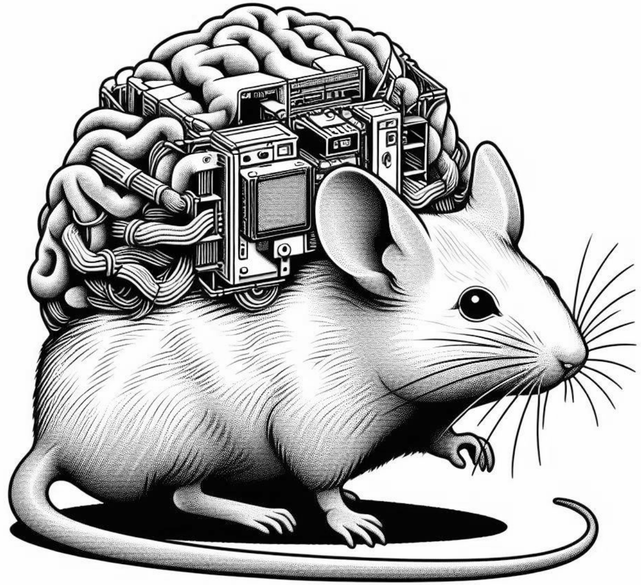
Foundation model of the mouse visual cortex

Bryan M. Li @bryalimy
Biomedical AI CDT, University of Edinburgh
Enrichment student, The Alan Turing Institute

Generated by Microsoft Image Creator



Journey

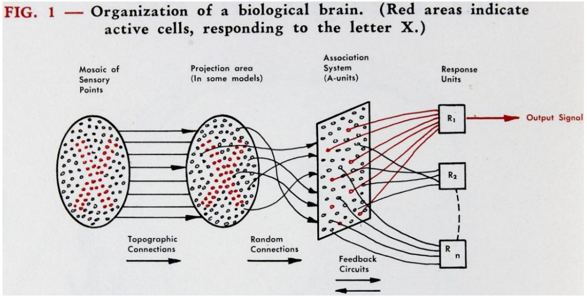
2023 - Present	Enrichment Scheme, The Alan Turing Institute
	- Contribute to the *autoemulate* digital twin project
2020 - Present	Biomedical AI PhD (CDT) at the University of Edinburgh
	- NeuroAI (machine learning and computational neuroscience)
	- Machine learning for healthcare
2019 - 2020	MSc by Research in AI at the University of Edinburgh
2018 - 2019	Research Engineer at Huawei Noah’s Ark Lab in Toronto
2017 - Present	FOR.ai (now Cohere For AI)
2016 - 2017	Software Engineer Intern at AMD (Embedded Team) in Toronto
2013 - 2018	BSc in Computer Science at the University of Toronto

Background

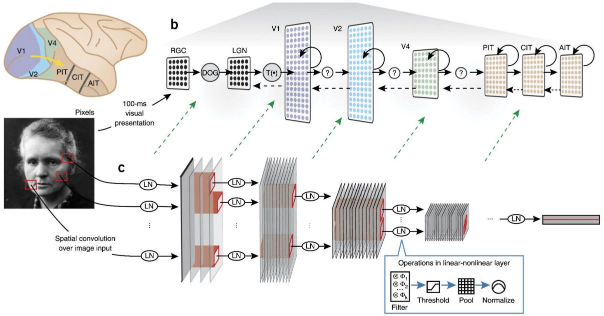
Machine learning and computational neuroscience

Neuroscience and machine learning

ML inspired by neuroscience



[1] Perceptron



[2] CNN and the ventral stream

ML application in neuroscience

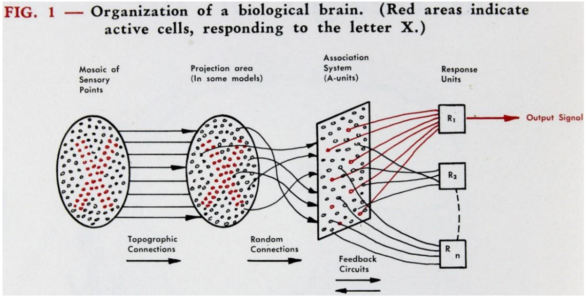
[3] decode behaviors from neural activities

[4] feature landscape of mouse visual cortex

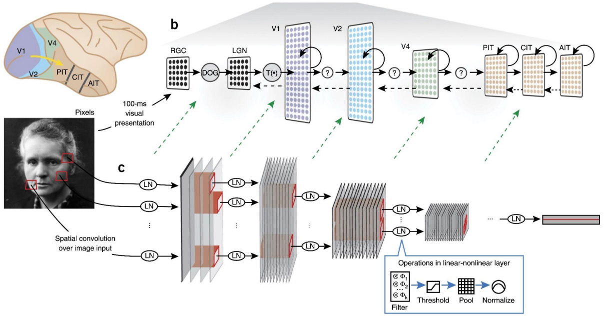
[1] Frank Rosenblatt 1958. [2] Yamins and DiCarlo, Nature Neuro. 2016 [3] Azabou et al. NeurIPS 2023.
[4] Tong et al. bioRxiv 2023.

Neuroscience and machine learning

ML inspired by neuroscience

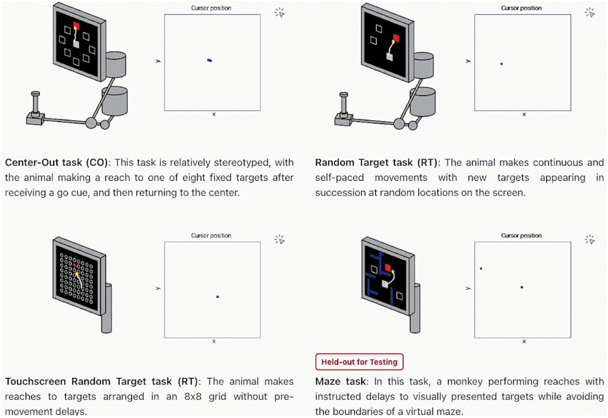


[1] Perceptron

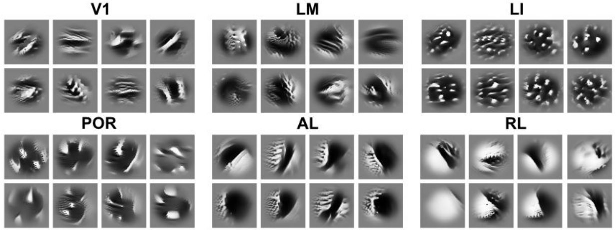


[2] CNN and the ventral stream

ML application in neuroscience



[3] decode behaviors from neural activities

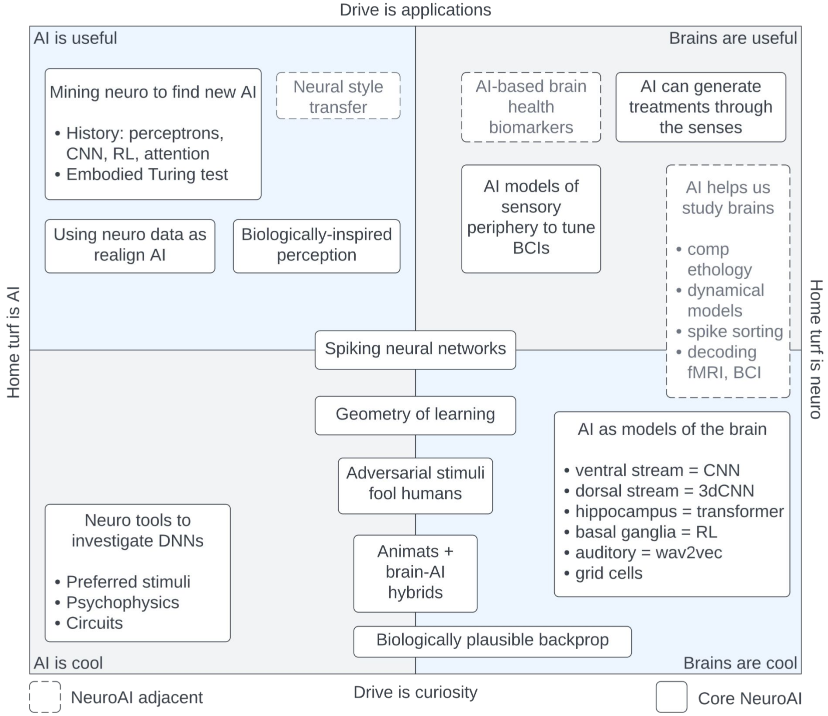


[4] feature landscape of mouse visual cortex

[1] Frank Rosenblatt 1958. [2] Yamins and DiCarlo, Nature Neuro. 2016 [3] Azabou et al. NeurIPS 2023.
[4] Tong et al. bioRxiv 2023.

Emergence of NeuroAI

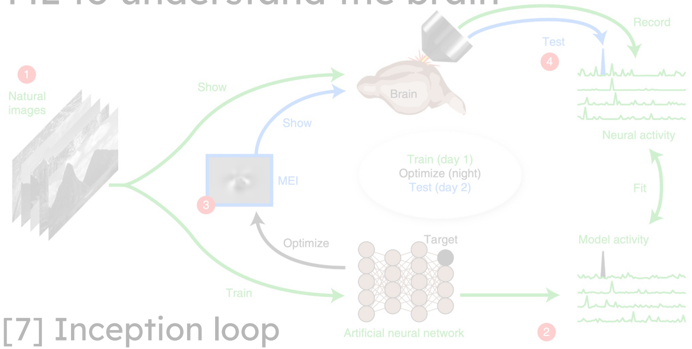
[5] Typology of NeuroAI



1. build better ML using knowledge of the brain

[6] Translate C. elegans locomotion circuits into swimming agent

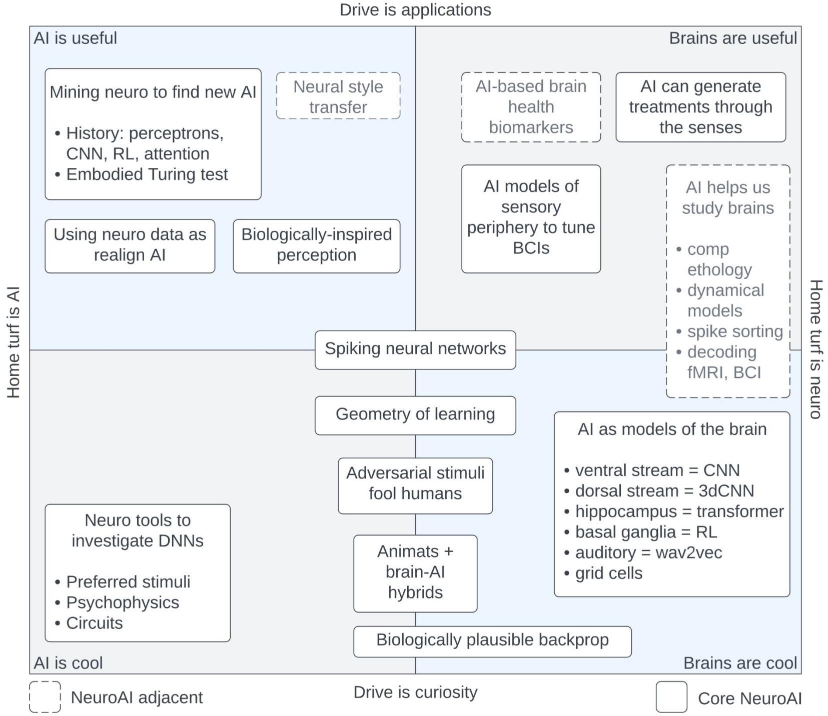
2. use ML to understand the brain



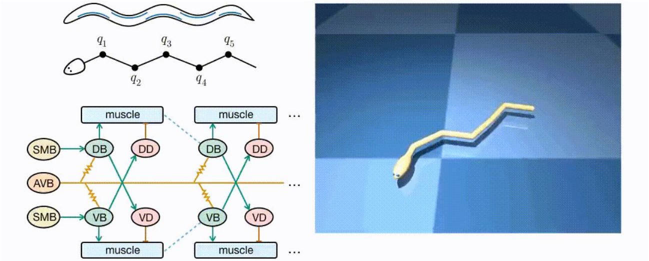
[7] Inception loop

Emergence of NeuroAI

[5] Typology of NeuroAI

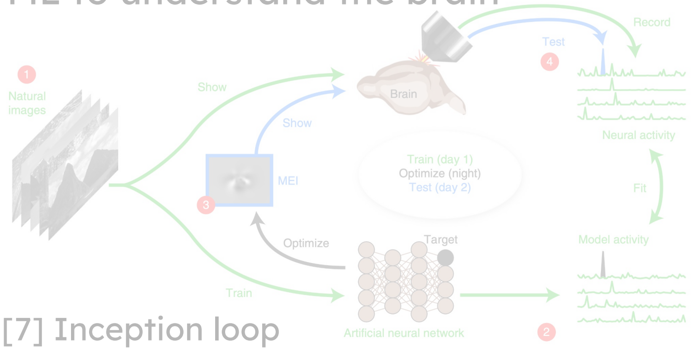


1. build better ML using knowledge of the brain



[6] Translate C. elegans locomotion circuits into swimming agent

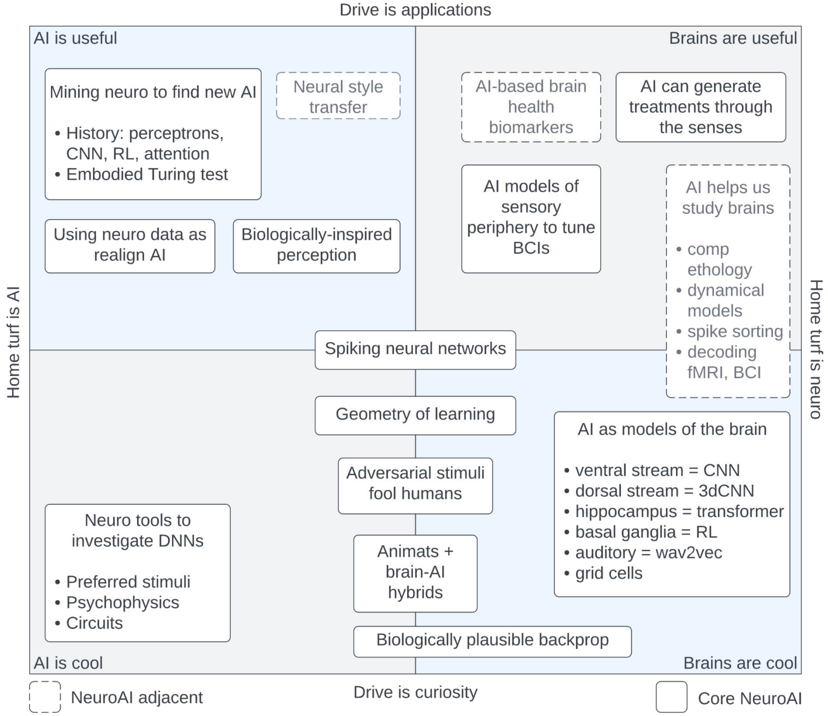
2. use ML to understand the brain



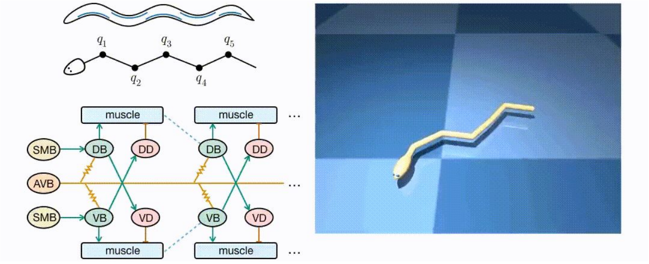
[7] Inception loop

Emergence of NeuroAI

[5] Typology of NeuroAI

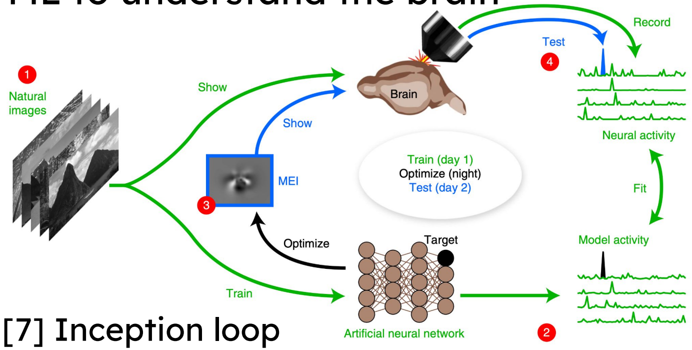


1. build better ML using knowledge of the brain



[6] Translate C. elegans locomotion circuits into swimming agent

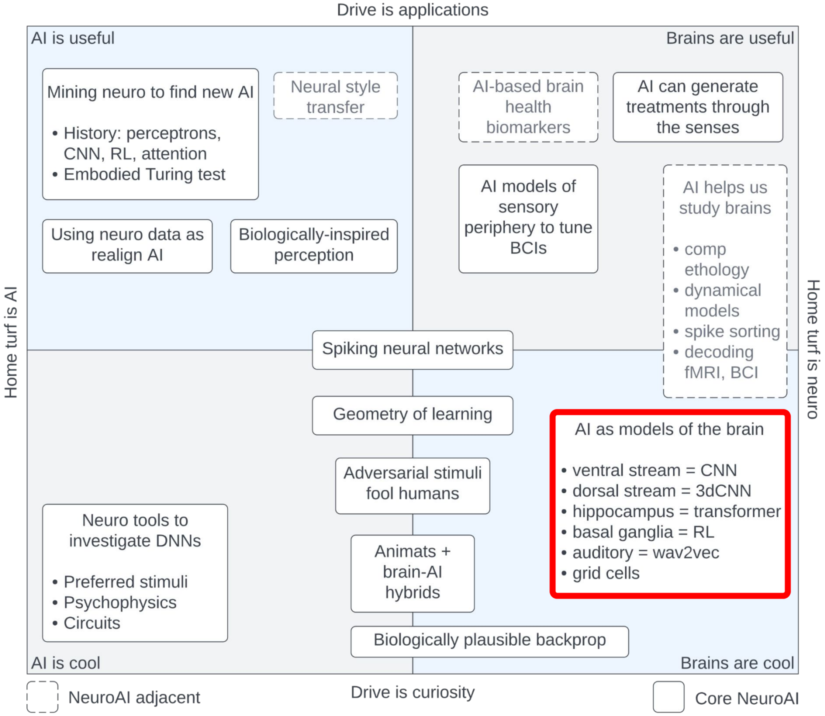
2. use ML to understand the brain



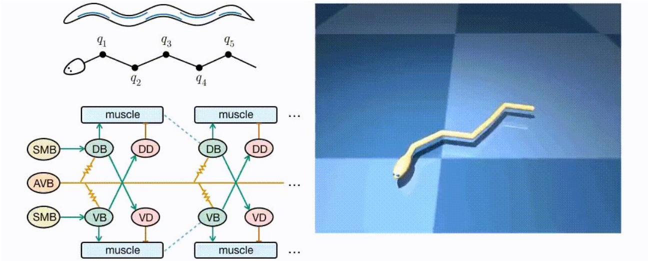
[7] Inception loop

Emergence of NeuroAI

[5] Typology of NeuroAI

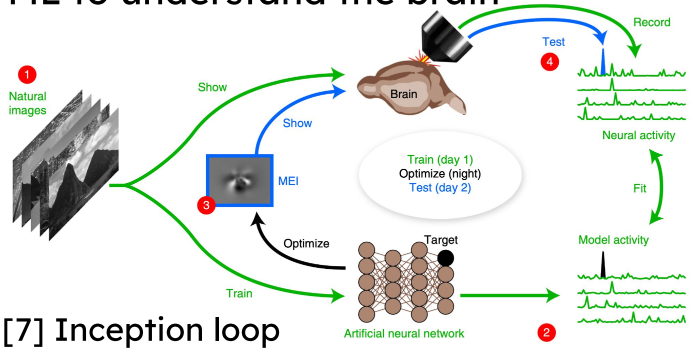


1. build better ML using knowledge of the brain



[6] Translate C. elegans locomotion circuits into swimming agent

2. use ML to understand the brain



[7] Inception loop

ML as a model of the brain

Visual response prediction

Visual response prediction

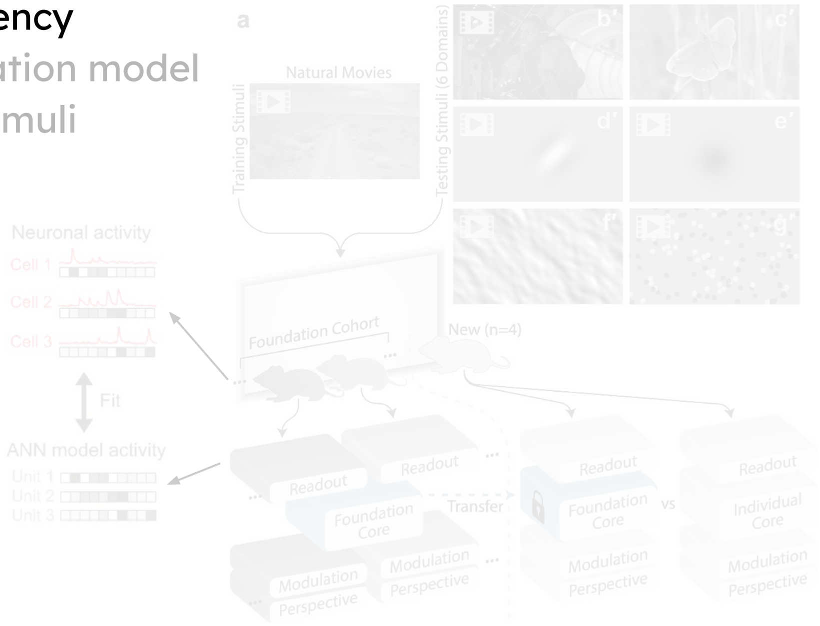
Understanding how the visual system processes information is a fundamental challenge

- Nonlinearity in neural activities w.r.t. visual stimuli
- Limited animal experiments in duration and frequency

Goal: build digital twin of the visual system, a foundation model

- accurately prediction visual response to natural stimuli
- generalize across animals and stimuli
- ML challenge: Brain score, Sensorium, ...

[8] foundation model of mouse V1



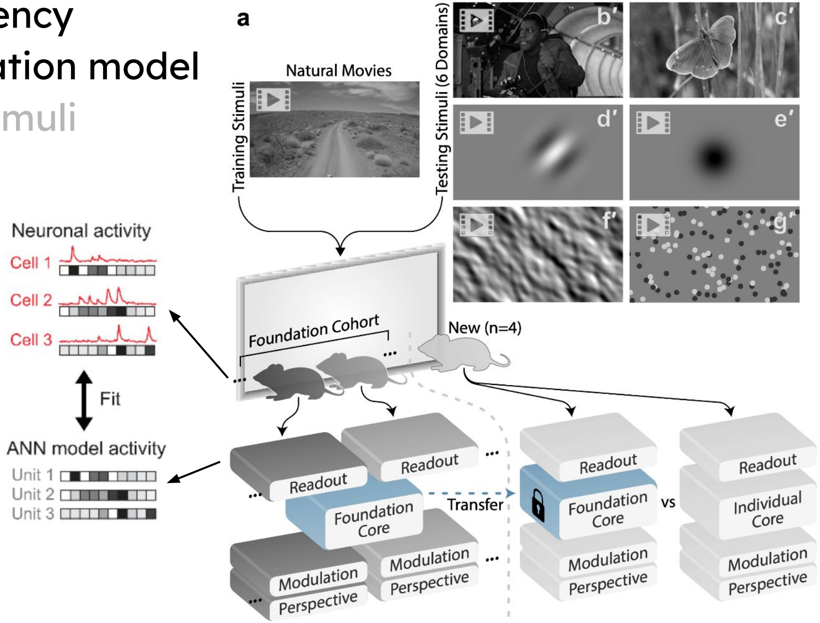
[8] Wang et al. bioRxiv 2023. [9] Schrimpf et al. bioRxiv 2020. [10] Willeke et al. arXiv 2022.
[11] Li et al. TMLR 2023

Visual response prediction

Understanding how the visual system processes information is a fundamental challenge
- Nonlinearity in neural activities w.r.t. visual stimuli
- Limited animal experiments in duration and frequency

Goal: build digital twin of the visual system, a foundation model
- accurately prediction visual response to natural stimuli
- generalize across animals and stimuli
- ML challenge: Brain score, Sensorium, ...

[8] foundation model of mouse V1



Visual response prediction

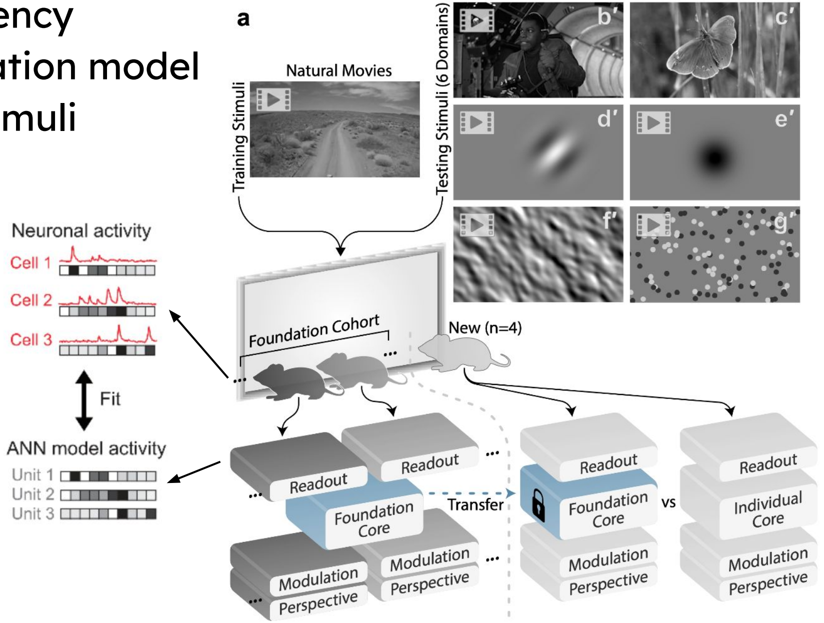
Understanding how the visual system processes information is a fundamental challenge

- Nonlinearity in neural activities w.r.t. visual stimuli
- Limited animal experiments in duration and frequency

Goal: build digital twin of the visual system, a foundation model

- accurately prediction visual response to natural stimuli
- generalize across animals and stimuli
- ML challenge: Brain score, Sensorium, ...

[8] foundation model of mouse V1

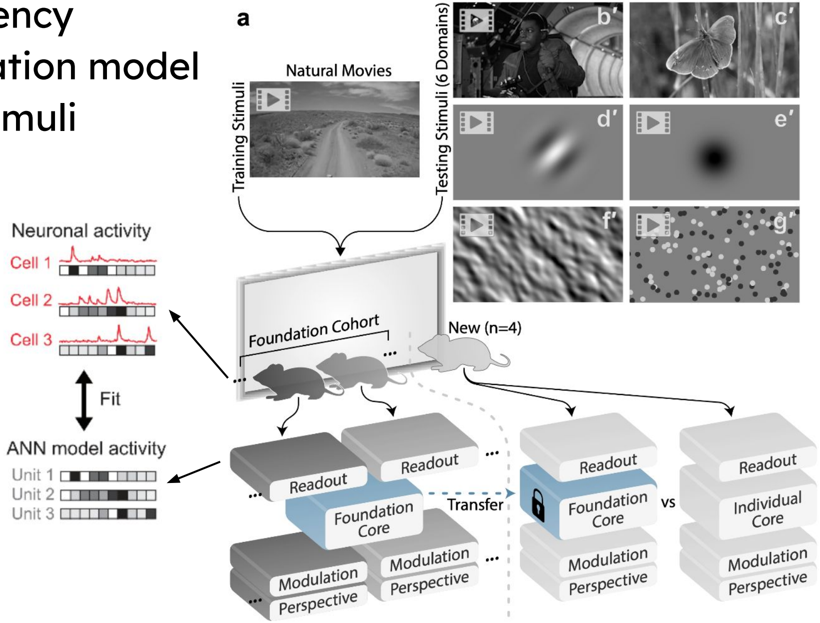


Visual response prediction

Understanding how the visual system processes information is a fundamental challenge
- Nonlinearity in neural activities w.r.t. visual stimuli
- Limited animal experiments in duration and frequency

Goal: build digital twin of the visual system, a foundation model
- accurately prediction visual response to natural stimuli
- generalize across animals and stimuli
- ML challenge: Brain score, Sensorium, ...

[8] foundation model of mouse V1

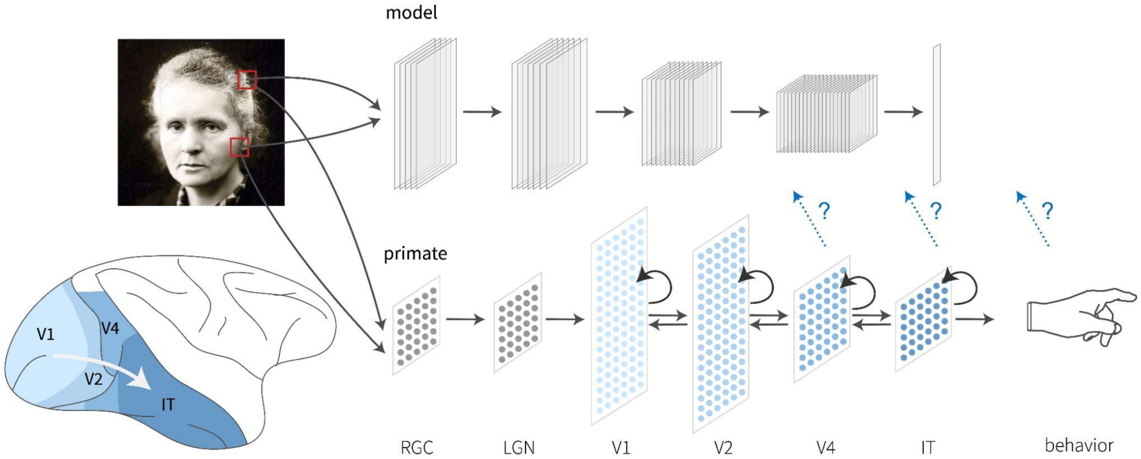


Visual response prediction

Understanding how the visual system processes information is a fundamental challenge

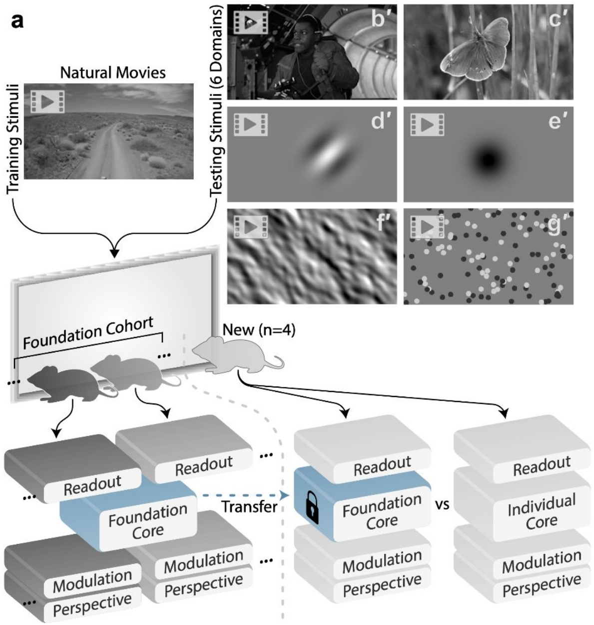
Goal: build digital twin of the visual system, a foundation model

- ML challenge: Brain score



[9] align model response with macaque monkeys V4 and IT neurons.

[8] foundation model of mouse V1

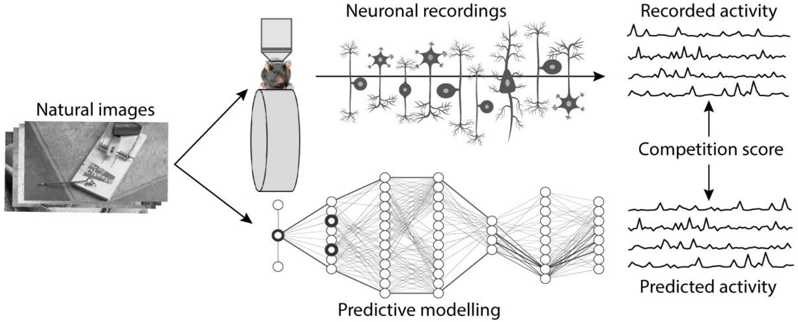


Visual response prediction

Understanding how the visual system processes information is a fundamental challenge

Goal: build digital twin of the visual system, a foundation model

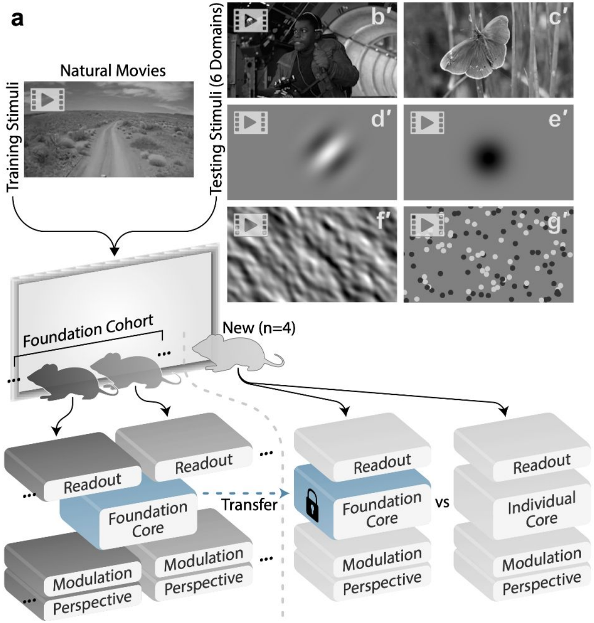
- ML challenge: Sensorium (NeurIPS competition track)
  - Sensorium 2022: static stimuli
  - Sensorium 2023: dynamic stimuli



[10] predict mouse V1 response to natural stimuli

[11] V1T: V1 response prediction using ViT

[8] foundation model of mouse V1

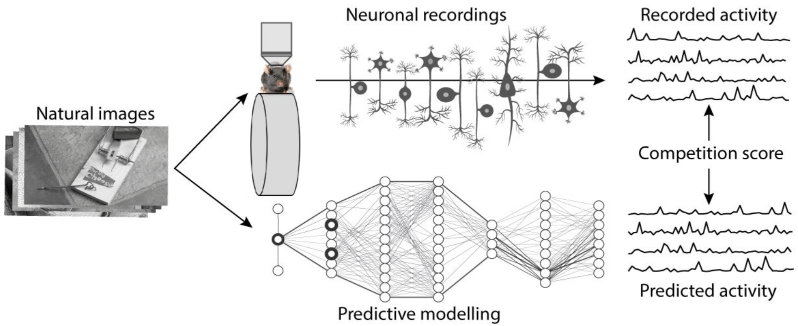


Visual response prediction

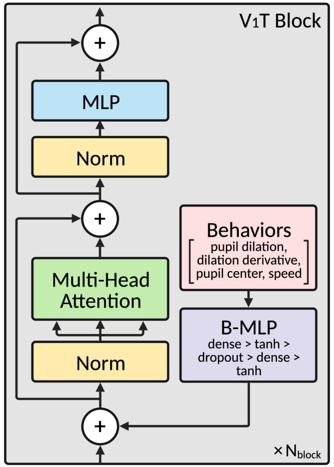
Understanding how the visual system processes information is a fundamental challenge

Goal: build digital twin of the visual system, a foundation model

- ML challenge: Sensorium (NeurIPS competition track)
  - Sensorium 2022: static stimuli
  - Sensorium 2023: dynamic stimuli

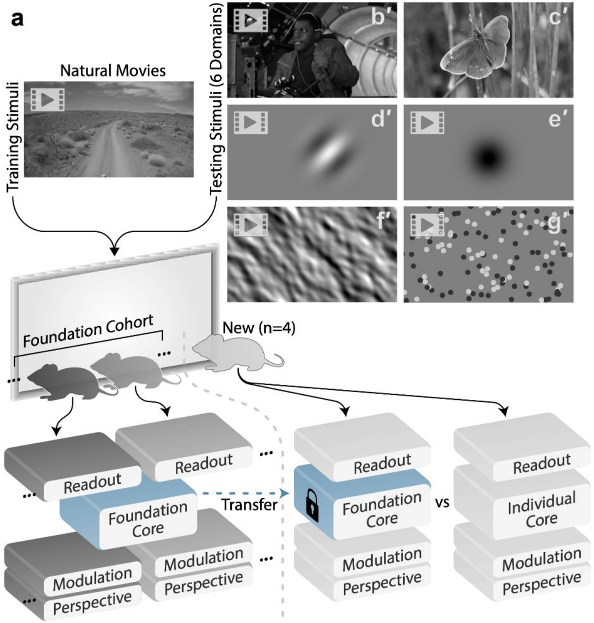


[10] predict mouse V1 response to natural stimuli



[11] V1T: V1 response prediction using ViT

[8] foundation model of mouse V1

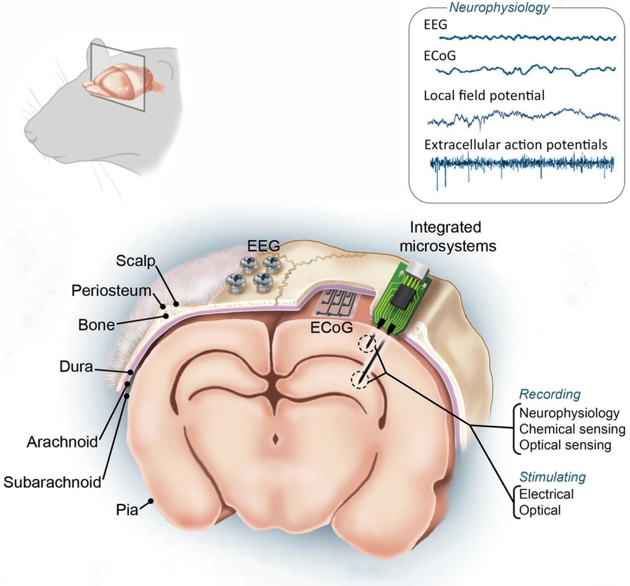


[8] Wang et al. bioRxiv 2023. [9] Schrimpf et al. bioRxiv 2020. [10] Willeke et al. arXiv 2022.
[11] Li et al. TMLR 2023

In vivo recordings from behaving mice

I am NOT an experimentalist

Neural recordings

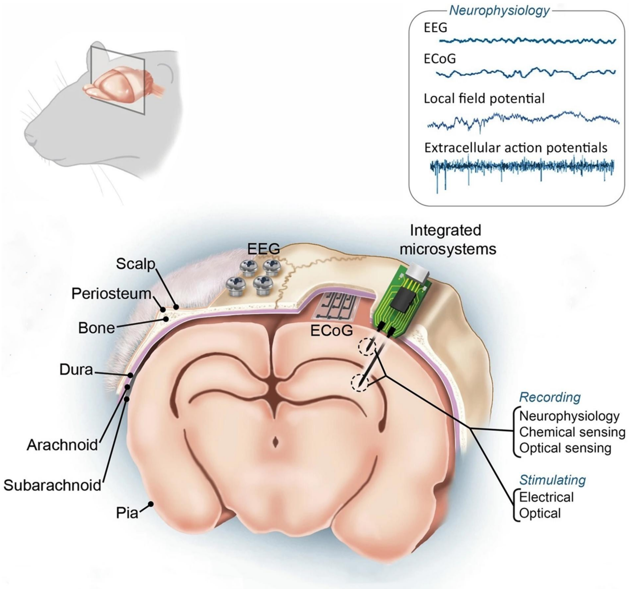


Type of neural recordings [12]

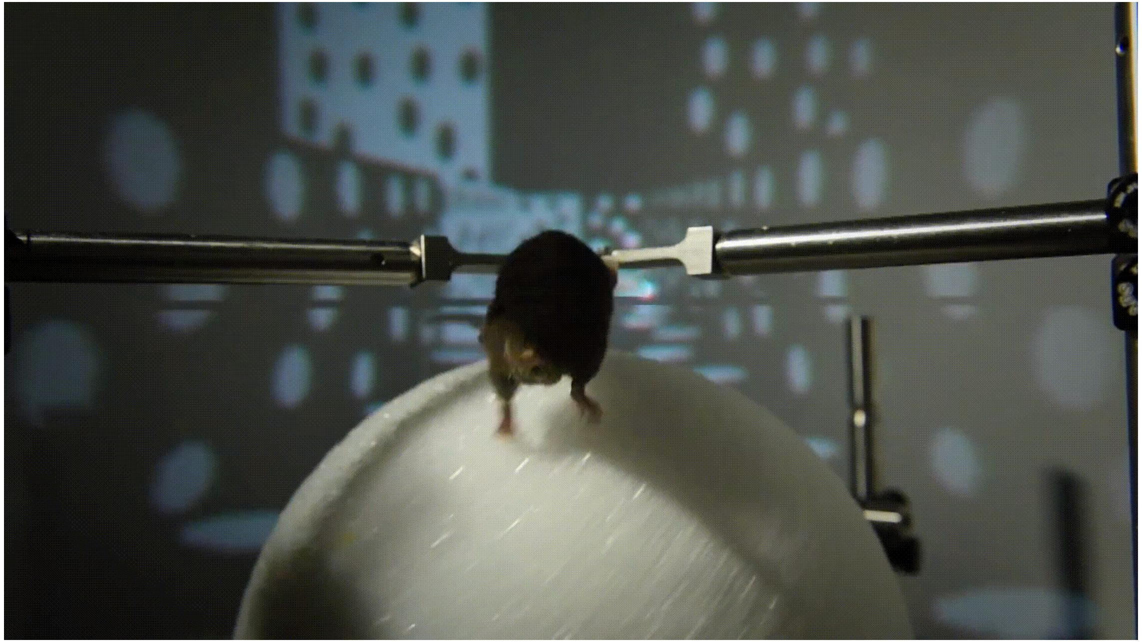
Head fixed mouse in virtual maze navigation experiment [13]

[12] Seymour et al. Nature Microsystems & Nanoengineering 2017. [13] Ekaterina Pesheva, Harvard Medical School 2016.

Neural recordings



Type of neural recordings [12]



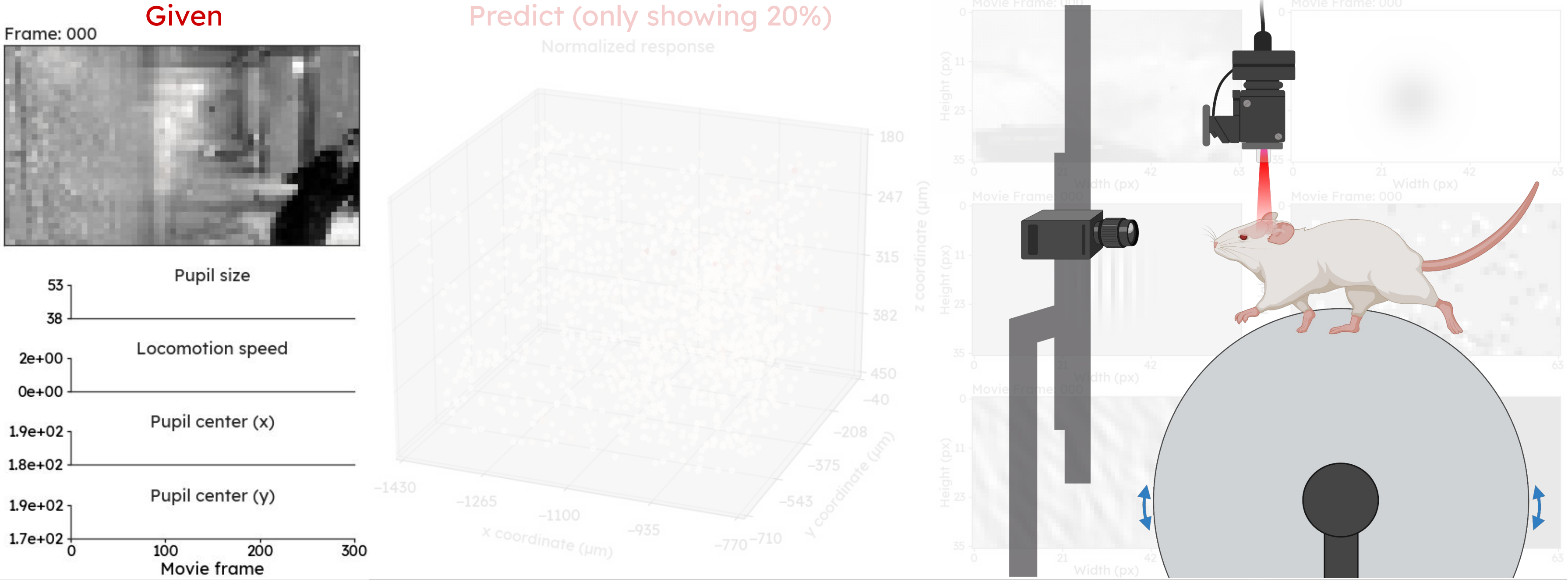
Head fixed mouse in virtual maze navigation experiment [13]

[12] Seymour et al. Nature Microsystems & Nanoengineering 2017. [13] Ekaterina Pesheva, Harvard Medical School 2016.

Sensorium 2023 [14] (NeurIPS competition track)

Predict mouse V1 response to dynamic stimuli
10 mice, ~8000 neurons per mouse.

Train on movie (10s clips, ~300 clips per mouse)
Test on unseen movies and artificial patterns

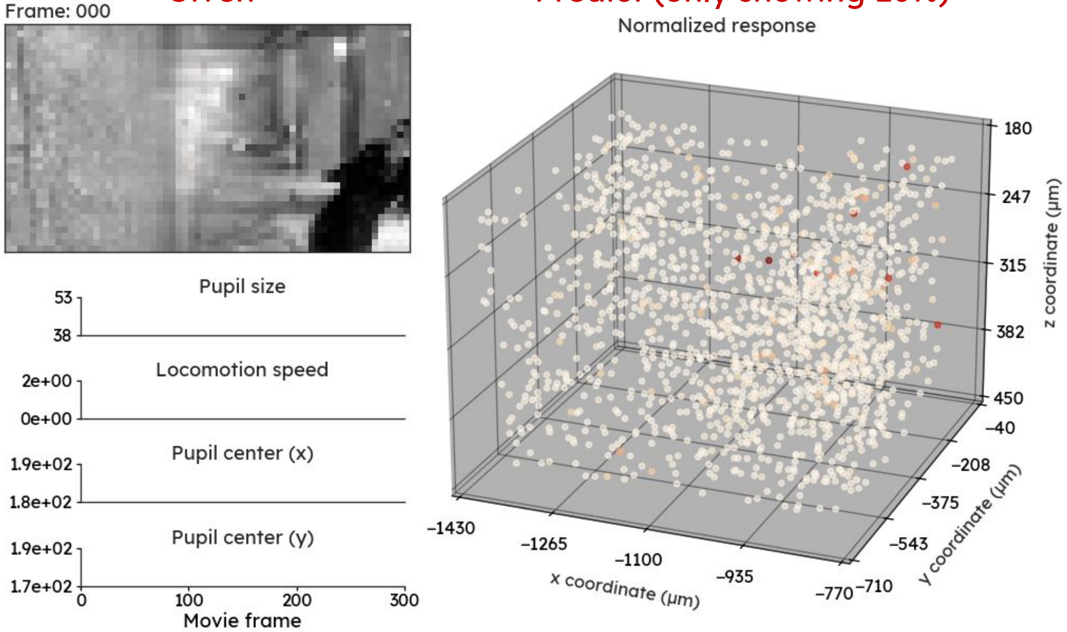


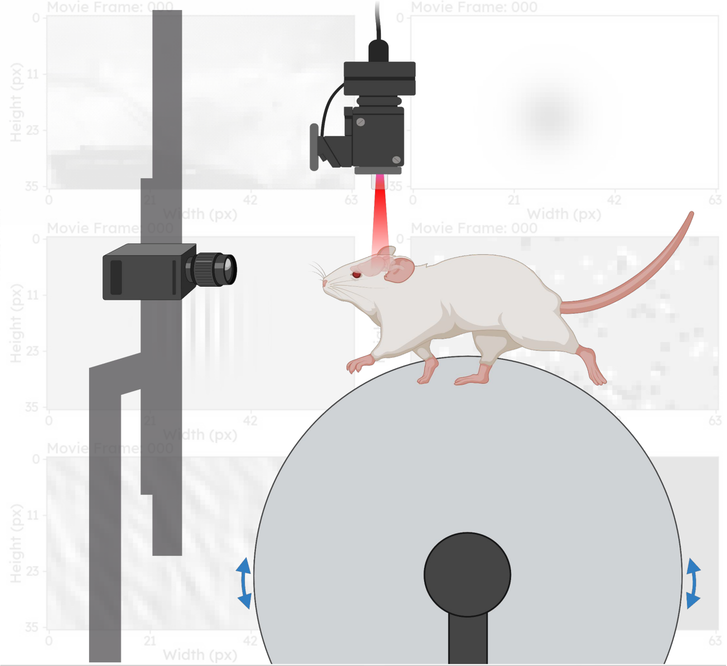
[14] Turishcheva et al. arXiv 2023.

Sensorium 2023 [14] (NeurIPS competition track)

Predict mouse V1 response to dynamic stimuli
10 mice, ~8000 neurons per mouse.

Train on movie (10s clips, ~300 clips per mouse)
Test on unseen movies and artificial patterns





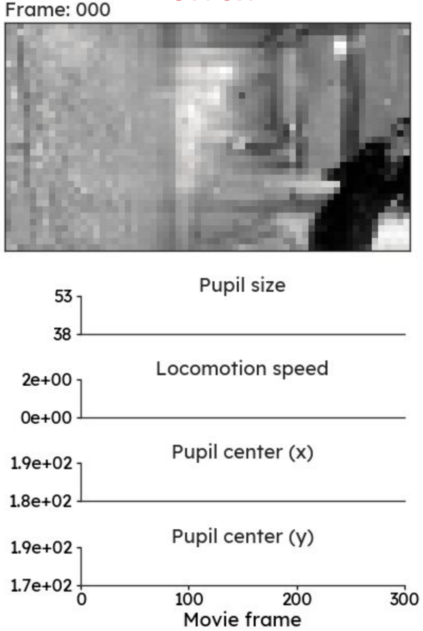
[14] Turishcheva et al. arXiv 2023.

Sensorium 2023 [14] (NeurIPS competition track)

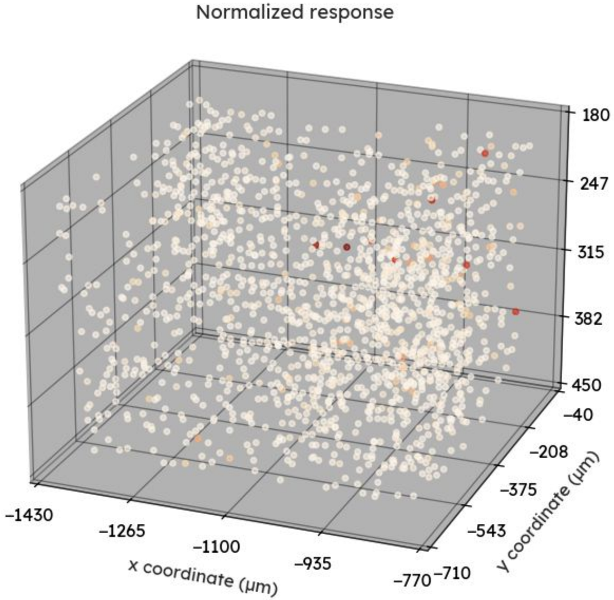
Predict mouse V1 response to dynamic stimuli
10 mice, ~8000 neurons per mouse.

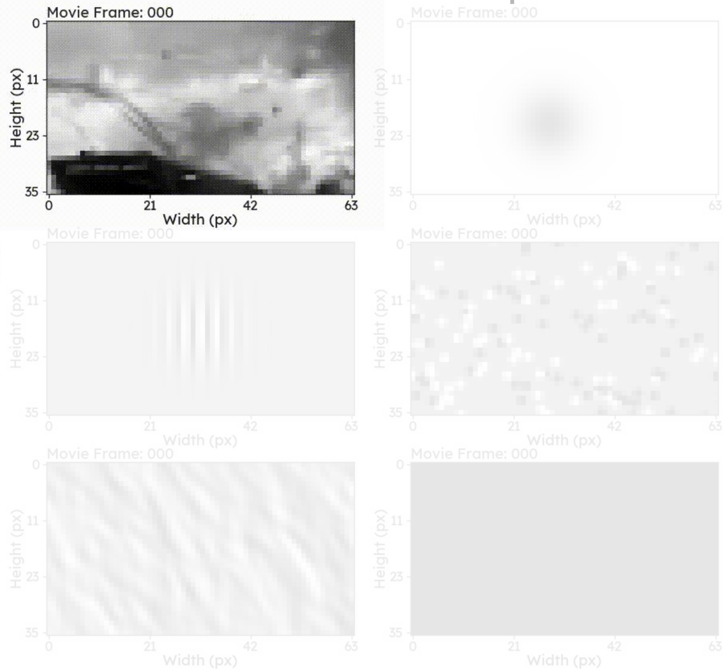
Train on movie (10s clips, ~300 clips per mouse)
Test on unseen movies and artificial patterns

Given



Predict (only showing 20%)



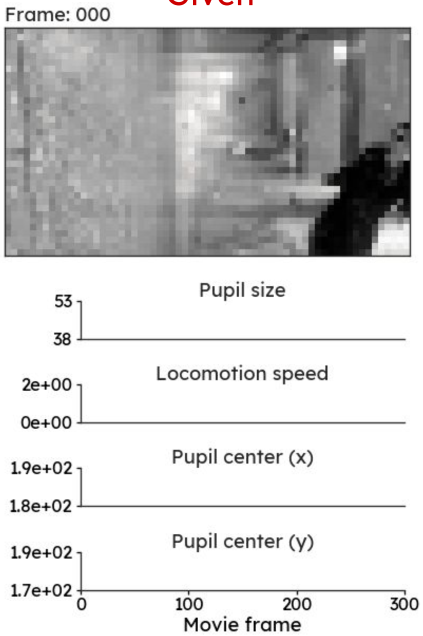


Sensorium 2023 [14] (NeurIPS competition track)

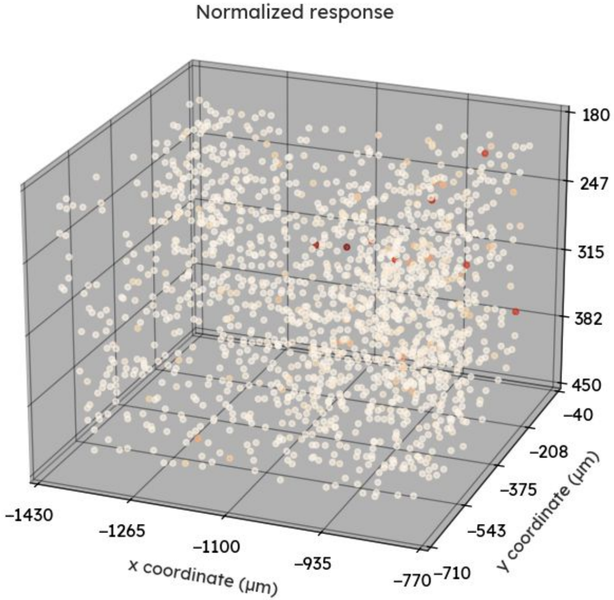
Predict mouse V1 response to dynamic stimuli
10 mice, ~8000 neurons per mouse.

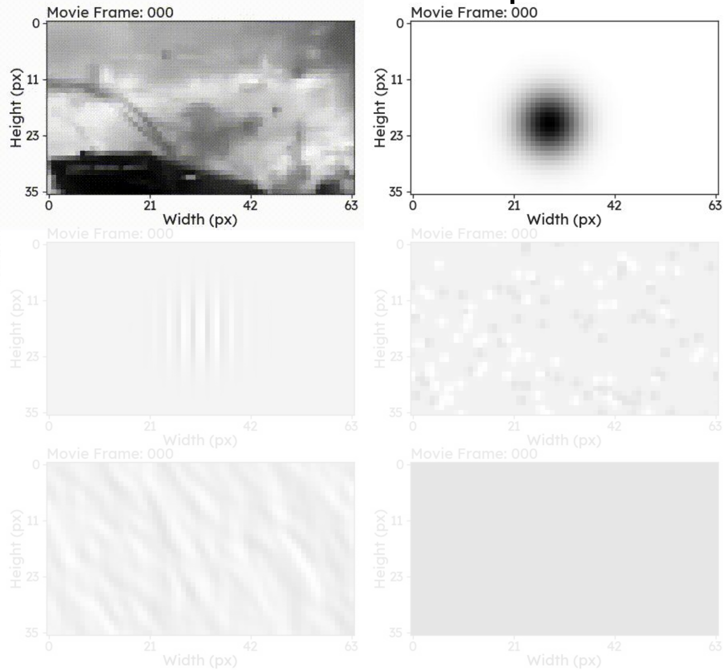
Train on movie (10s clips, ~300 clips per mouse)
Test on unseen movies and artificial patterns

Given



Predict (only showing 20%)



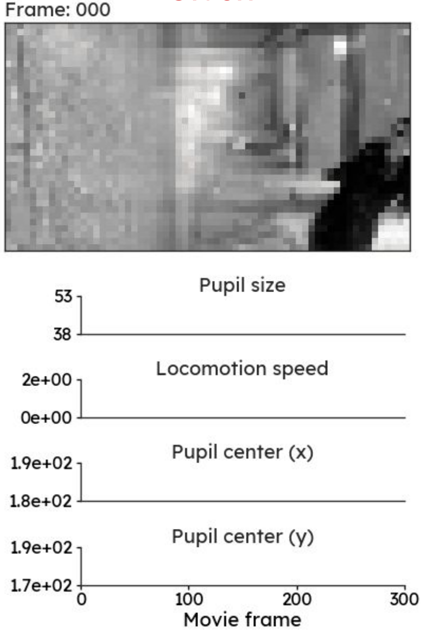


Sensorium 2023 [14] (NeurIPS competition track)

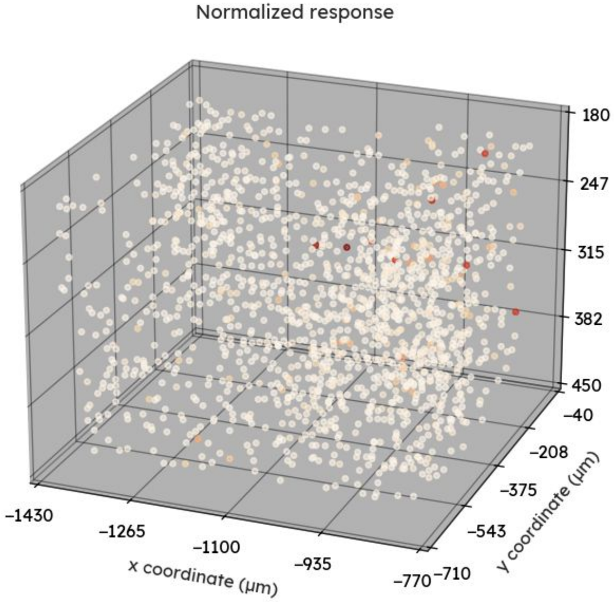
Predict mouse V1 response to dynamic stimuli
10 mice, ~8000 neurons per mouse.

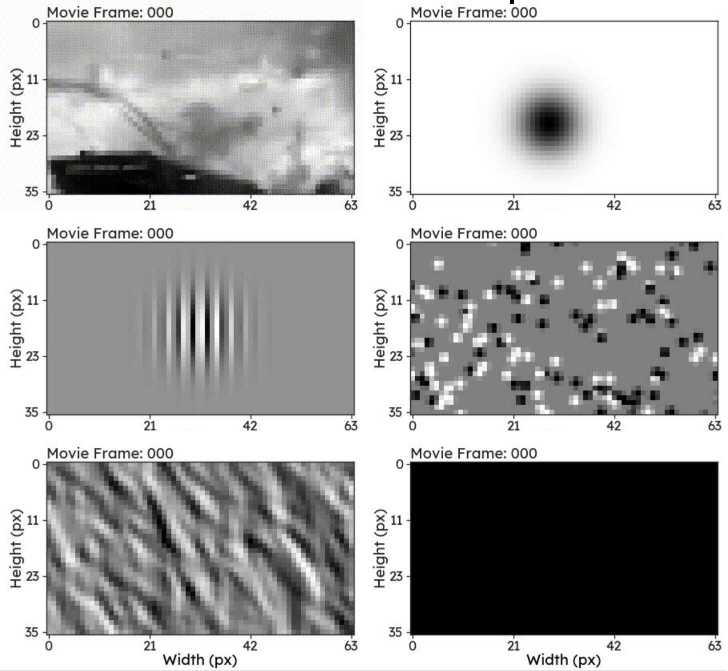
Train on movie (10s clips, ~300 clips per mouse)
Test on unseen movies and artificial patterns

Given



Predict (only showing 20%)





Video vision Transformer
for dynamic response prediction

slides removed

Best single model and 🥉 overall in the NeurIPS2023 challenge

movie test set

	Team	single trial correlation
1	lRomul	0.301
2	YuZhu	0.276
3	dunedin	0.254
4	gautam8387	0.231
5	dengkw	0.228
6	JBauer	0.210

OOD test set

	Team	single trial correlation
1	lRomul	0.217
2	YuZhu	0.180
3	dunedin	0.175
4	dengkw	0.147
5	gautam8387	0.141
6	BENCHMARK	0.128

Same training procedure as prev SOTA [20], +12% and +19% in movies and artificial patterns

🥉 in movie and artificial patterns in the competition overall (incl. ensembles)

🥇 in single model setting

[20] Höfling et al. bioRxiv 2022.

Best single model and 🥉 overall in the NeurIPS2023 challenge

movie test set

	Team	single trial correlation
1	lRomul	0.301
2	YuZhu	0.276
3	dunedin	0.254
4	gautam8387	0.231
5	dengkw	0.228
6	JBauer	0.210

OOD test set

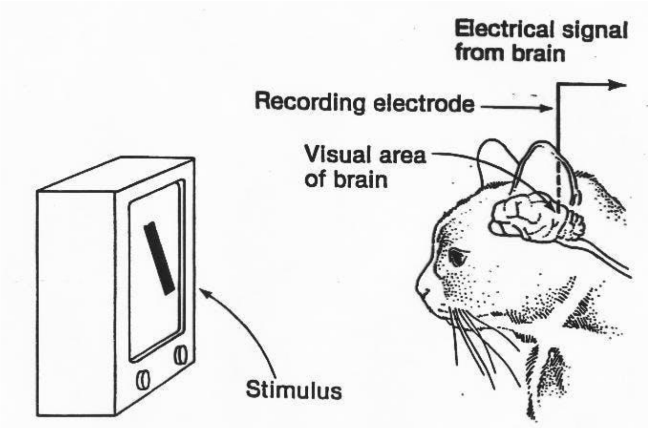
	Team	single trial correlation
1	lRomul	0.217
2	YuZhu	0.180
3	dunedin	0.175
4	dengkw	0.147
5	gautam8387	0.141
6	BENCHMARK	0.128

Same training procedure as prev SOTA [20], +12% and +19% in movies and artificial patterns
🥉 in movie and artificial patterns in the competition overall (incl. ensembles)
🥇 in single model setting

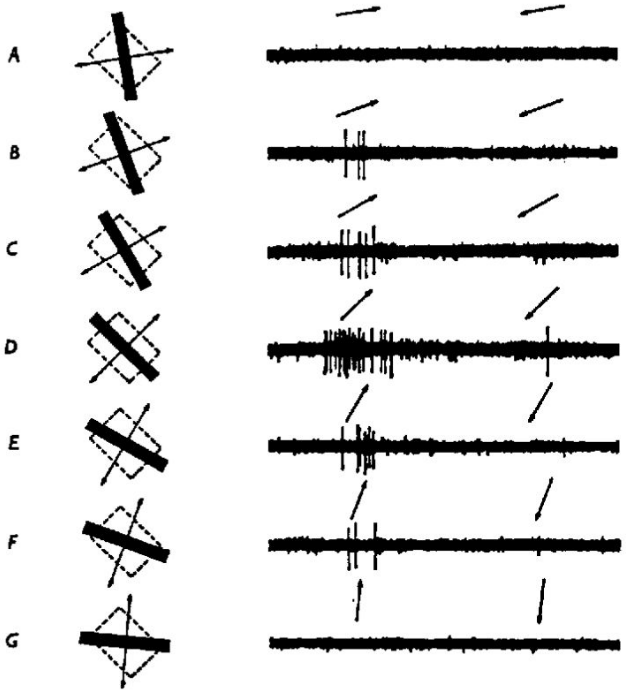
[20] Höfling et al. bioRxiv 2022.

Functional twin of the mouse visual cortex

Model learns tuning properties of mouse V1 neurons



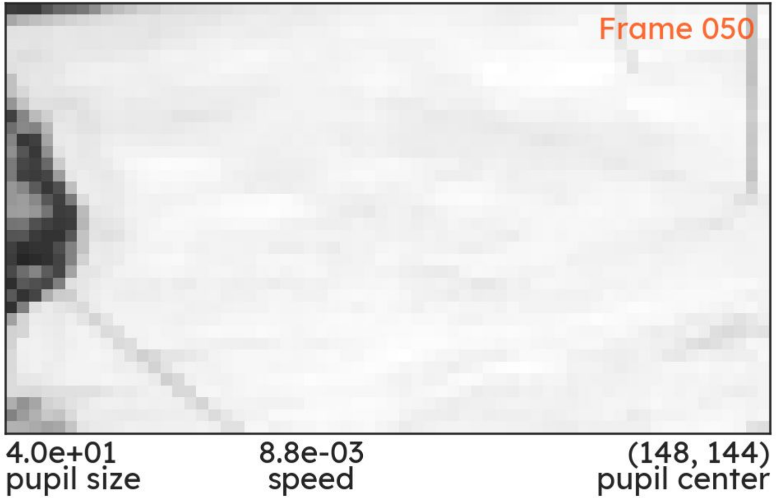
Specialized neurons that respond only to certain sensory information (e.g. orientation and direction of movement) [21]



slides removed

A platform to investigate the computation in the visual system *in silico*

What/where is the model looking at?



Visualize area(s) in the visual stimulus that drive response prediction via attention rollout [23]

“attention” **strictly** refers to the attention mechanism in Transformer [24]!

[23] Abnar and Zuidema, arXiv 2021. [24] Vaswani et al., NeurIPS 2017.

slides removed

Are Transformers better than CNNs in modelling the visual cortex?

IMO not necessarily

What is a good model of the visual cortex?

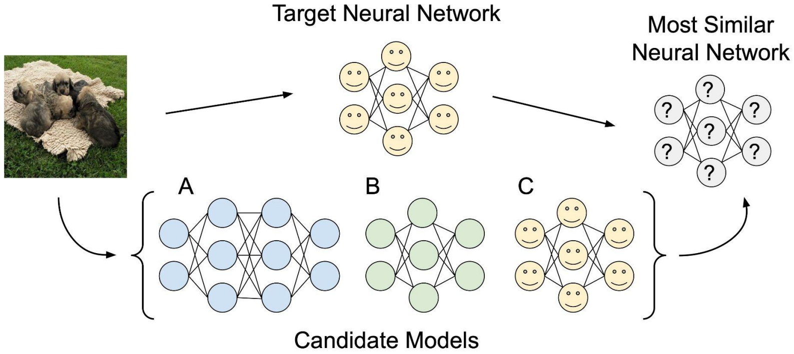


Figure 1. **Illustration of the research question:** If one of the candidate models matches the target neural network (brain) closely, can the current model evaluation methods accurately find that out?

“What I cannot create, I do not understand.” - Richard Feynman

What I can create, I do not necessarily understand?

What is a good model of the visual cortex?

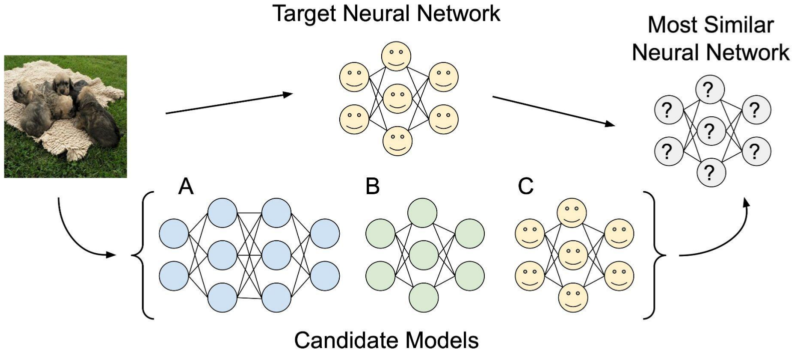


Figure 1. **Illustration of the research question:** If one of the candidate models matches the target neural network (brain) closely, can the current model evaluation methods accurately find that out?

“What I cannot create, I do not understand.” - Richard Feynman

What I can create, I do not necessarily understand?

What is a good model of the visual cortex?

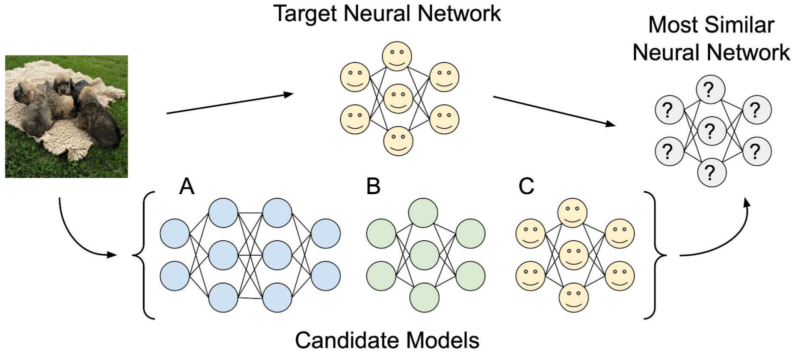
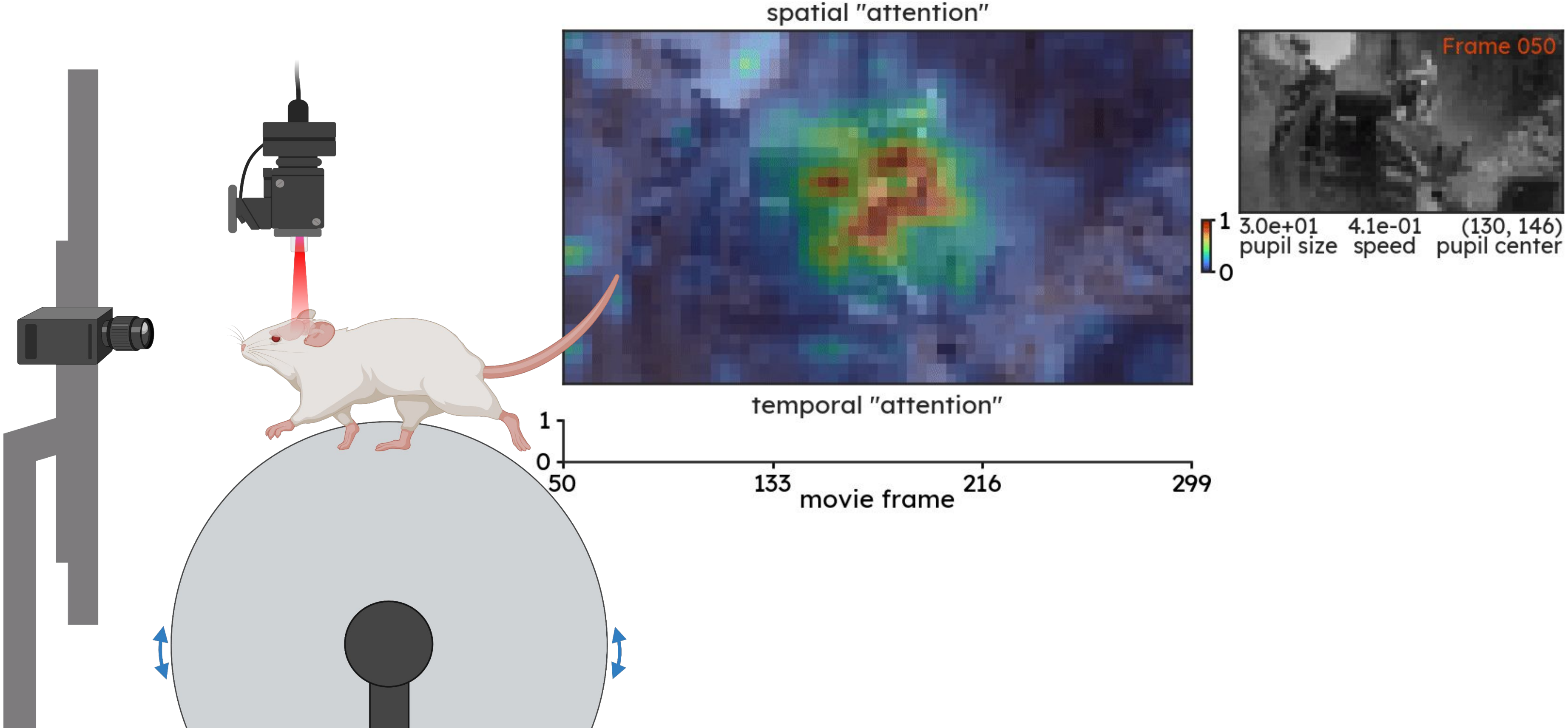


Figure 1. **Illustration of the research question:** If one of the candidate models matches the target neural network (brain) closely, can the current model evaluation methods accurately find that out?

"What I cannot create, I do not understand." - Richard Feynman

What I can create, I do not necessarily understand?

Find me @bryanlimy



slides removed

V1T

Backup

Cross-animal generalization

Cross-animal: pre-train on 4 animals, fine-tune readout on the left-out animal in Dataset S

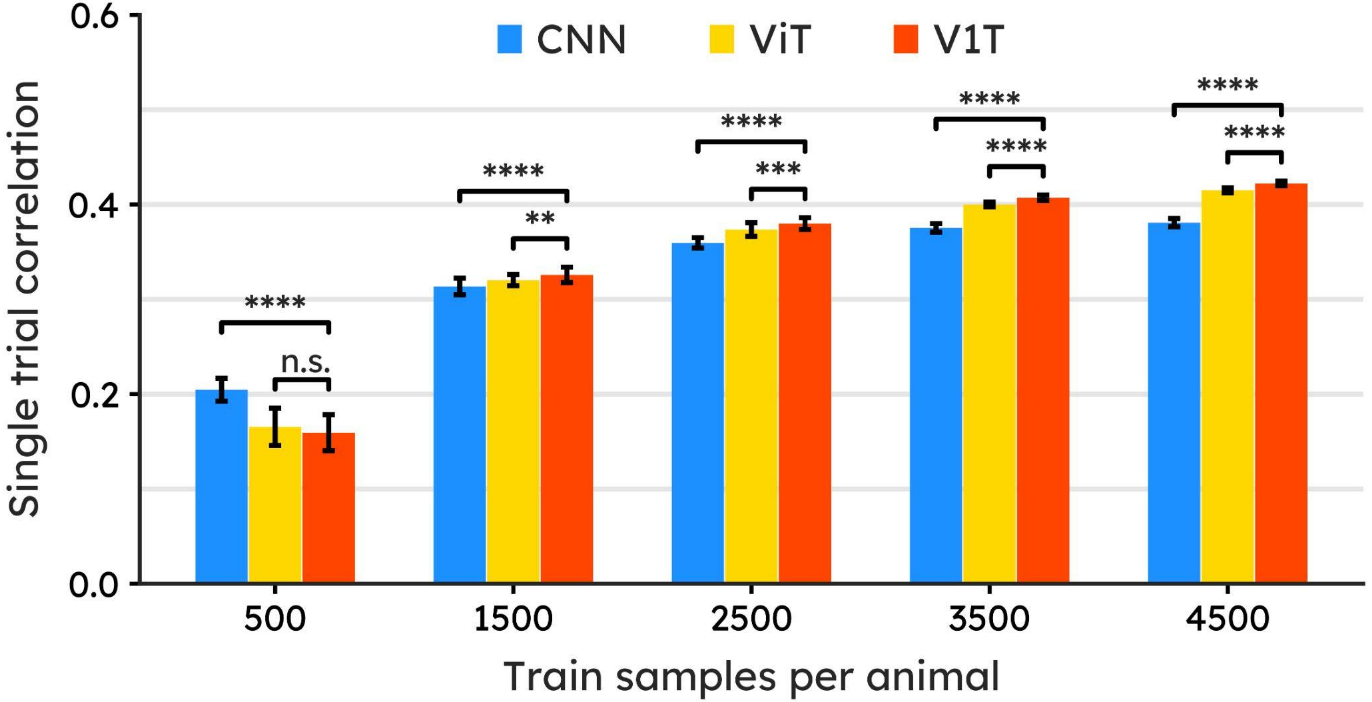
	A	B	C	D	E	mean	std
			CNN				
Original	0.350	0.424	0.385	0.371	0.360	0.378	0.029
1-out	0.353	0.394	0.373	0.365	0.333	**0.364**	0.023
% diff	0.920	-7.102	-3.101	-1.570	-7.479	**-3.666**	**3.609**
			V1T				
Original	0.401	0.464	0.430	0.436	0.401	0.426	0.027
1-out	0.388	0.450	0.415	0.417	0.384	**0.411**	0.027
% diff	-3.175	-2.885	-3.557	-4.381	-4.139	**-3.628**	**0.630**

Cross-dataset generalization

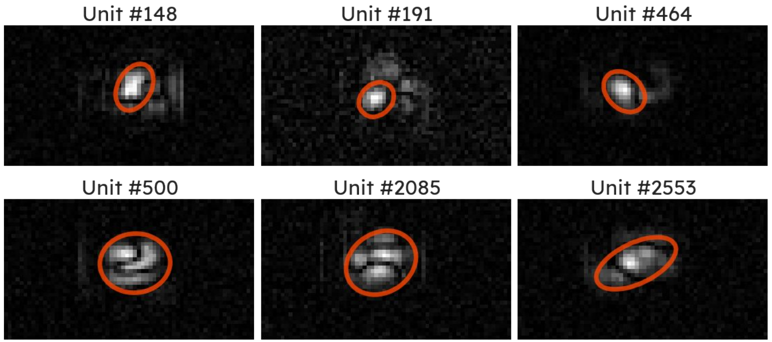
Cross-dataset: pre-train core on gray-scale Dataset F → fine-tune readouts on Dataset S

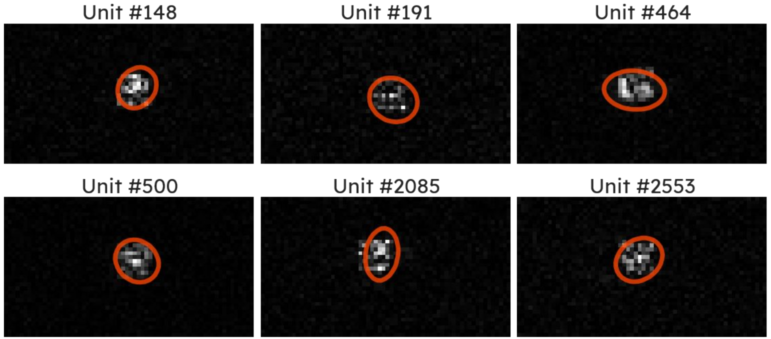
	A	B	C	D	E	mean	std
			CNN				
Original	0.350	0.424	0.385	0.371	0.360	0.378	0.029
Fine-tune	0.314	0.353	0.337	0.316	0.327	**0.329**	0.016
% diff	-10.182	-16.912	-12.373	-14.746	-9.363	**-12.715**	**3.143**
			V1T				
Original	0.401	0.464	0.430	0.436	0.401	0.426	0.027
Fine-tune	0.327	0.382	0.347	0.343	0.328	**0.345**	0.022
% diff	-18.291	-17.590	-19.310	-21.495	-18.114	**-18.960**	**1.548**

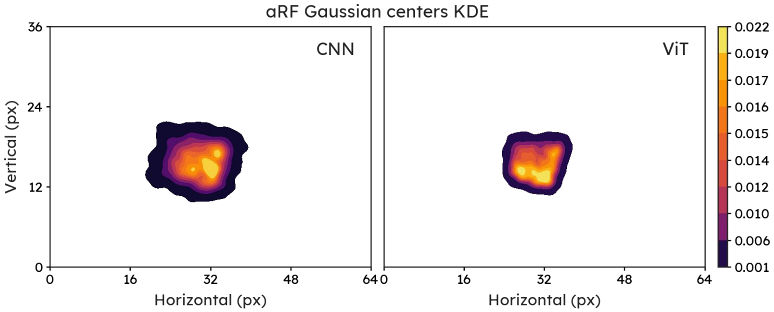
Sample efficiency

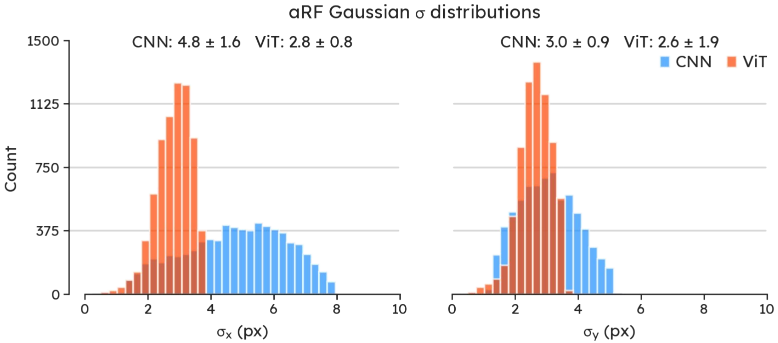


CNN vs ViT in visual response prediction

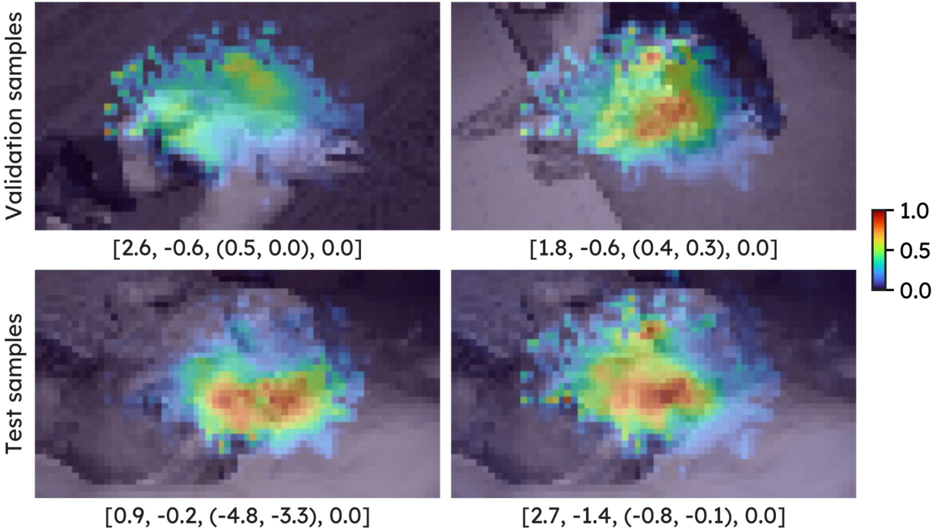








Attention rollout maps



Correlation between
pupil centers and attention map center of mass

Mouse	Horizontal	Vertical
A	0.682 (****)	0.568 (****)
B	0.489 (****)	0.493 (****)
C	0.505 (****)	0.370 (****)
D	0.484 (****)	0.310 (****)
E	0.464 (****)	0.302 (****)
Avg.	0.525	0.409

Application

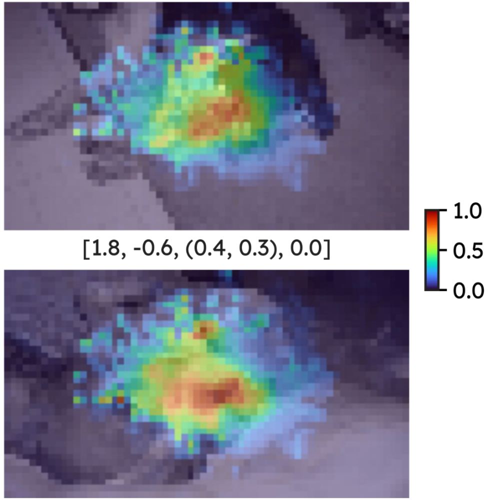
Investigate the coding of visual information across visual cortical areas

- e.g. larger receptive fields in higher visual areas

Determine which part of a visual scene was relevant when performing more specific tasks

- e.g. object recognition, decision-making, or spatial navigation

“Attention-guided” visual predictive models



[1.8, -0.6, (0.4, 0.3), 0.0]

[2.7, -1.4, (-0.8, -0.1), 0.0]