Building models of visual pathway: the ingredients

Task-driven deep neural network models are built from three basic components:

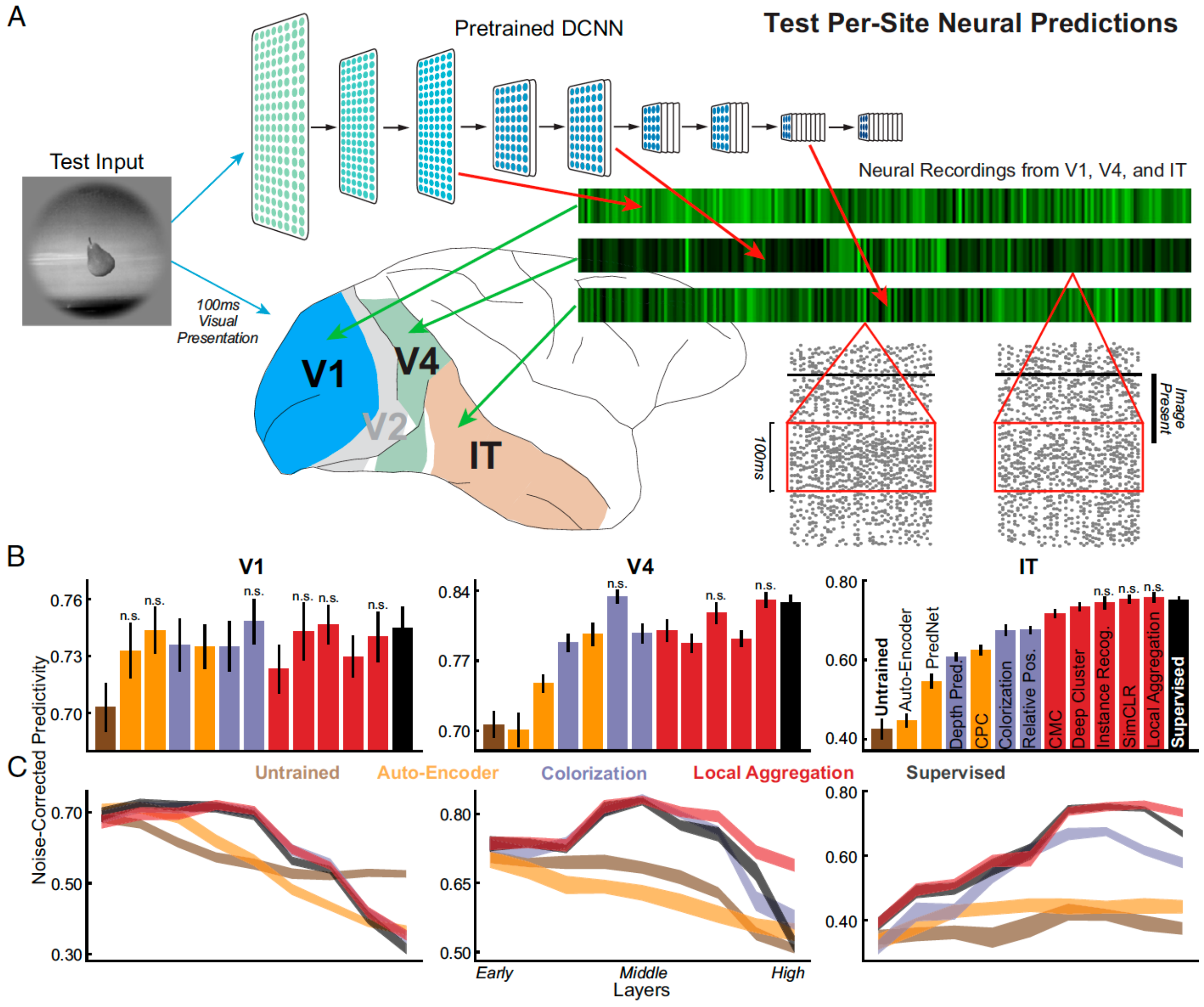
1. **model architecture class** from which the system is built, formalizing knowledge about the brain's anatomical and functional connectivity;
2. a **behavioral goal** that the system must accomplish, such as object categorization; and
3. **a learning rule** that optimizes parameters within the model class to achieve the behavioral goal.



Yamins & DiCarlo (2016)

Learning, architecture search, and model comparison across visual areas

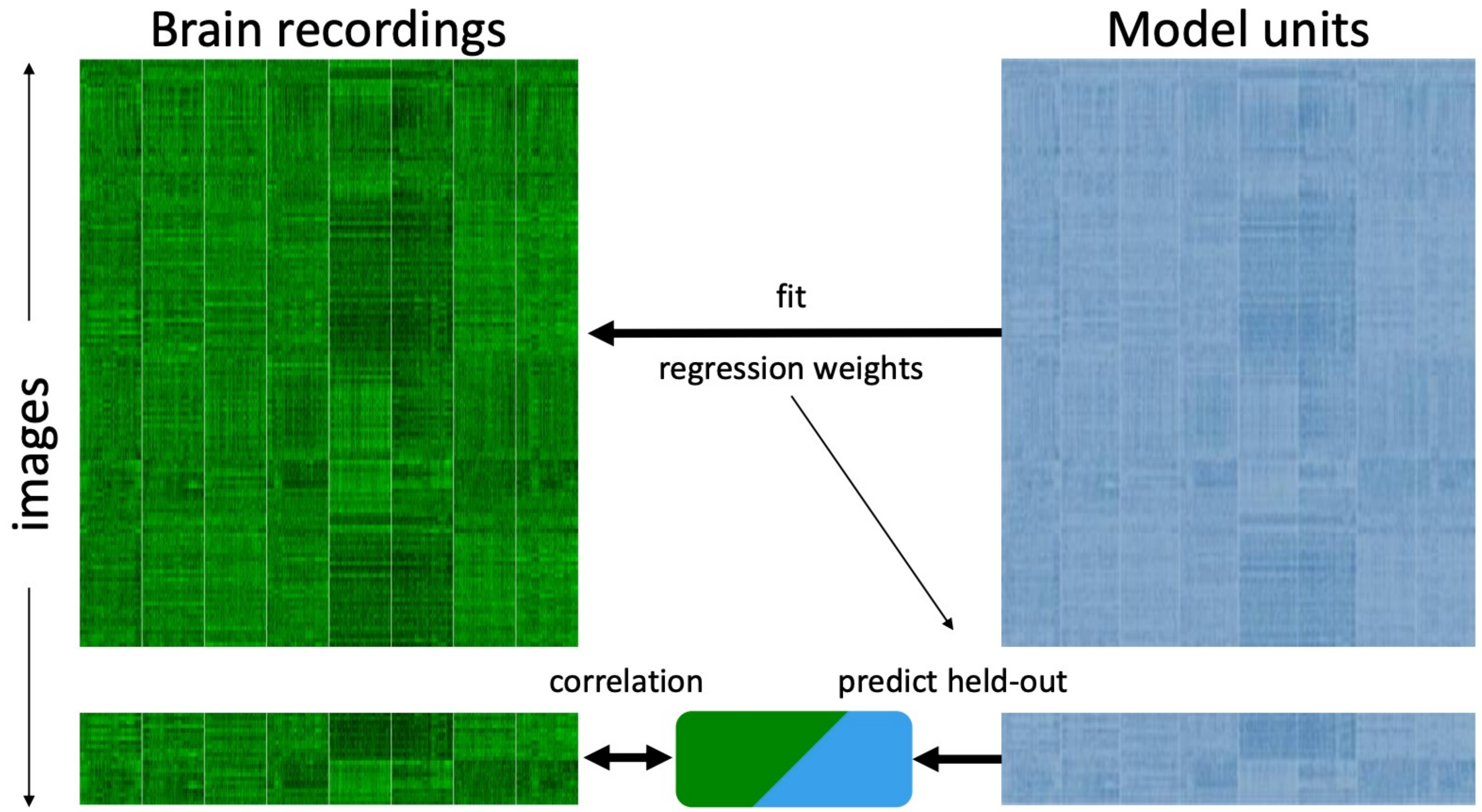
- recordings from across the visual hierarchy
- trained models outperform untrained models
- Supervised ImageNet models do very well, and so do unsupervised models…
- Early-Middle layers best predict V1, middle layer predict V4 and High layers best predict IT



Zhuang et al., PNAS 2020

Representational similarity analysis

similarity metric
neural predictivity



Yamins*, Hong*, et al. (PNAS 2014)   Schrimpf*, Kubilius*, et al. (bioRxiv 2018)

Slide from Martin Schrimpf

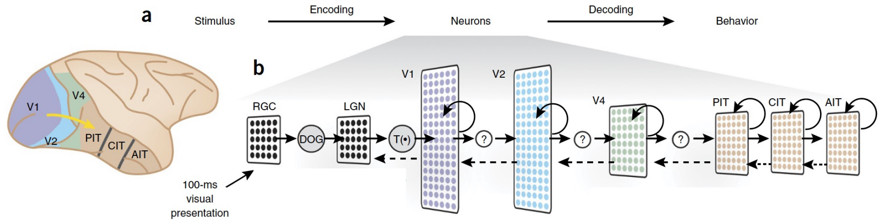
Intermediate Take Homes:

- NN models can be trained on different visual tasks to make hypotheses about the goal of the visual system
- Better NN models at the categorization task predicted IT neurons better
- Task mattered more than architecture or depth of networks
- Three points to consider when comparing:

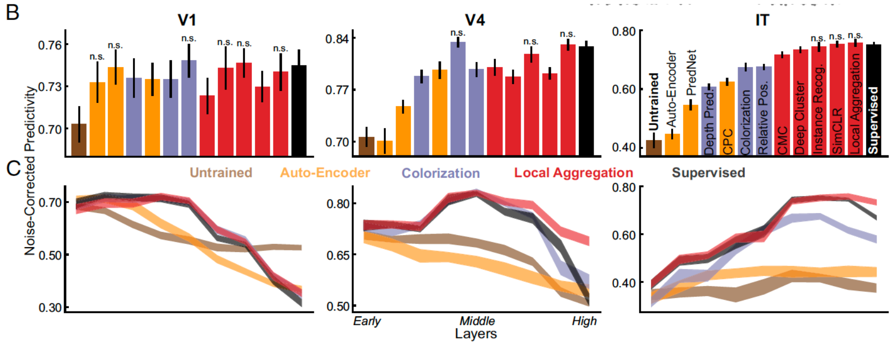
Task information consistency

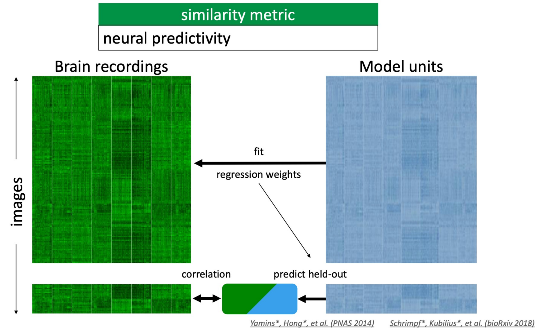
Single-unit response predictivity

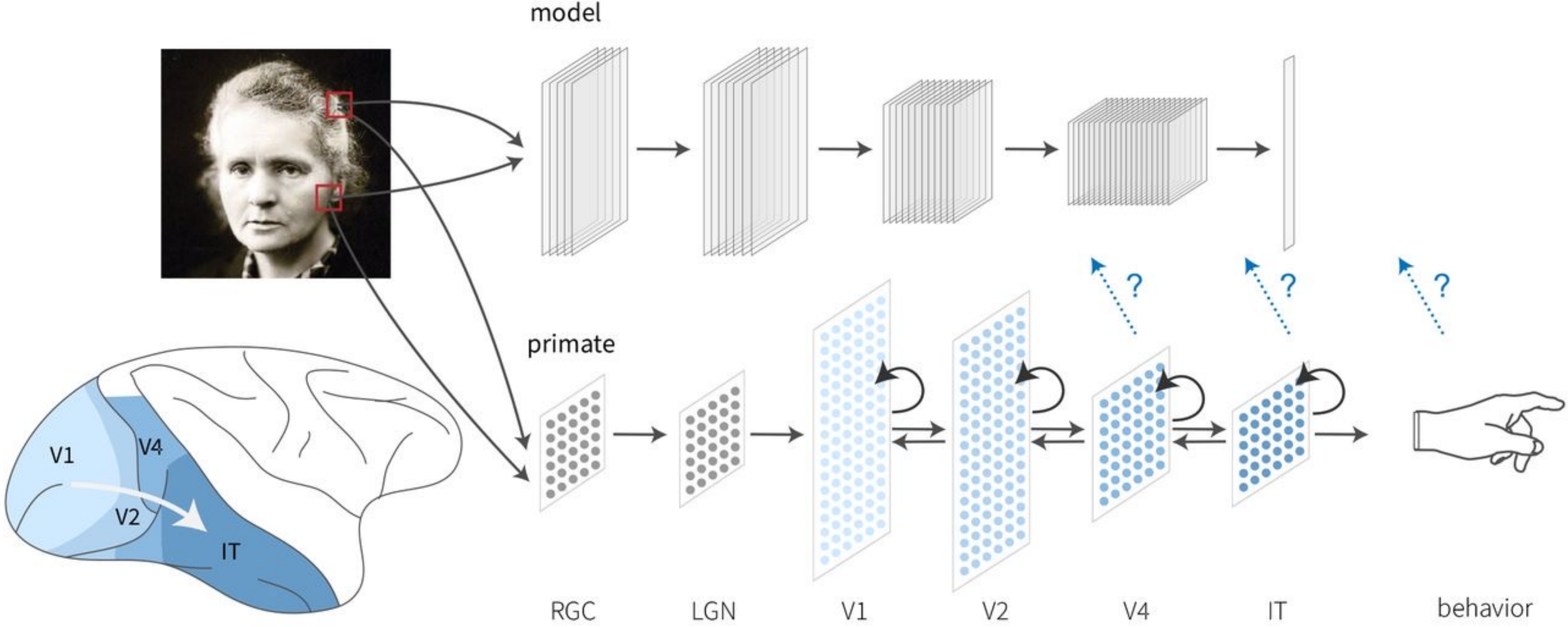
Population representational similarity



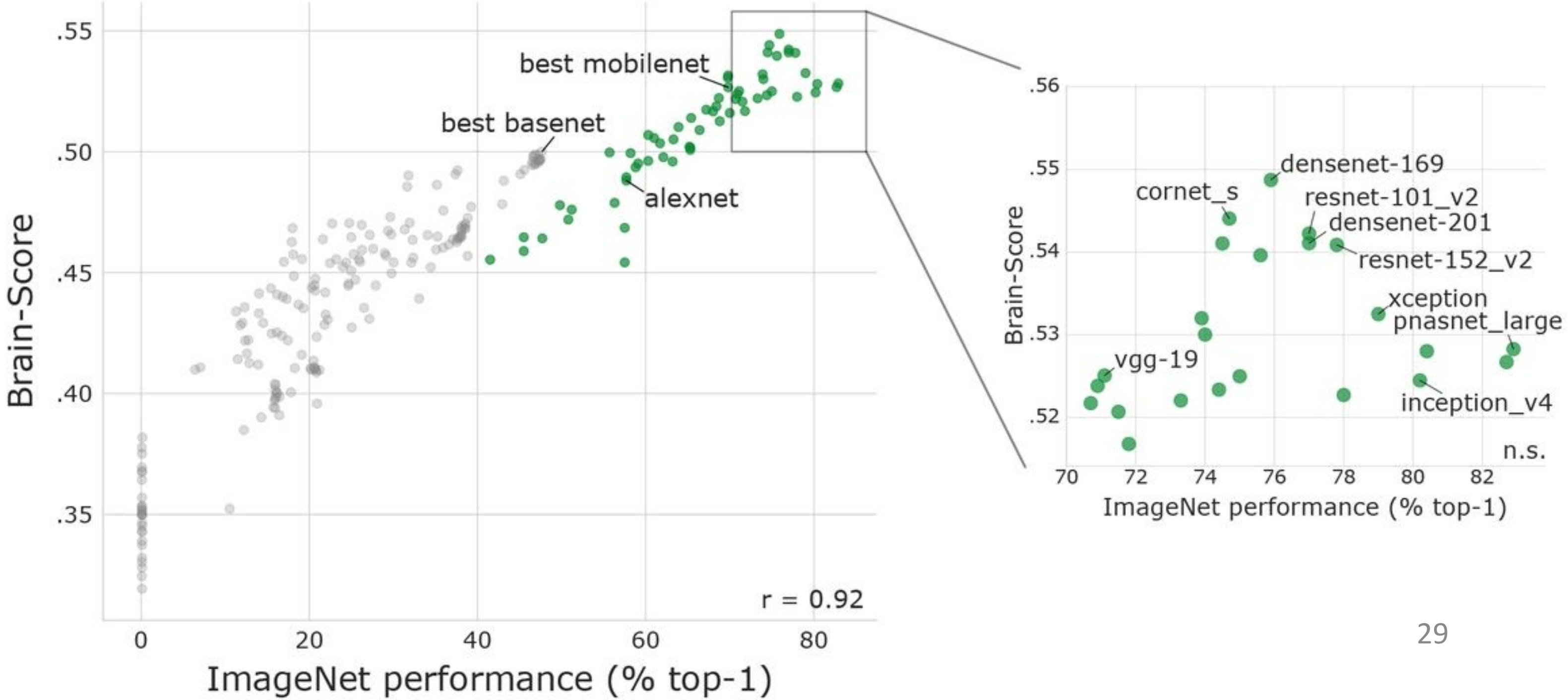
Yamins and Di Carlo, Nat Neuro 2016





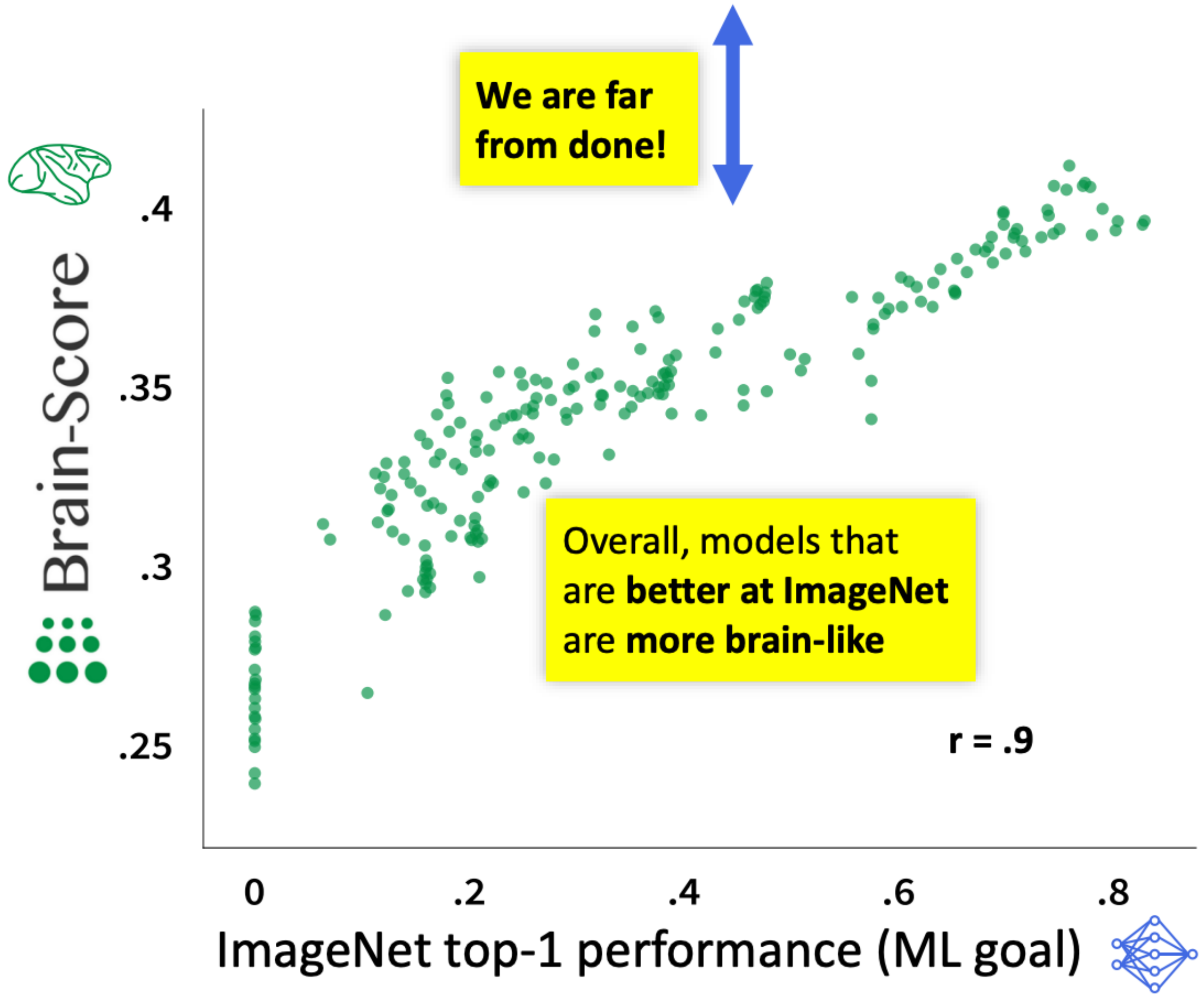


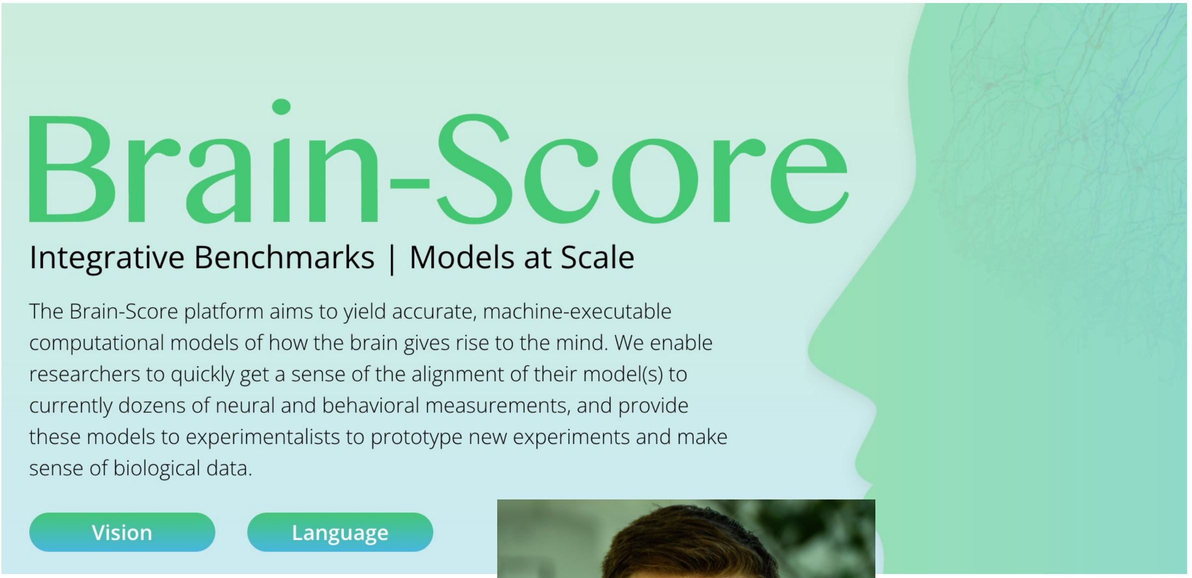
The hunt continues ...



Schrimpf et al. Neuron 2020

The hunt continues … integrative benchmarking across datasets



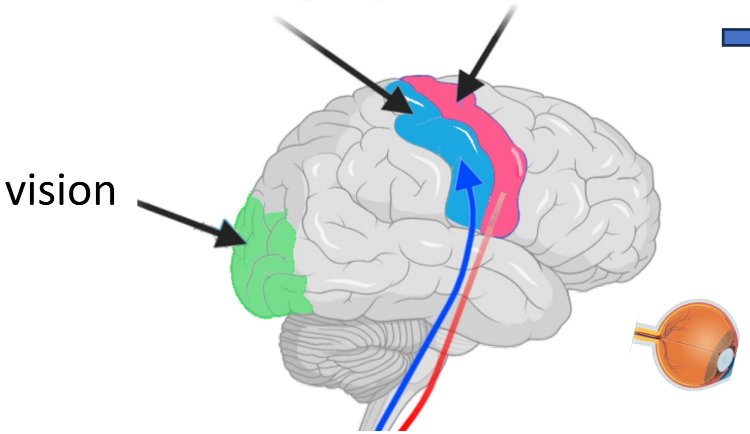


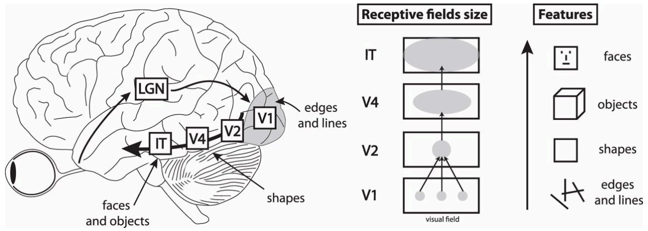
https://www.epfl.ch/labs/schrimpflab/

What other tasks? What other stimuli is the brain (visual) encoding?

Biological: recordings in visual system

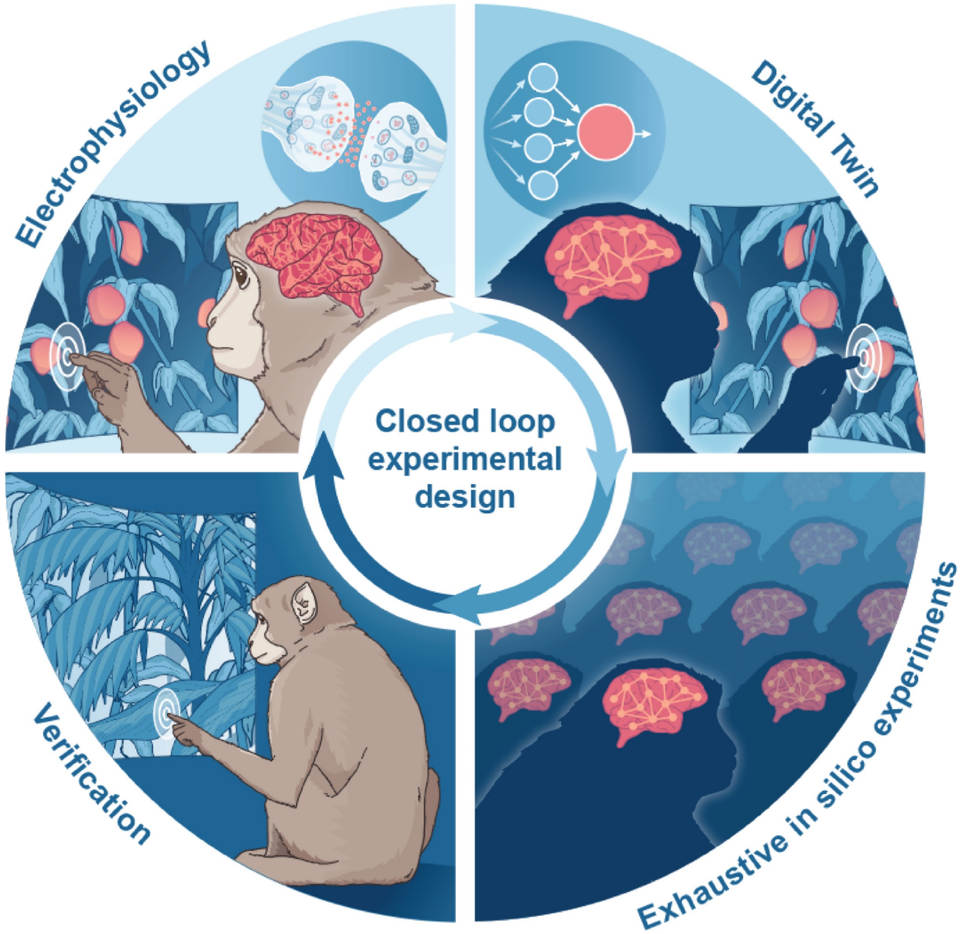
Building "digital twins", NN models of the system



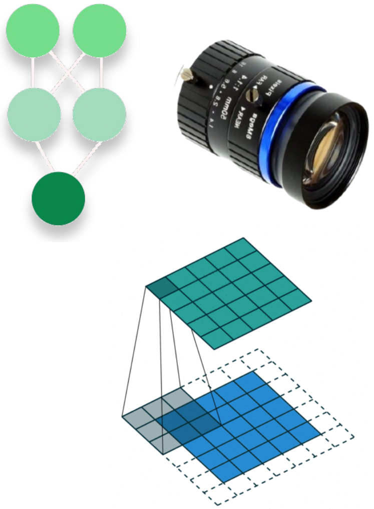


Hubel & Wiesel discoveries in cat V1 inspired convolutional neural networks

We now now a lot more (faces, motion, value coding) but we never can give enough stimuli ….
What would the ideal stimulus be for a given neuron?



Mathis, Perez Rotondo, Chang, Tolias, Mathis (unpublished)



Can we use our NN to produce predictions of optimal stimuli? Can this help reveal anew computational principle, or validate a discovered rule?

Section paper: developing deep predictive models for causal testing!

nature neuroscience

Explore content ⌄ About the journal ⌄ Publish with us ⌄

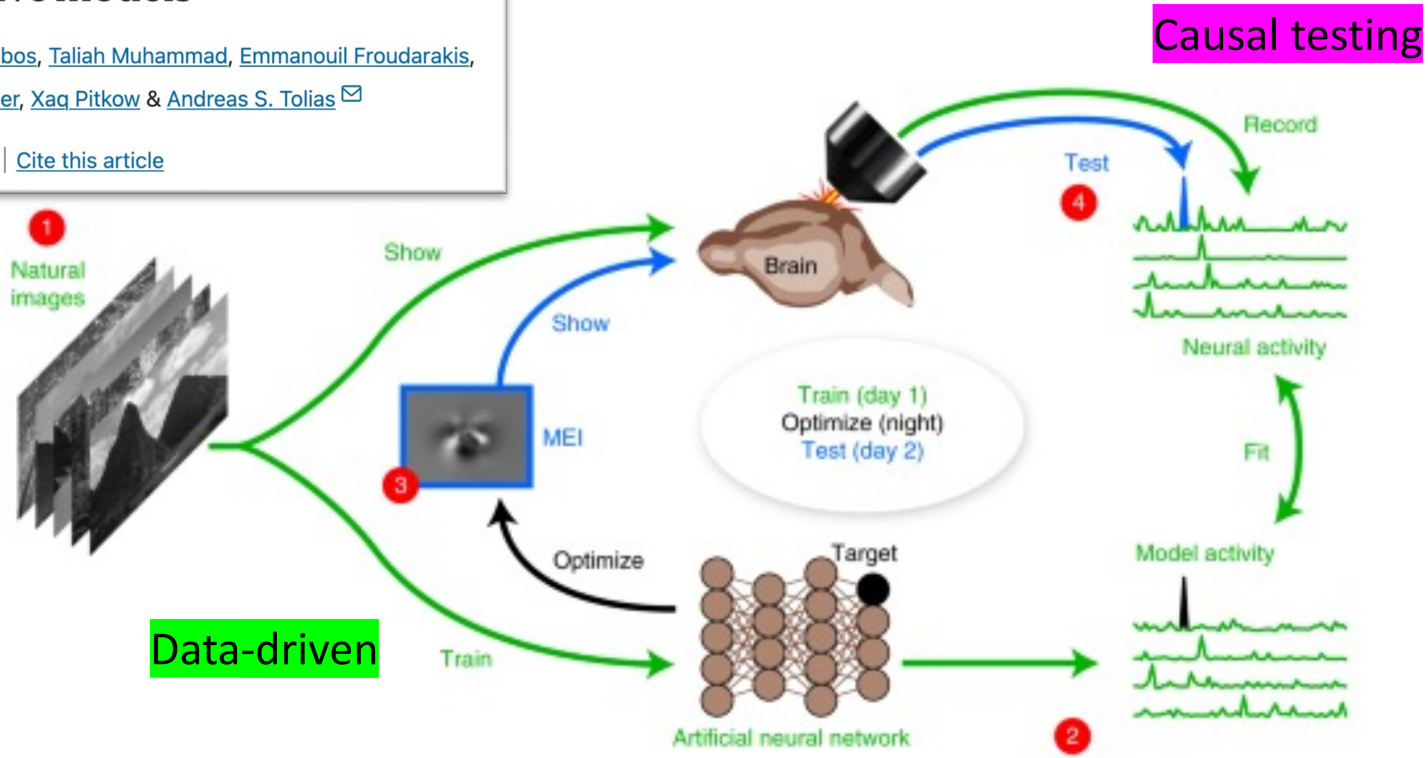
nature › nature neuroscience › articles › article

Article | Published: 04 November 2019

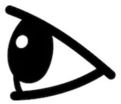
Inception loops discover what excites neurons most using deep predictive models

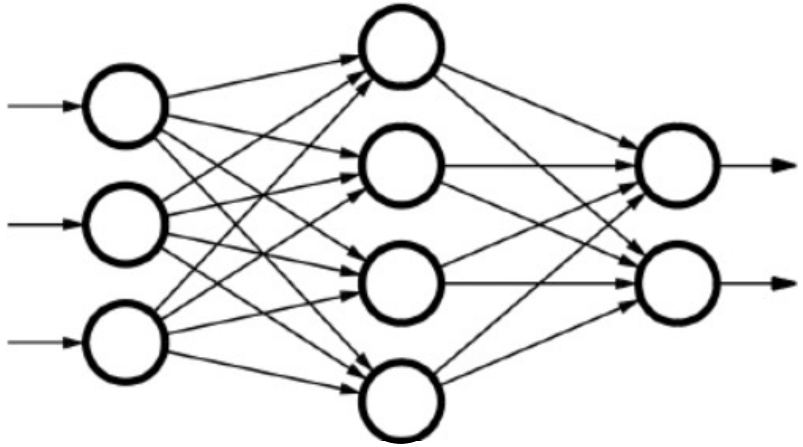
Edgar Y. Walker, Fabian H. Sinz, Erick Cobos, Taliah Muhammad, Emmanouil Froudarakis, Paul G. Fahey, Alexander S. Ecker, Jacob Reimer, Xaq Pitkow & Andreas S. Tolias

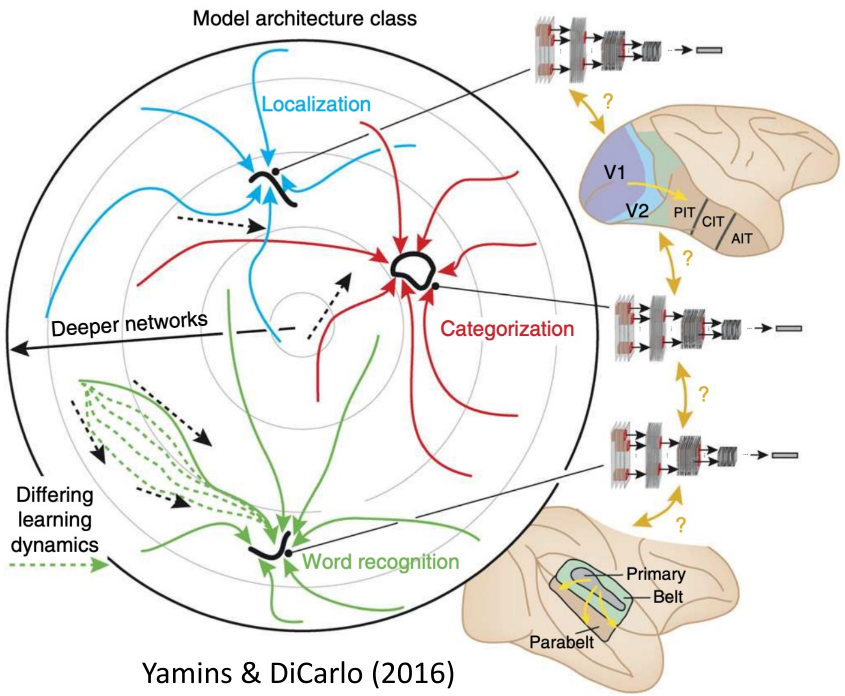
Nature Neuroscience **22**, 2060–2065 (2019) | Cite this article



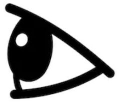
Using deep neural networks as task-driven models of a system





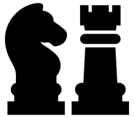


Yamins & DiCarlo (2016)


Vision: Yamins et al. (2014)


Audition: Kell et al. (2018)- speech recognition, speaker identification, natural sound identification

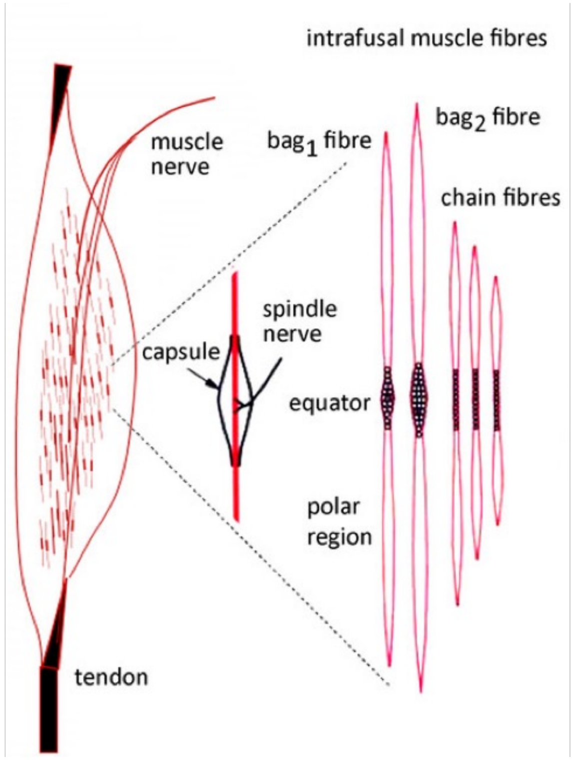

Barrel Cortex: Zhuang et al. (2017)

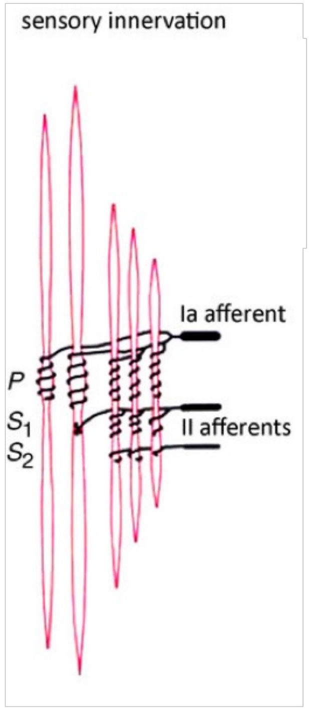

Cognition: Mante et al. (2013)


Proprioception: Sandbrink et al. (2023), Marin Vargas*, Bisi* et al. (2024)

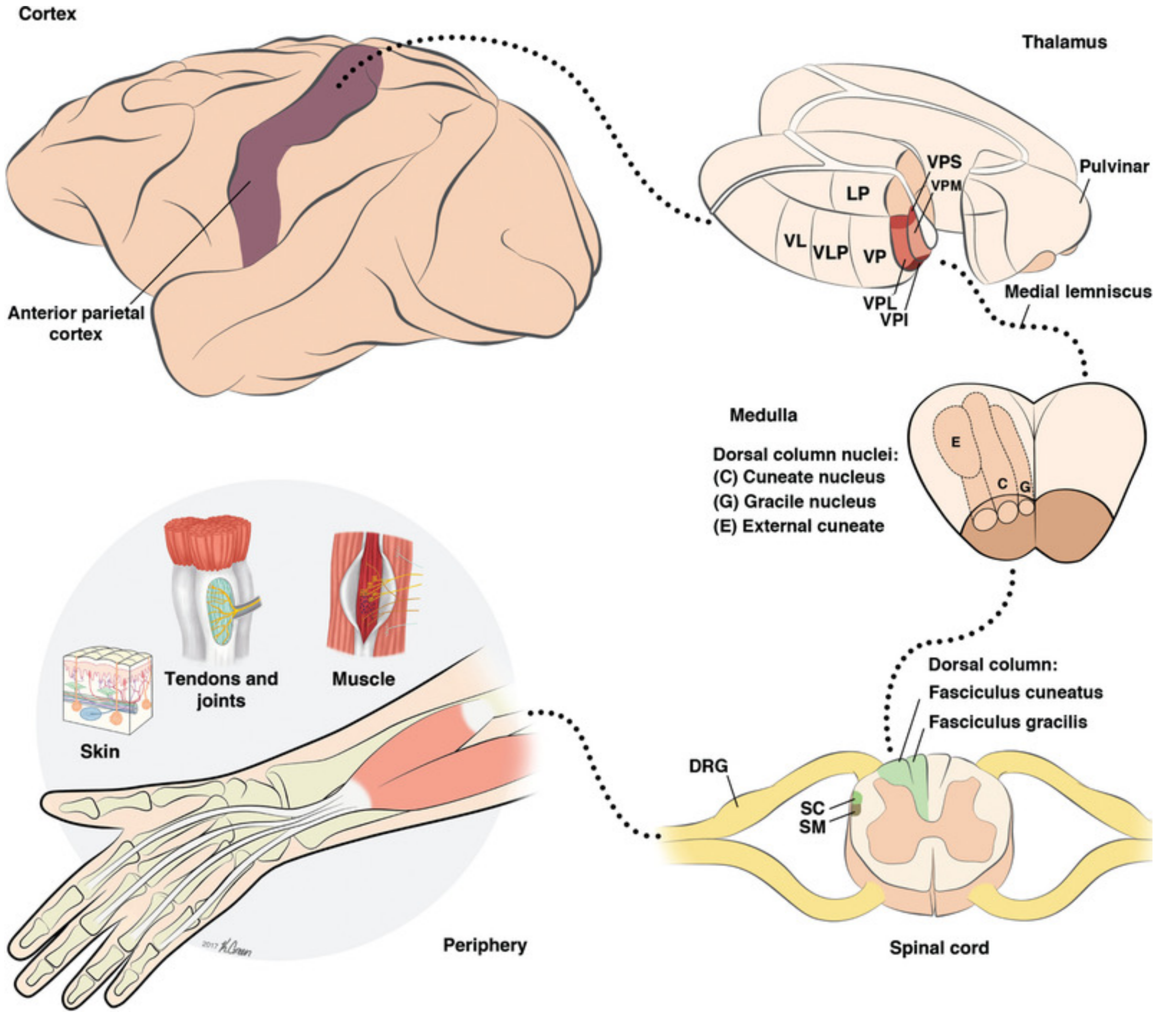
Proprioception (the sense of posture)

Muscle spindle



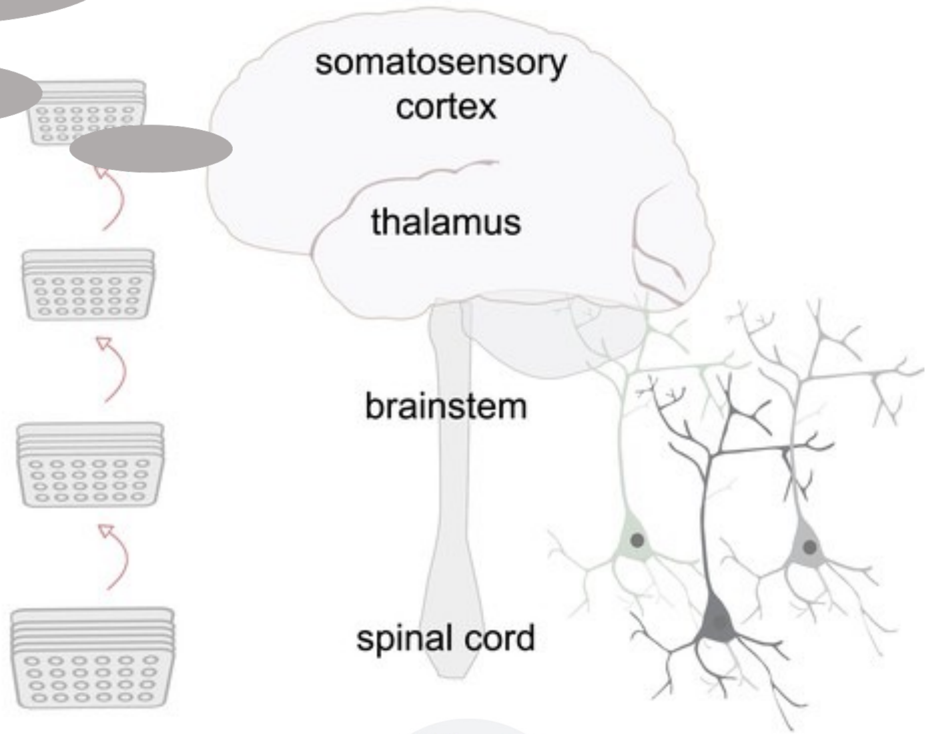


Banks 2020

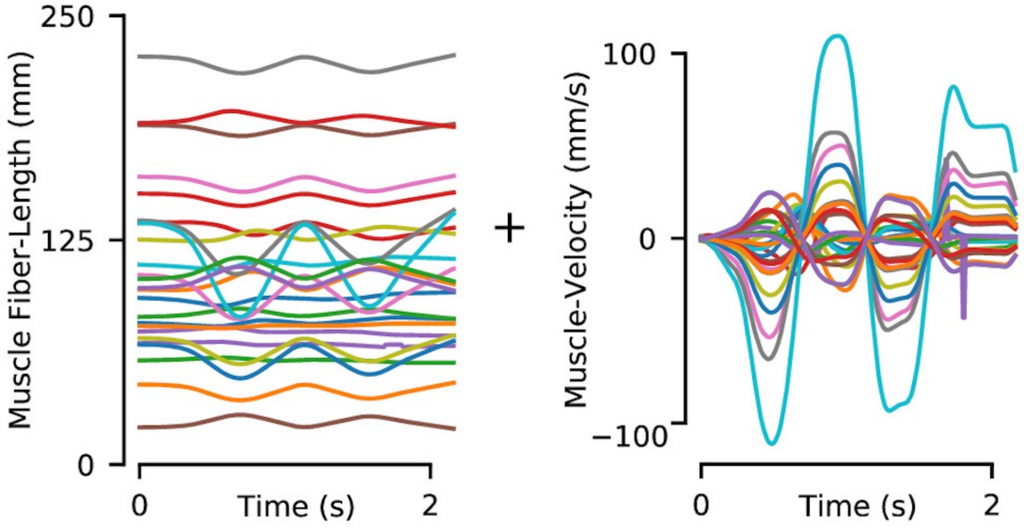


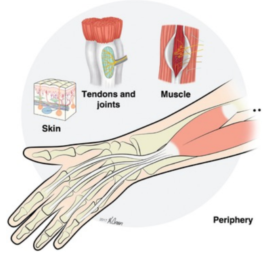
Delhaye et al., 2018

What's the trajectory of my limbs?



What is the integrative logic of proprioception?



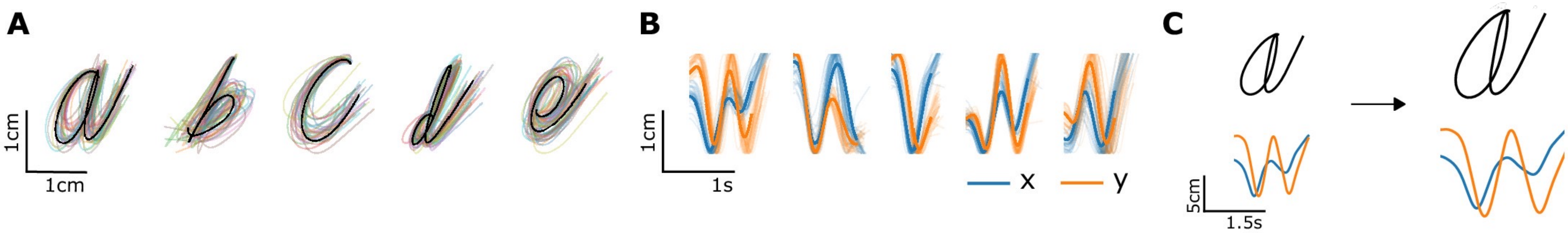


Sandbrink et al. 2023

Delhaye et al., 2018

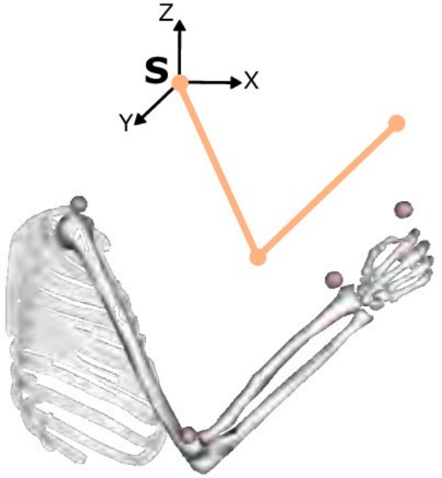
Slide courtesy of A. Mathis

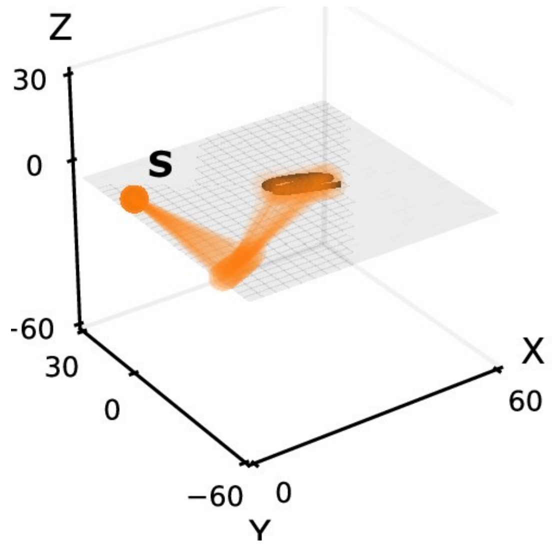
Task development: arm movements in 3D space

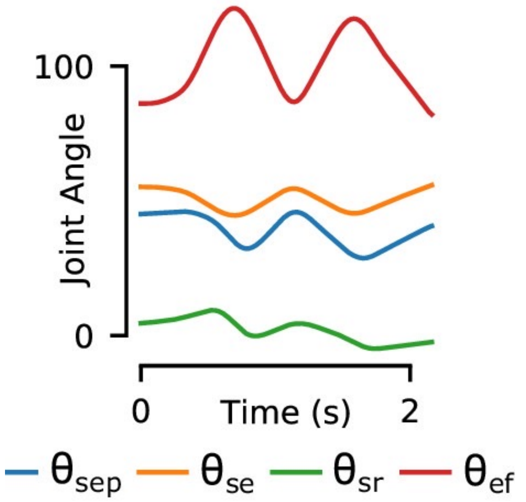


Human pen-tip movements writing Latin alphabet

Generate synthetic arm trajectories to re-create movement in 3D space



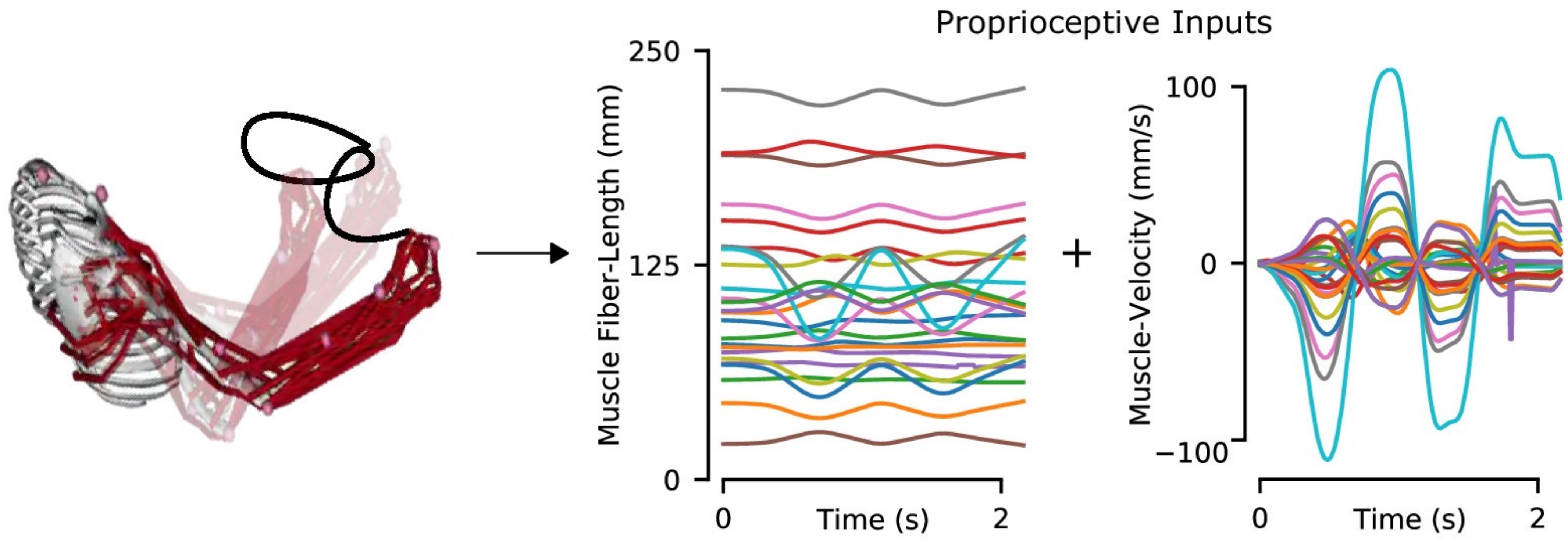




Sandbrink, Mamidanna et al. 2020, 2022 biorxiv / 2023 eLife

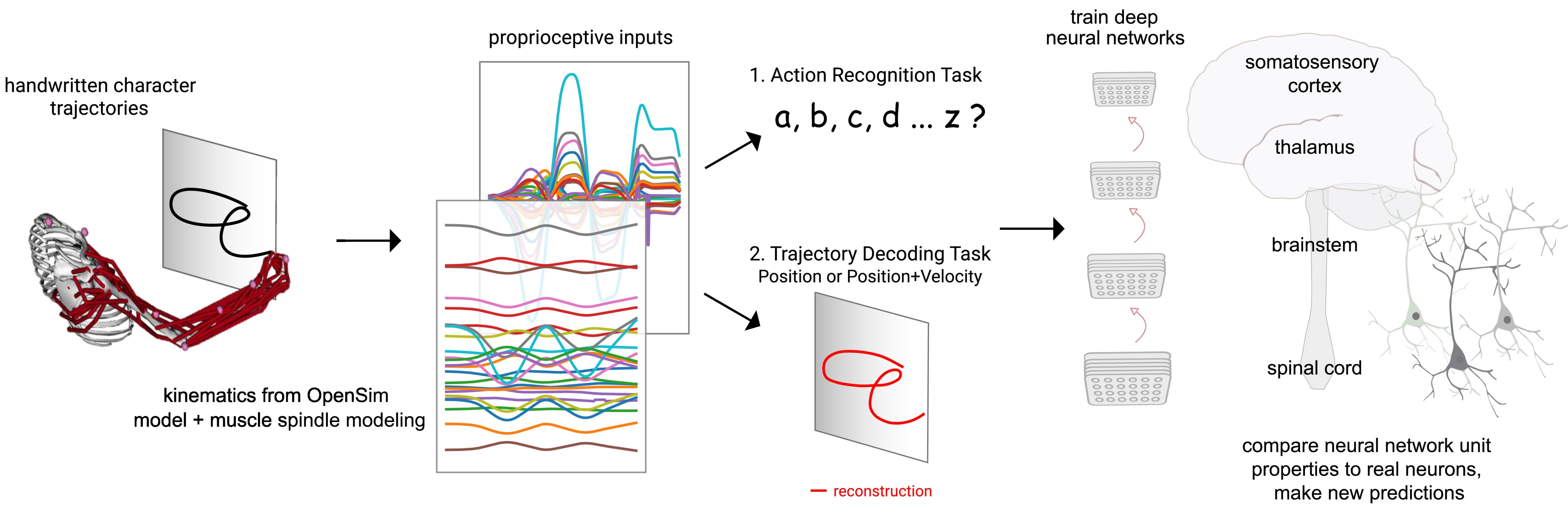
Derive muscle spindle activity for 1M trajectories

-



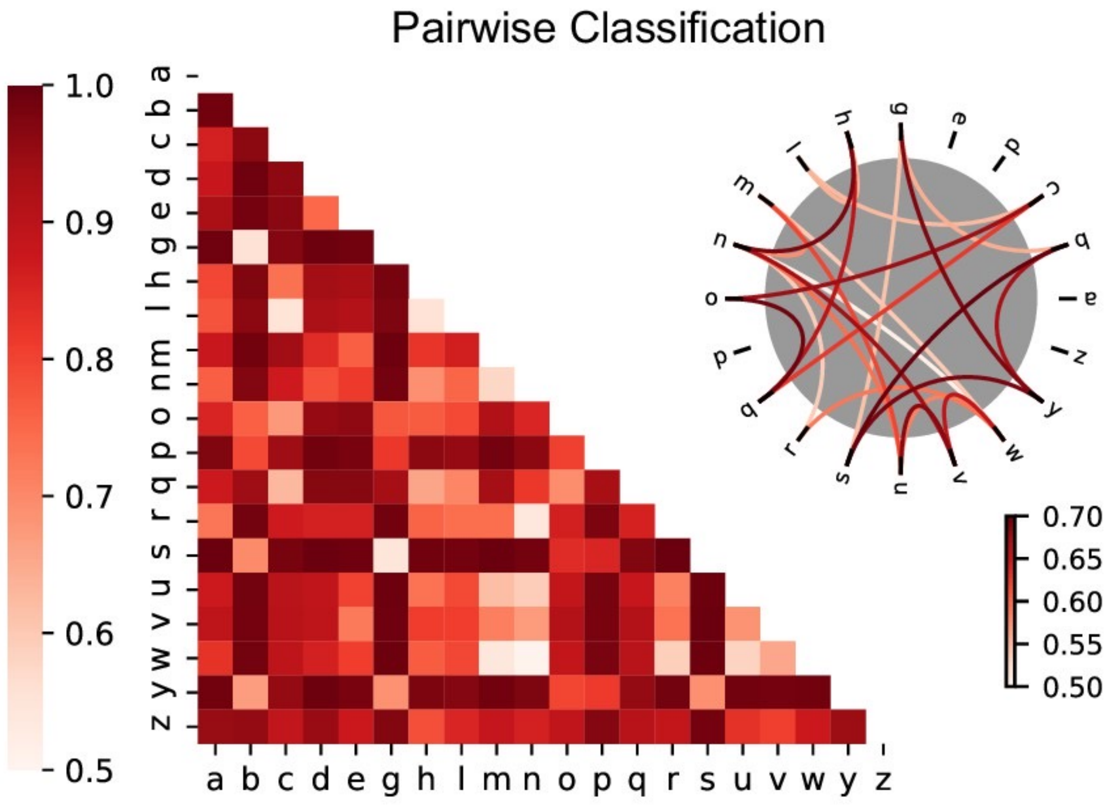
Sandbrink, Mamidanna et al. 2020, 2022 biorxiv / 2023 eLife

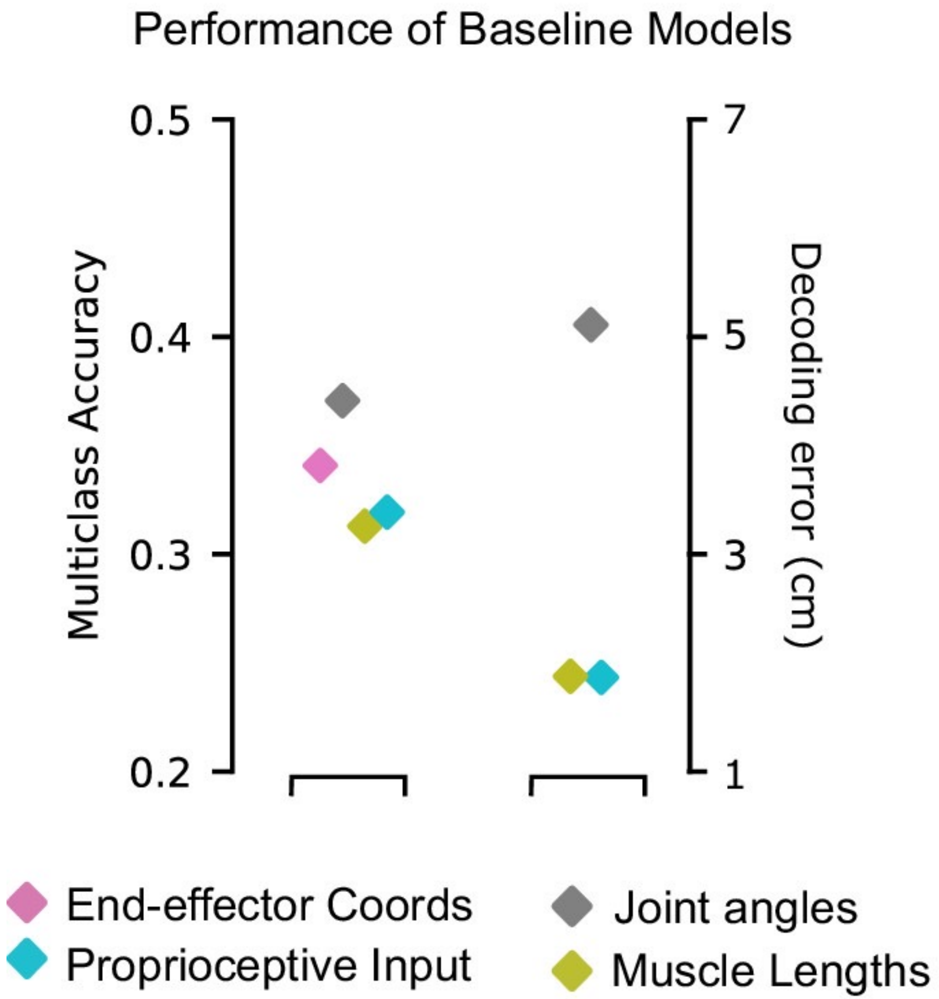
Task-driven modelling for the sensori(motor) system



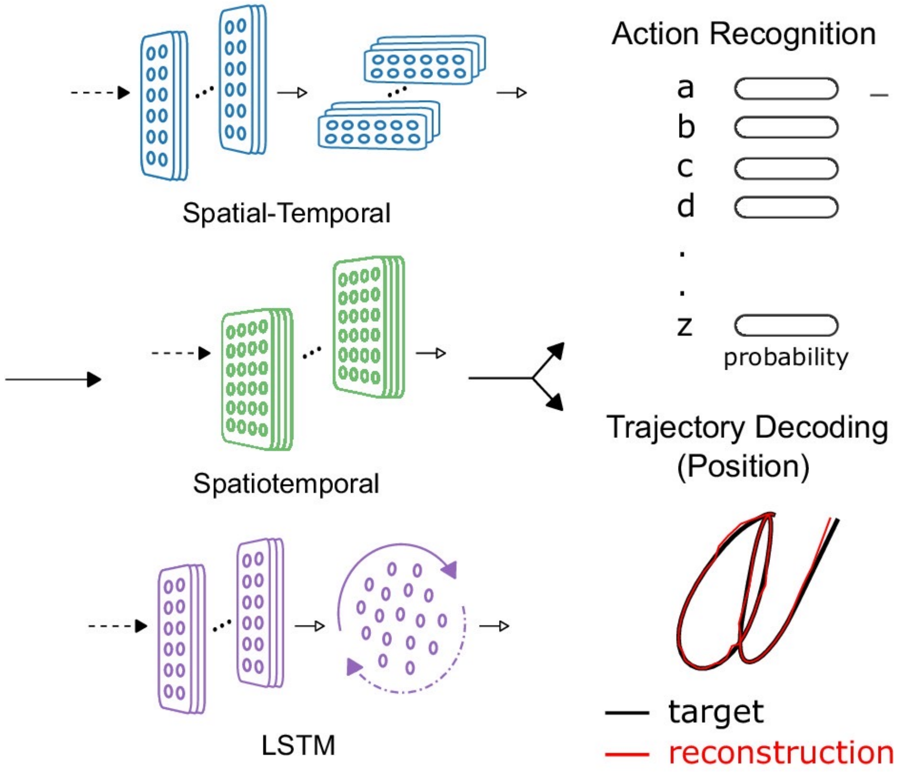
Sandbrink, Mamidanna et al. 2020, 2022 biorxiv / 2023 eLife

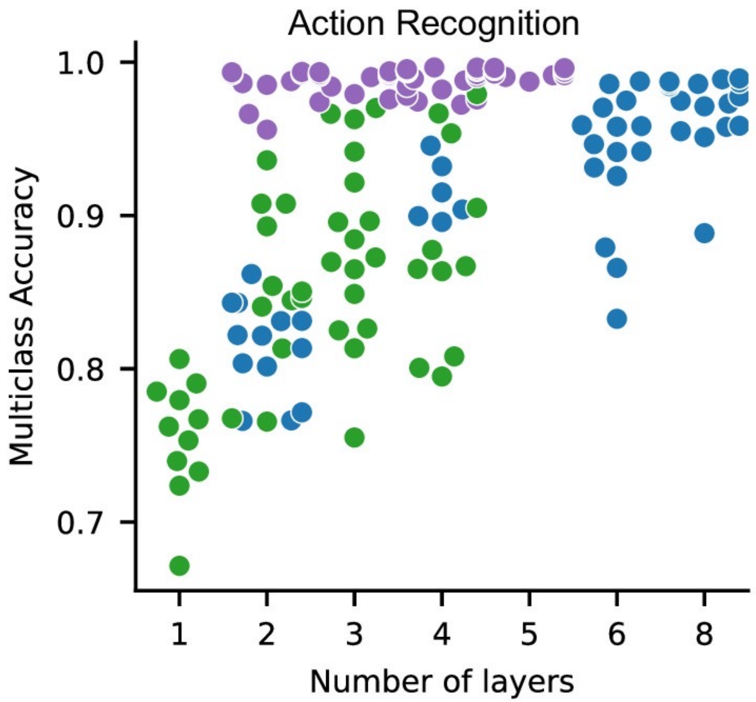
multi-class SVMs trained on proprioceptive inputs show action recognition is a hard task

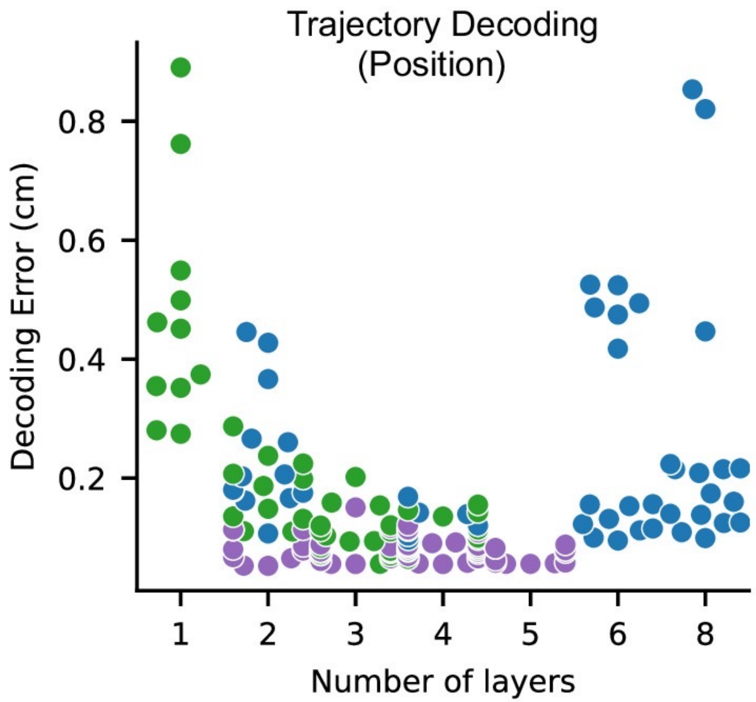




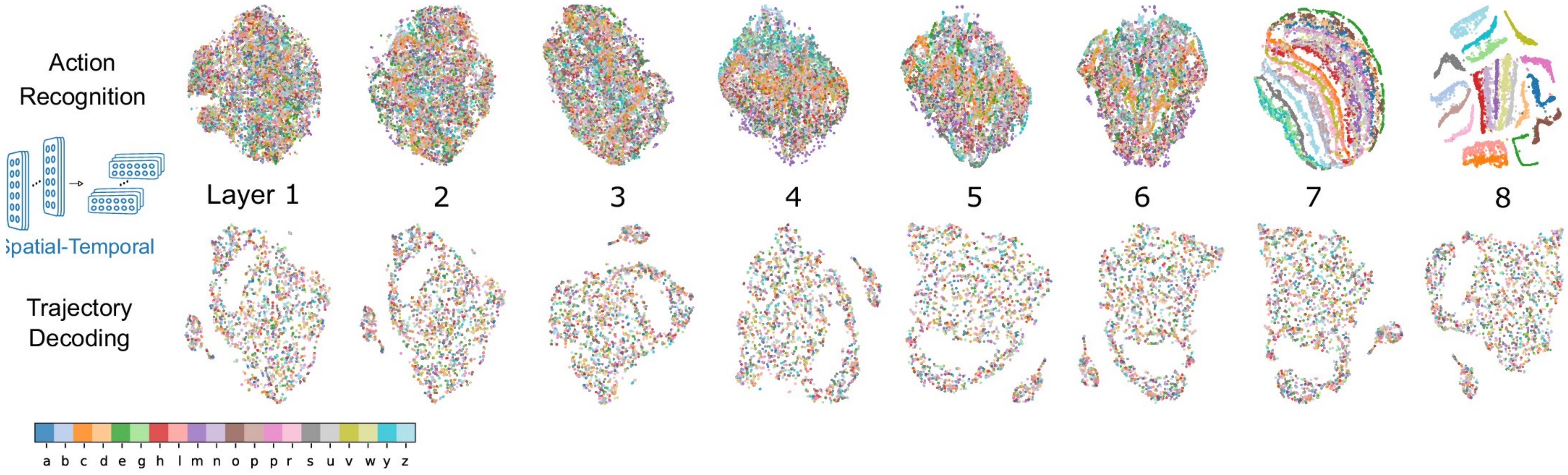
Neural networks can readily solve both action recognition and trajectory decoding





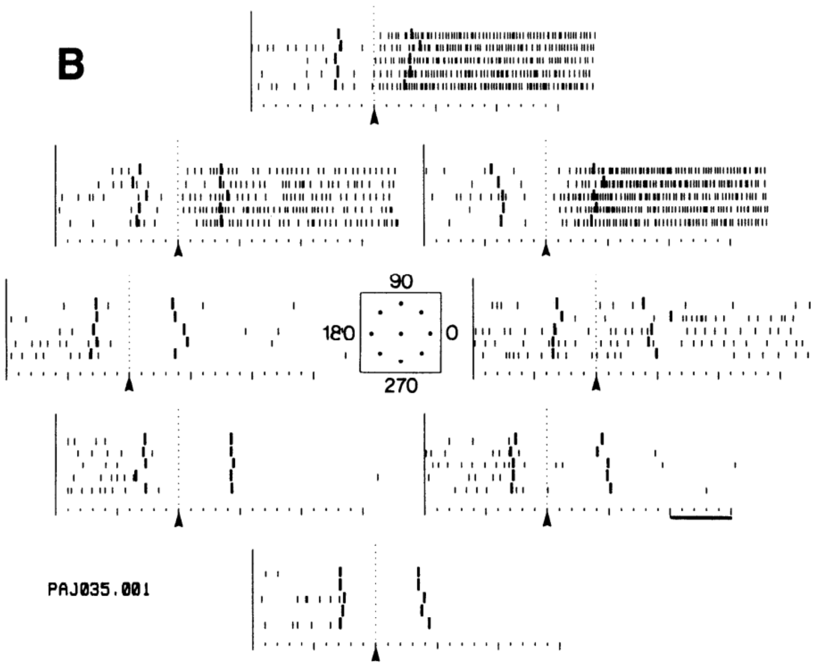


Different representations “emerge” in NNs trained on action recognition or trajectory decoding

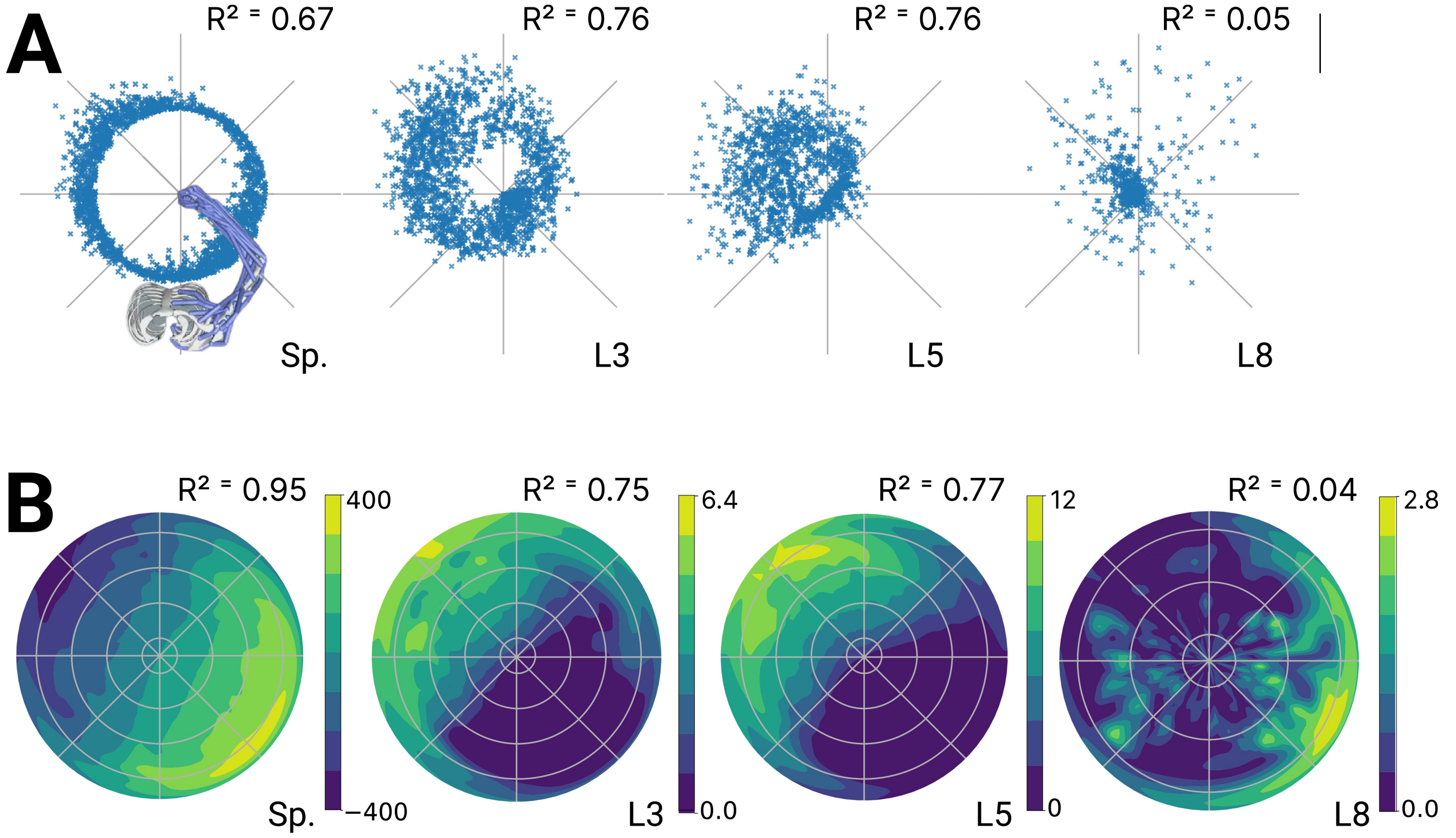


- Here, we used tSNE (nonlinear method) to look at changes in Label clustering across NN layers

Tuning curve analysis of NN units show canonical S1-like tuning



Prud'Homme and Kalaska (1994)

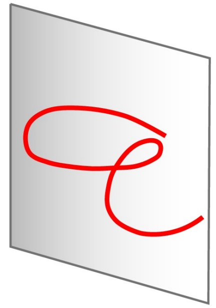


ART-trained models contain direction selective units, but TDT-trained do not ...

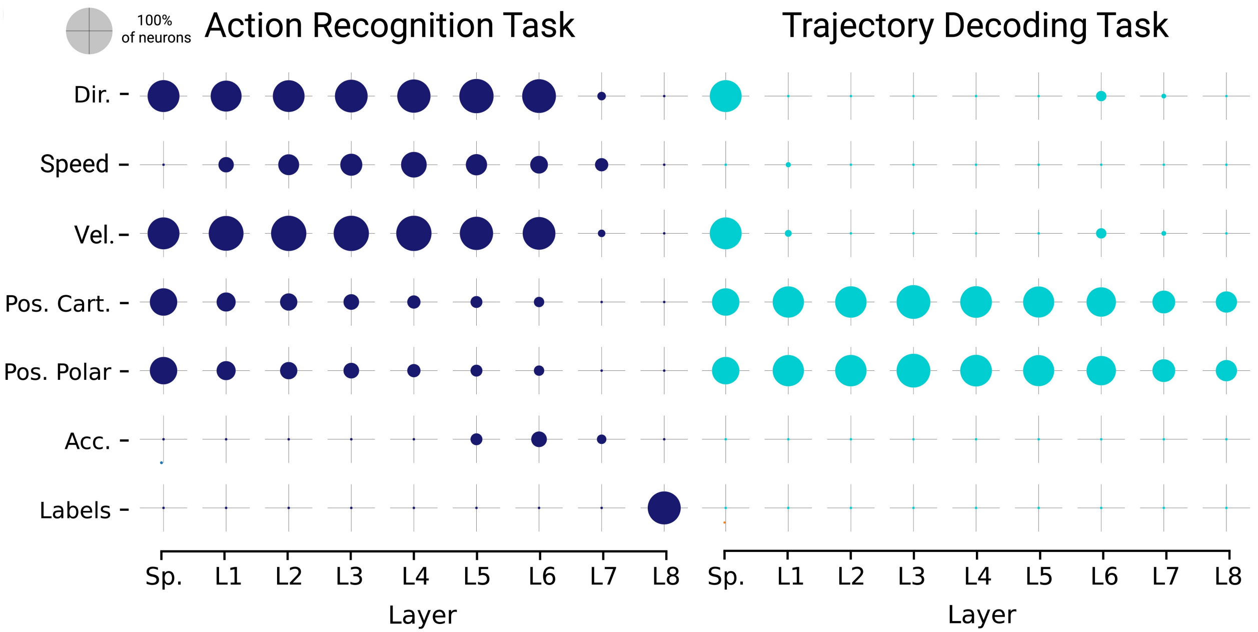
1. Action Recognition Task

a, b, c, d ... z ?

2. Trajectory Decoding Task
Position or Position+Velocity

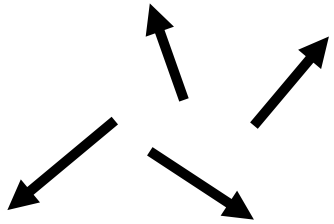


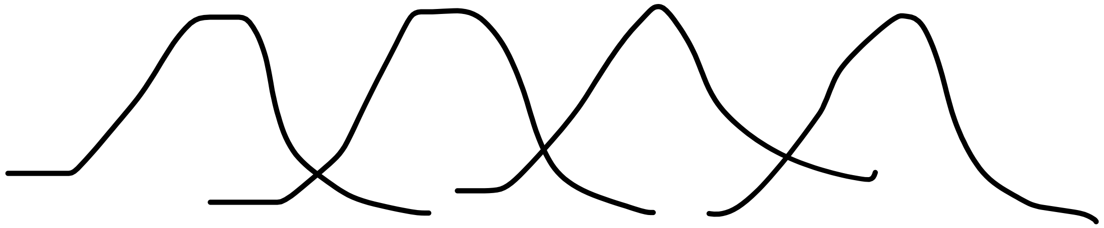
— reconstruction



Textbooks tells us:

- proprioception is there to tell us where our body is in space
- We should see direction selectivity & positional information …
- **If we train only on ART we see direction, *less* position; traj → position and no dir .. independent of ref. framework**



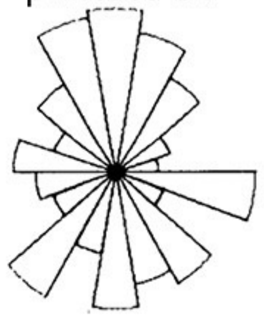


Population vector analysis of NN units show canonical S1-like tuning, in trained models

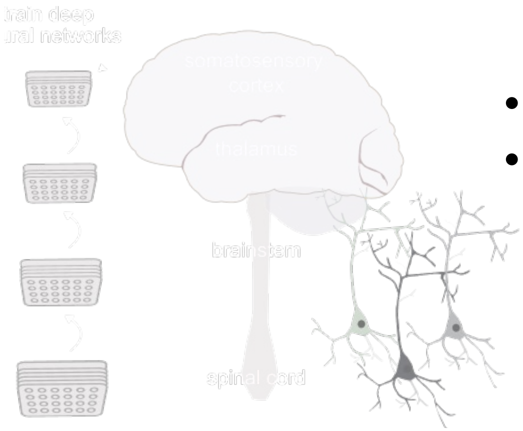
. Action Recognition Task

a, b, c, d … z ?

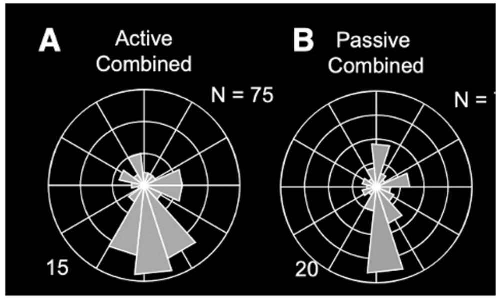
primate S1



Prud'Homme and Kalaska (1994)

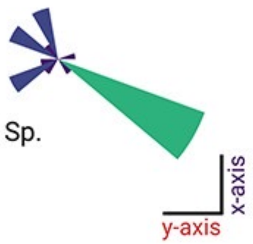


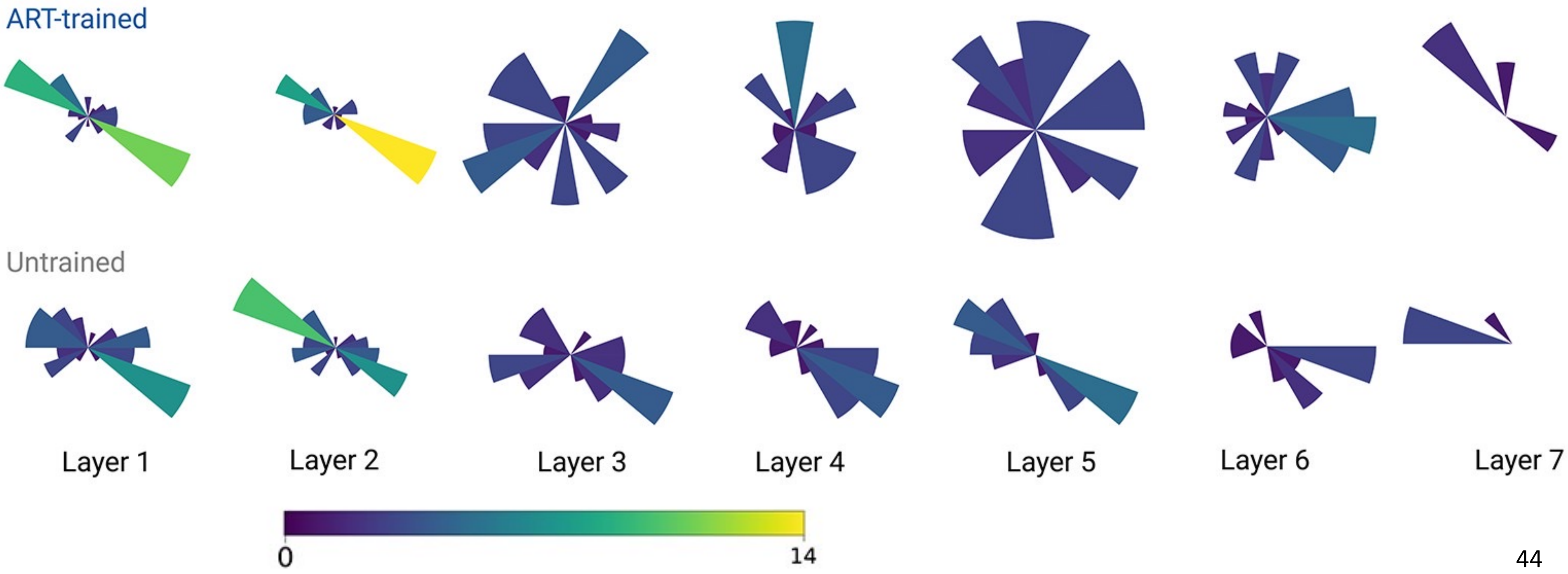
- **ART Layer 4-5 most S1 like**
- Hypothesized that:
    - cuneate nucleus closer to layer 1-2
    - S2 deepest layers (6+)



Versteeg et al., Journal of Neurophysiology (2021)

Spindles



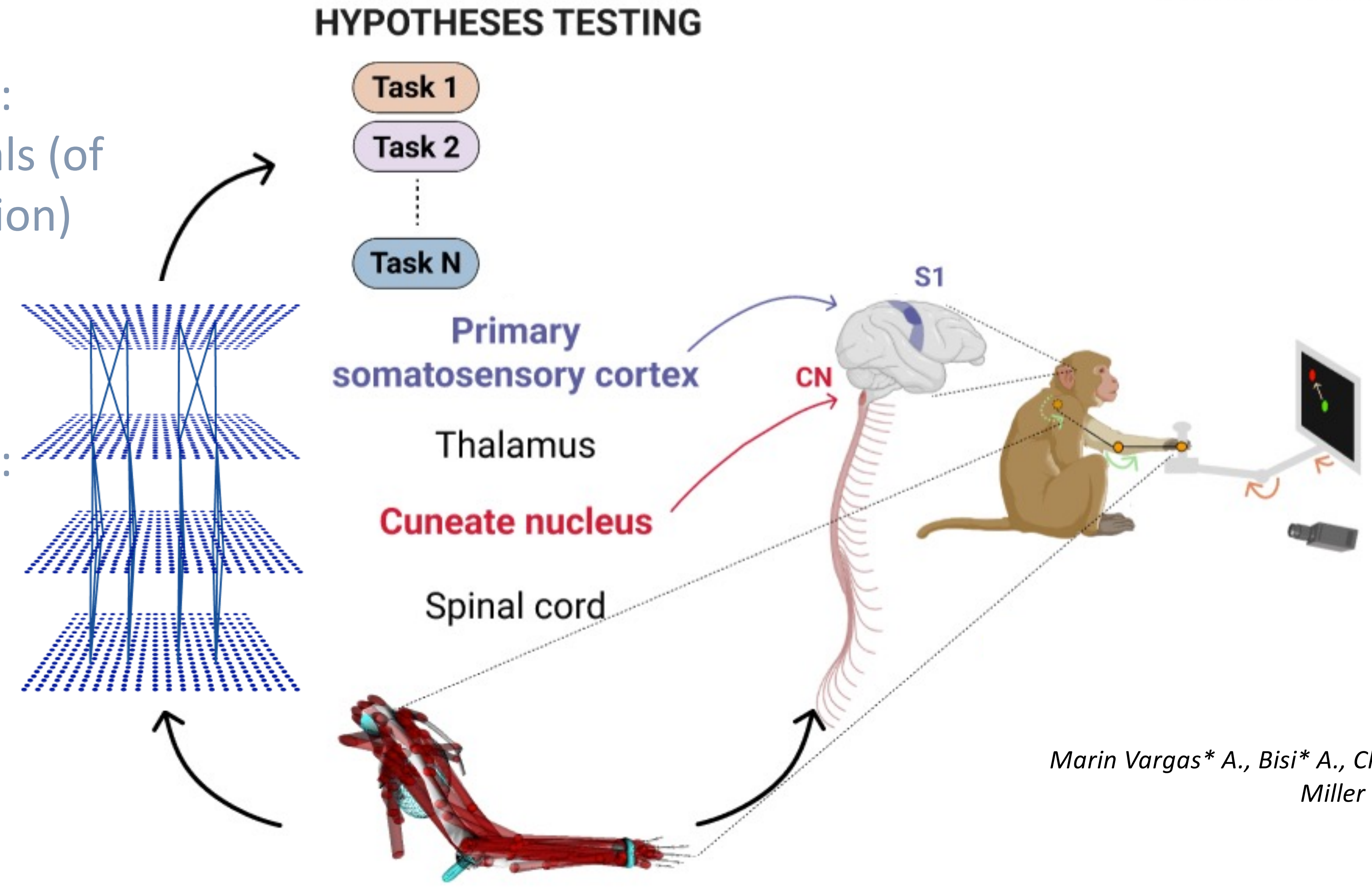


"Biomechanics only" show direction tuning…

Creating *even more* task-driven models of proprioception

Ingredient 1: Putative goals (of proprioception)

Ingredient 2: ANNs

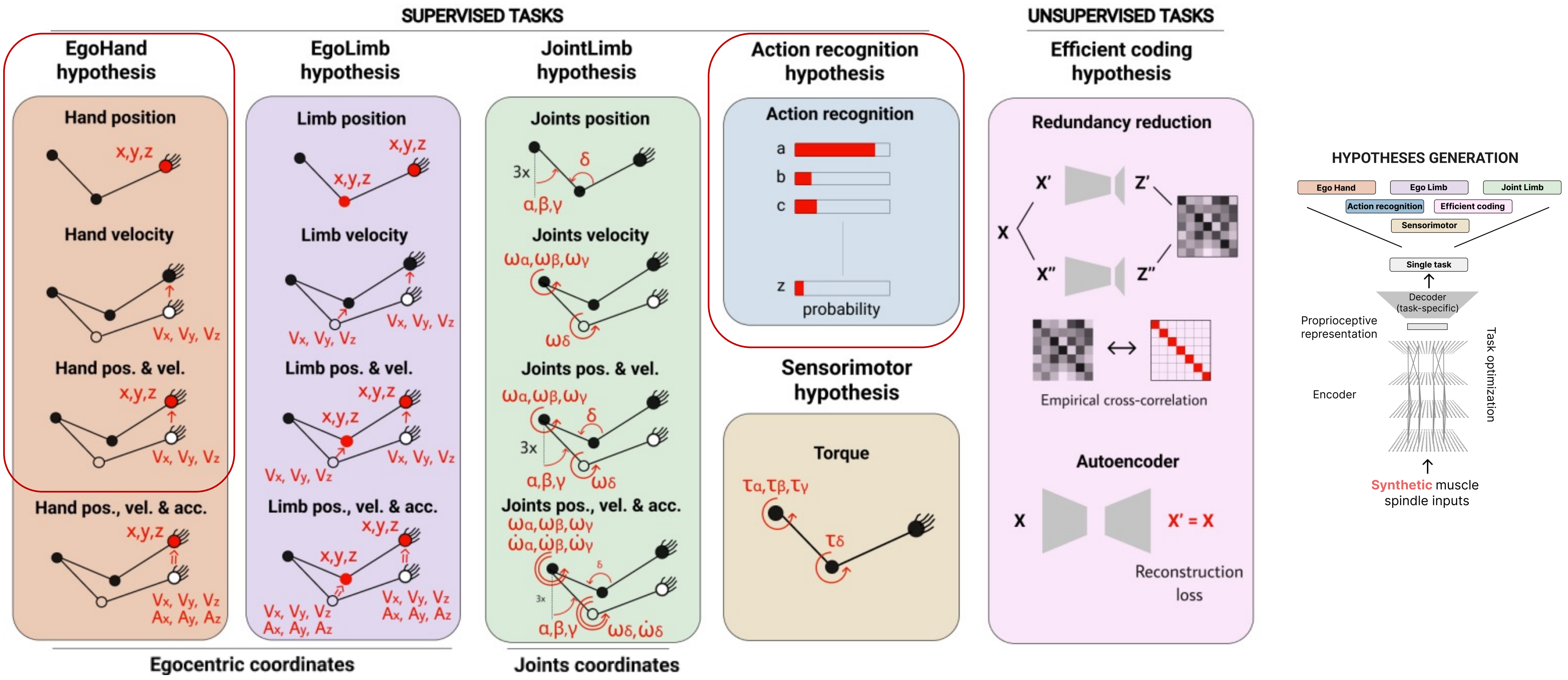


Marin Vargas A., Bisi* A., Chiappa A., Versteeg C., Miller L., Mathis A. Cell 2024*

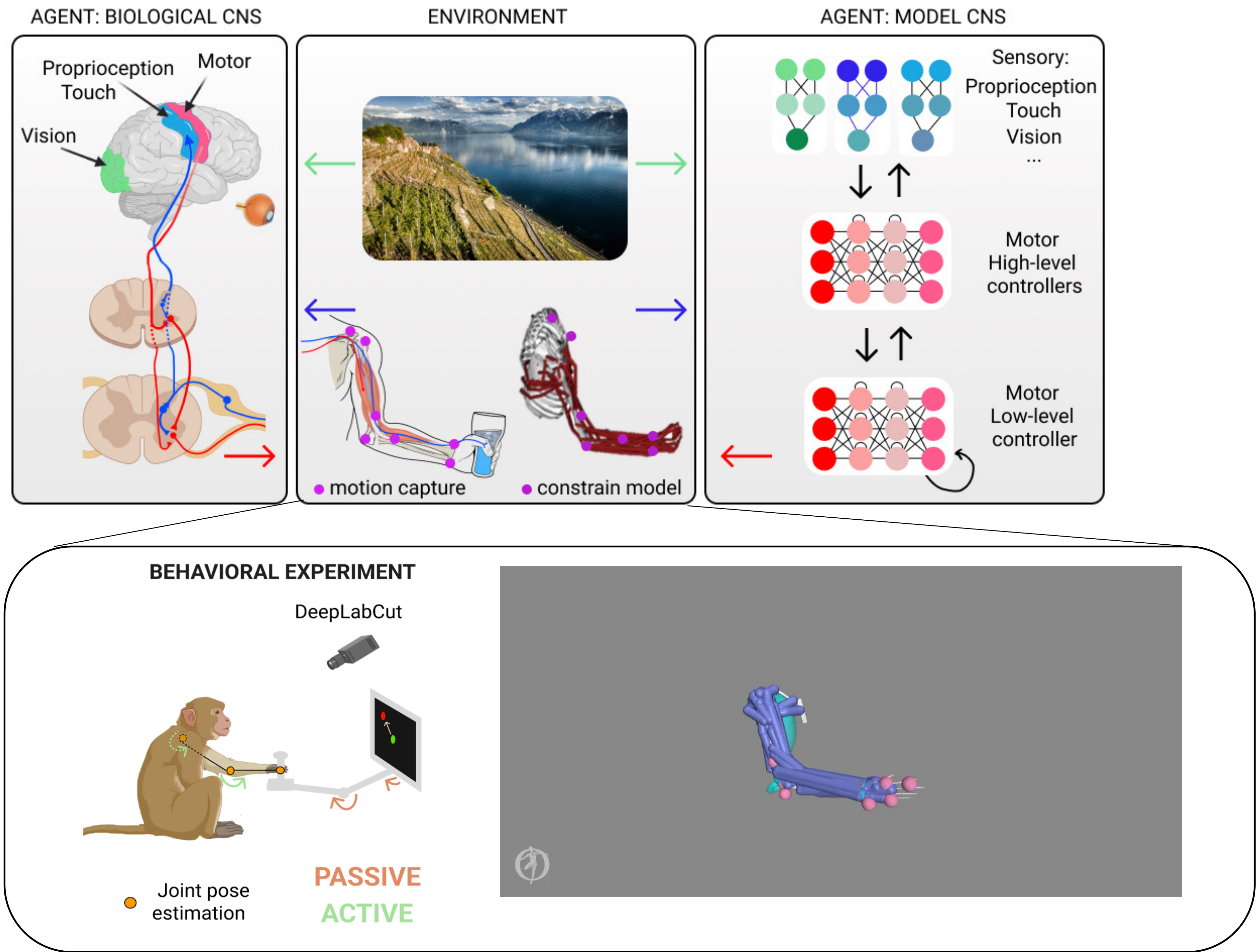
Ingredient 3: simulating spindle dynamics at scale

Slide courtesy of A. Mathis

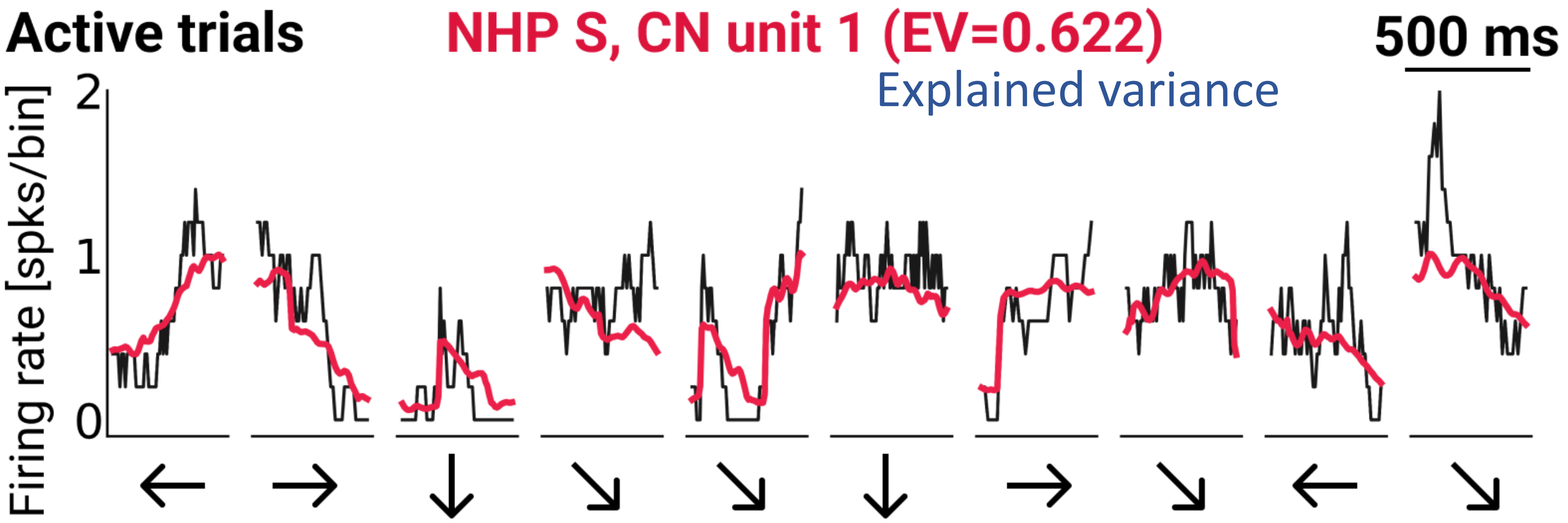
16 computational tasks to create candidate models

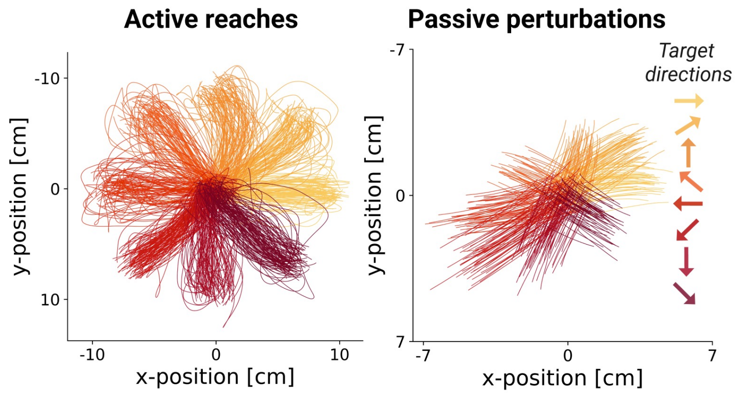


Marin Vargas A., Bisi* A., Chiappa A., Versteeg C., Miller L., Mathis A. Cell 2024*


AGENT: BIOLOGICAL CNS
Proprioception
Touch
Motor
Vision
ENVIRONMENT
motion capture
constrain model
AGENT: MODEL CNS
Sensory:
Proprioception
Touch
Vision
...
Motor
High-level
controllers
Motor
Low-level
controller
BEHAVIORAL EXPERIMENT
DeepLabCut
Joint pose
estimation
PASSIVE
ACTIVE

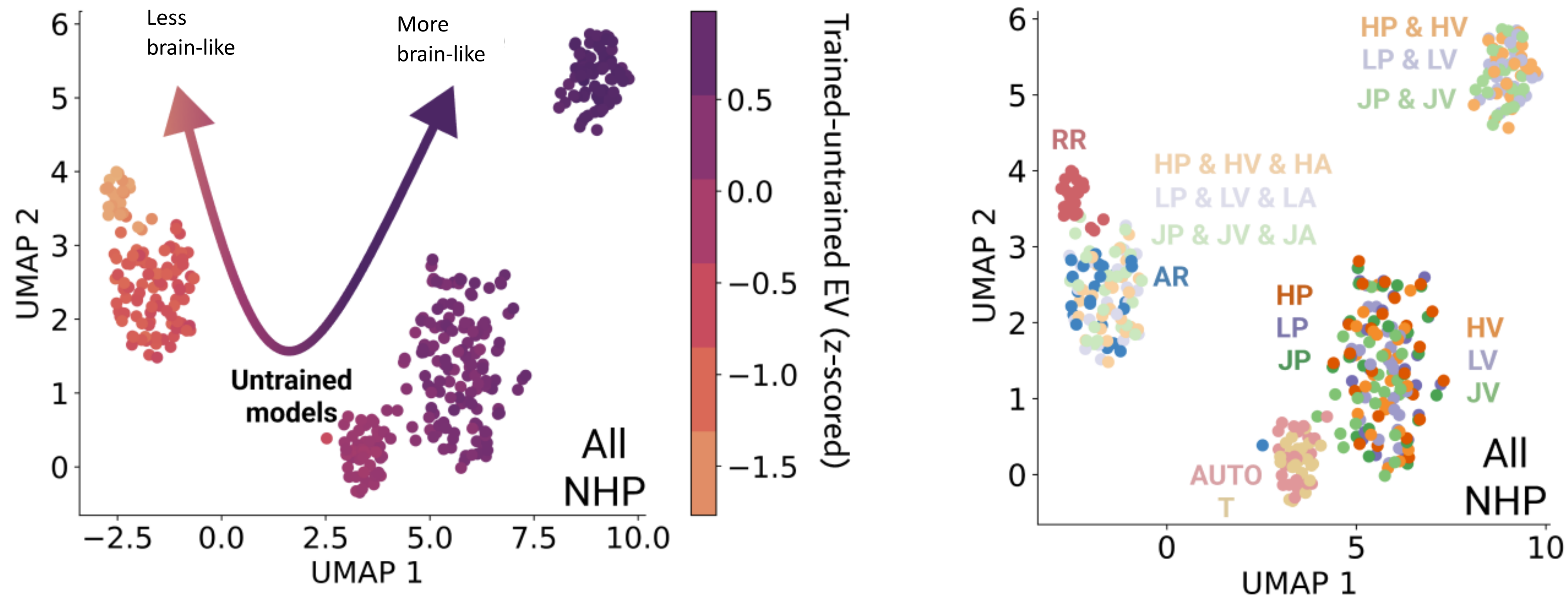
Task-trained models predict single neuron dynamics!





Comparing network representations with unsupervised UMAP clustering

ACTIVE



HP = Hand position
HV = Hand velocity
HA = Hand acceleration

LP = Limb position
LV = Limb velocity
LA = Limb acceleration

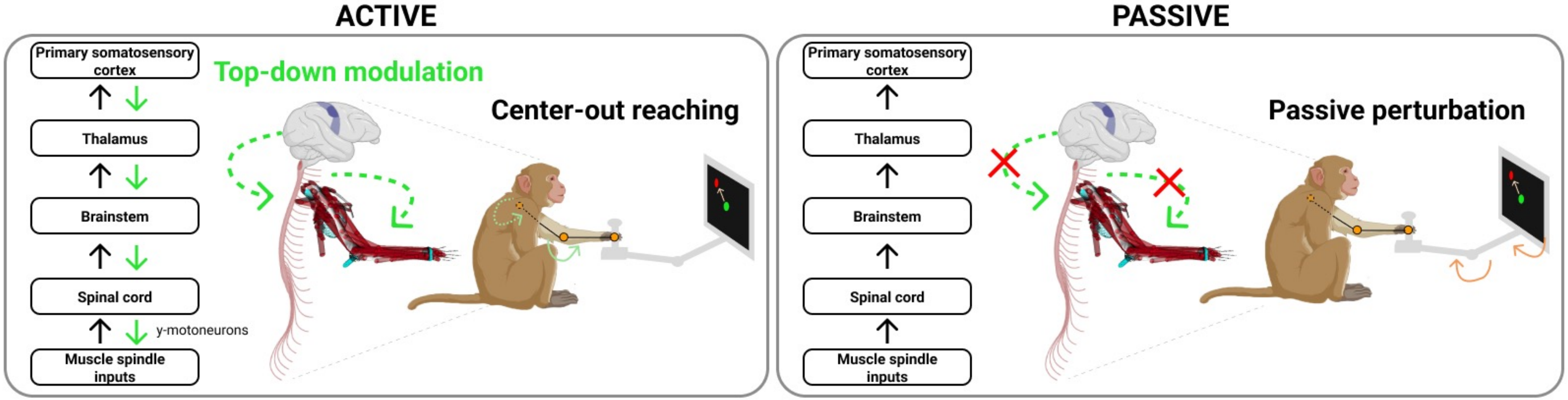
JP = Joint position
JM = Joint velocity
JA = Joint acceleration

RR = Redundancy Reduction
AUTO = Autoencoder

AR = Action Recognition
T = Torque

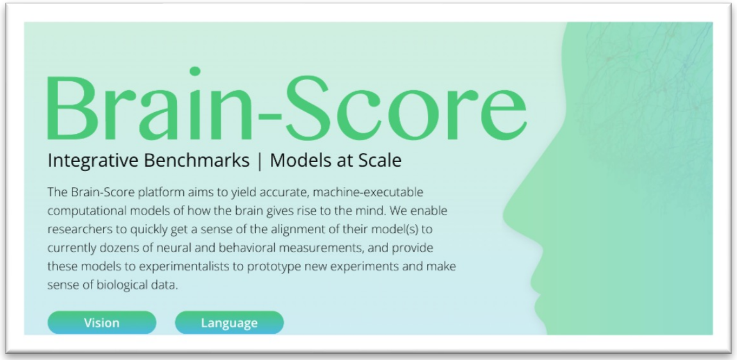
Task-trained models predict neuron dynamics!

- For all 16 hypotheses, if a *computation* is better learned on synthetic, passive spindle data, then the model also generalizes better to neural data
- Neural data (*in the active case*) is best explained by the **hypothesis that proprioception is optimized to encode the location and velocity of the body** (irrespective of coordinate framework (hand or limb)
- Lack of evidence for hierarchical processing; this suggests that proprioception even in the brain stem is dominated by efference copies …



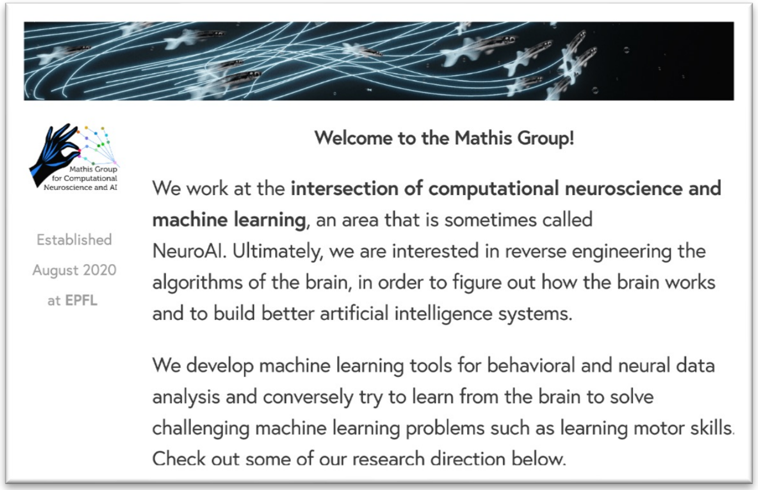
Slide courtesy of A. Mathis

neuroAI in labs at EPFL:



https://www.epfl.ch/labs/schrimpflab/

Prof. Dr. Martin Schrimpf



https://www.mathislab.org/

Prof. Dr. Alexander Mathis

https://www.mackenziemathislab.org/

Prof. Dr. Mackenzie Mathis

Summary

- NeuroAI is an emerging discipline that crosses across systems neuroscience and computer science
- It's goal is broadly to use neuro insights to build AI, and to develop AI for understanding the brain (neuro)
- It is needed as it is still VERY hard to develop embodied AI, human-like movement into robotics, and we still lack generally intelligent systems (although LLMs for language are impressive …)
- Key example in Neuroscience inspiring AI: convolutional neural networks (likely transformers too "attention"): this is a hot area in industry – using cognitive neuro approaches to study NN btw!
  - Interestingly, CNNs developed representations similar to the brain
- Key examples of AI influencing neuro: better behavioral analysis tools, better neural analysis tools (see also BCI week soon!)
- What is missing? NNs are very simple "neurons," that lack the complexity of what we find in the real brain: an opportunity awaits!
- Data-driven and task-driven modeling: key approaches in neuroAI
- How do we model sensory systems: examples in vision and proprioception
  - What to consider: both how close they are at single cell, task performance, and population level similarity
  - Ongoing efforts: Brain-Score, **Inception Loops …**