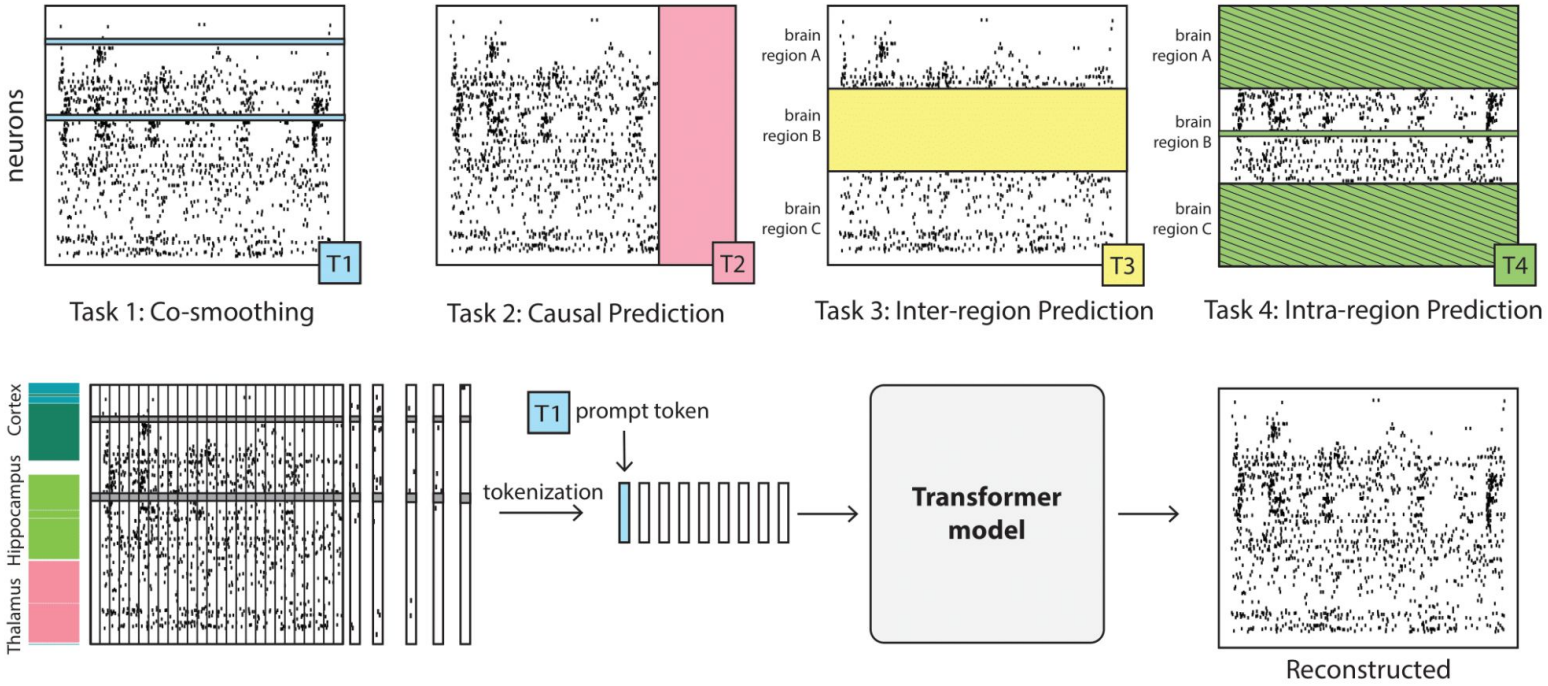
Towards a “universal translator” for neural dynamics at single-cell, single-spike resolution

Yizi Zhang, Yanchen Wang, Donato M. Jiménez-Benetó, Zixuan Wang, Mehdi Azabou, Blake Aaron Richards, Renee Tung, Olivier Winter, International Brain Laboratory, Eva L Dyer, Liam Paninski, Cole Hurwitz

Multi-task-masking (MtM) - a pre-training recipe for learning spatiotemporal dynamics



Generative process for MtM

$$M \sim \mathcal{U}(\text{causal, neuron, intra-region, inter-region})$$

$$Z = \text{Tokenizer}(M \odot X)$$

$$Z_{\text{prompt}} = [P, Z]$$

$$r = \text{Transformer}(Z_{\text{prompt}})$$

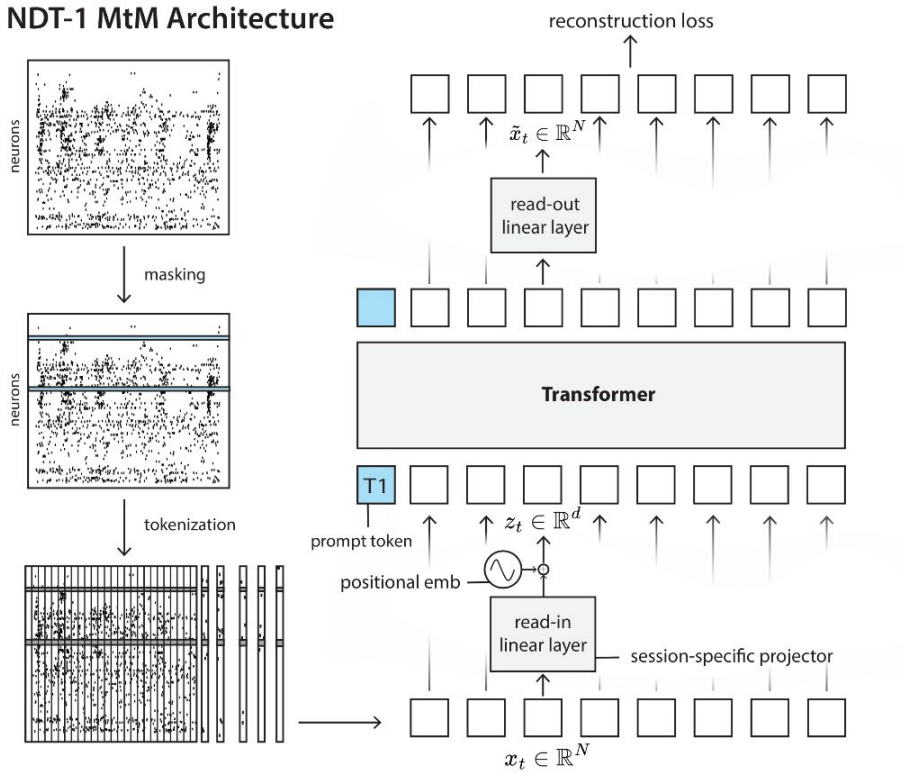
$$\hat{X} \sim \text{Poisson}(X \mid r)$$

Masking schemes + Architectures

Temporal masking (baseline): Mask and reconstruct randomly selected portions of time steps in the data.

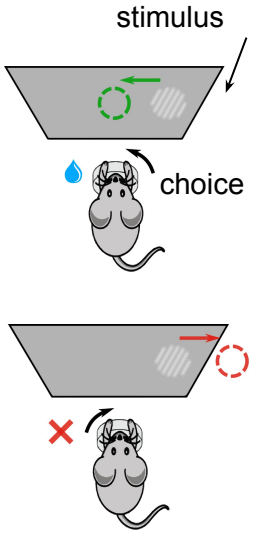
MtM (ours): Alternate between neuron, causal, inter-region, and intra-region masking schemes.

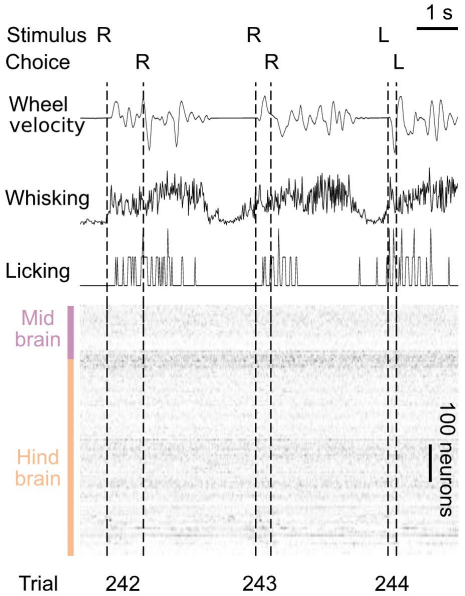
NDT-1: Transformer architecture for activity prediction which turns time steps into “tokens”.



Datasets

- IBL repeated site dataset - Neuropixels recordings targeting the same brain locations in mice performing a visual decision-making task.
- Trial-aligned, spike-sorted data from 39 mice and 26,376 neurons (~676 neurons per session).
- Neural activity is binned into 20 ms windows, with a fixed trial length of 2 seconds.





IBL et al. 2023

Metrics

Co-Smoothing: Predict the activity of a held-out neuron using all other neurons (Macke et al. 2011, Ye et al. 2021).

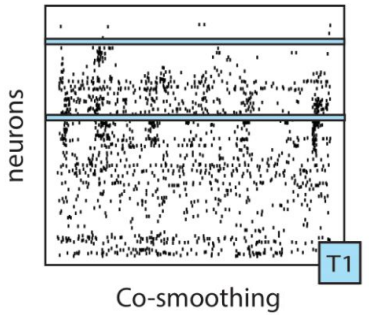
Forward prediction: Predict future activity from past activity. We predict the last 10% (200 ms) of the trial-aligned activity (2 seconds).

Intra-region co-smoothing: Predict the activity of a held-out neuron using neurons in the *same* brain region.

Inter-region co-smoothing: Predict the activity of a held-out neuron using neurons in *other* brain regions.

Choice decoding: Predict the choice the mouse makes using all trial-aligned neural activity.

Motion energy decoding: Predict motion energy of the mouse's whiskers using all trial-aligned neural activity.

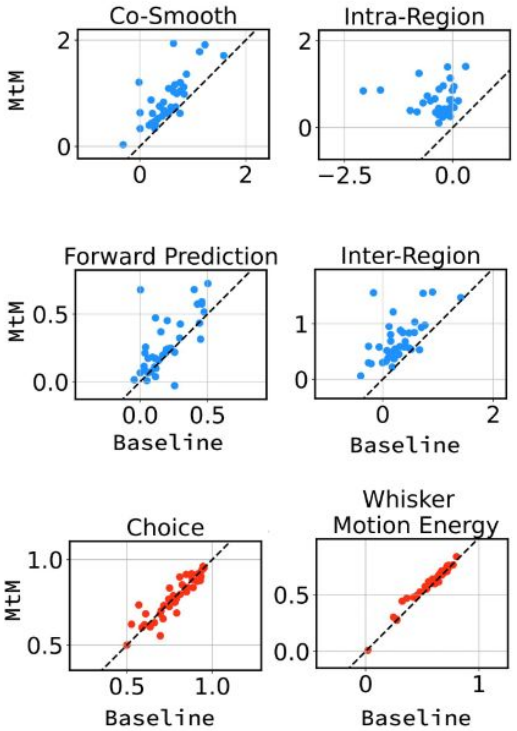


$$\text{bits/spike} = \frac{1}{n_{sp} \log 2} (\mathcal{L}(\boldsymbol{\lambda}; \hat{\boldsymbol{y}}_{n,t}) - \mathcal{L}(\bar{\boldsymbol{\lambda}}_{n,:}; \hat{\boldsymbol{y}}_{n,t}))$$

Single-session results for MtM vs. temporal masking

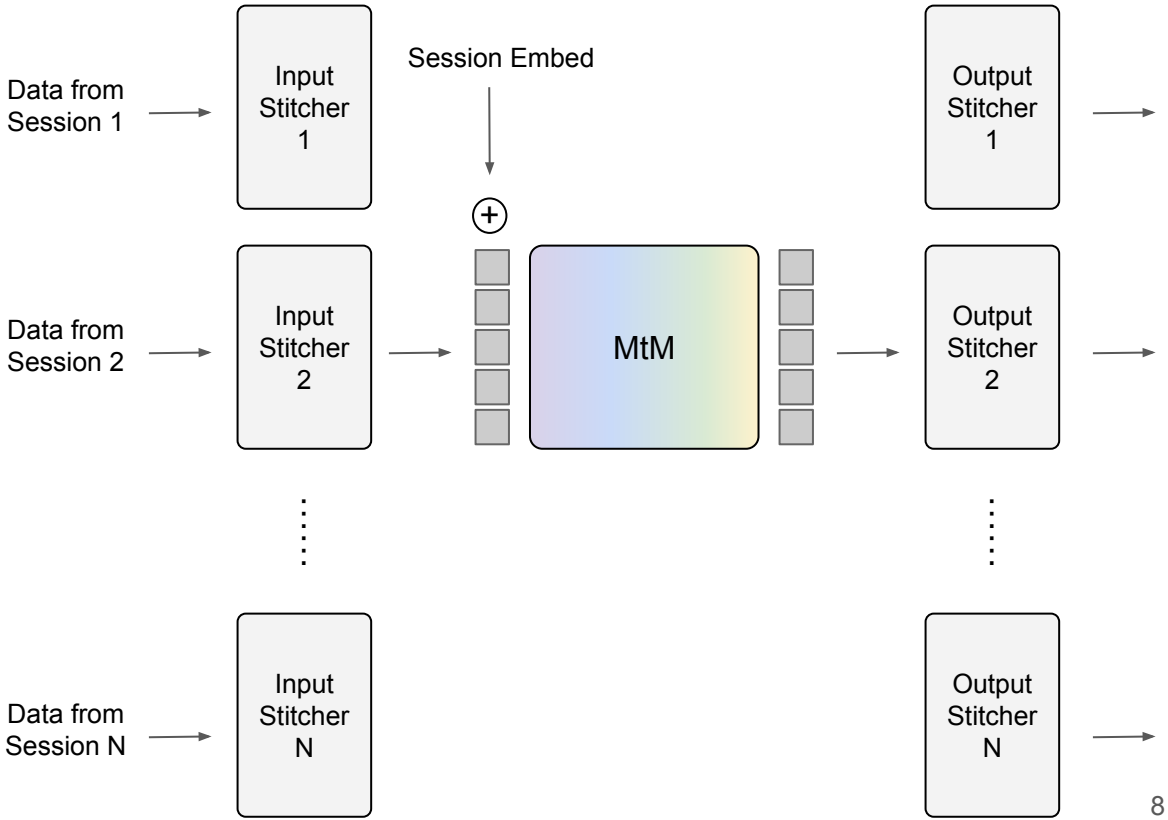
Masking scheme ablation

Masking	Activity Reconstruction			
	Co-Smooth	Forward Prediction	Intra-Region	Inter-Region
Temporal (Baseline)	0.84	0.42	-0.20	0.57
Neuron	**1.04**	-0.21	-0.22	0.78
Causal	0.44	0.48	-0.36	0.23
Intra-Region	-9.86	-2.97	0.32	-9.06
Inter-Region	0.92	0.01	-0.58	**0.90**
MtM (Not Prompted)	0.99	0.54	0.42	0.83
MtM (Prompted)	0.98	**0.57**	**0.43**	0.84



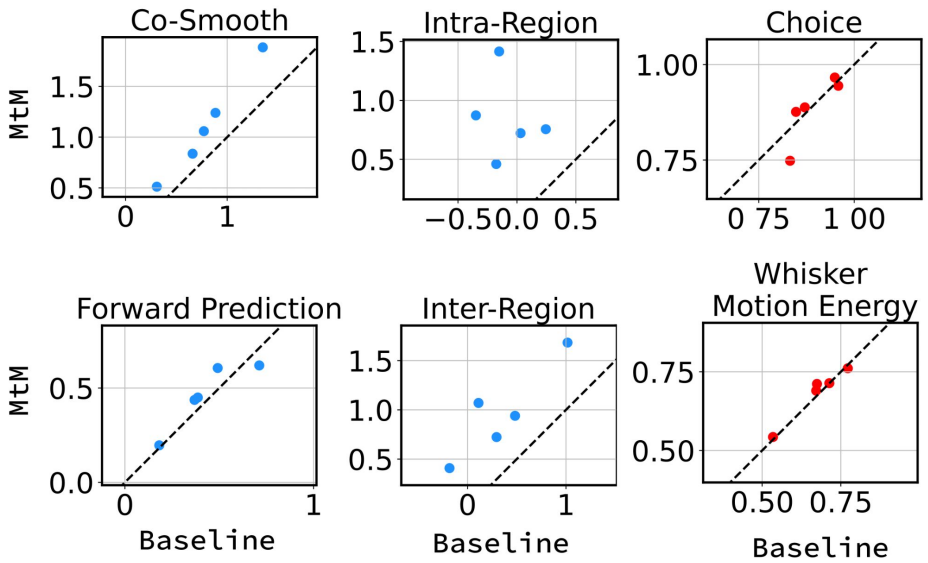
Pre-training MtM across many different animals + insertions

- Session-specific matrices (“stitchers”) map sessions with different numbers of neurons into latent representations with fixed dimensions.
- Session-embeddings are used to distinguish sessions.

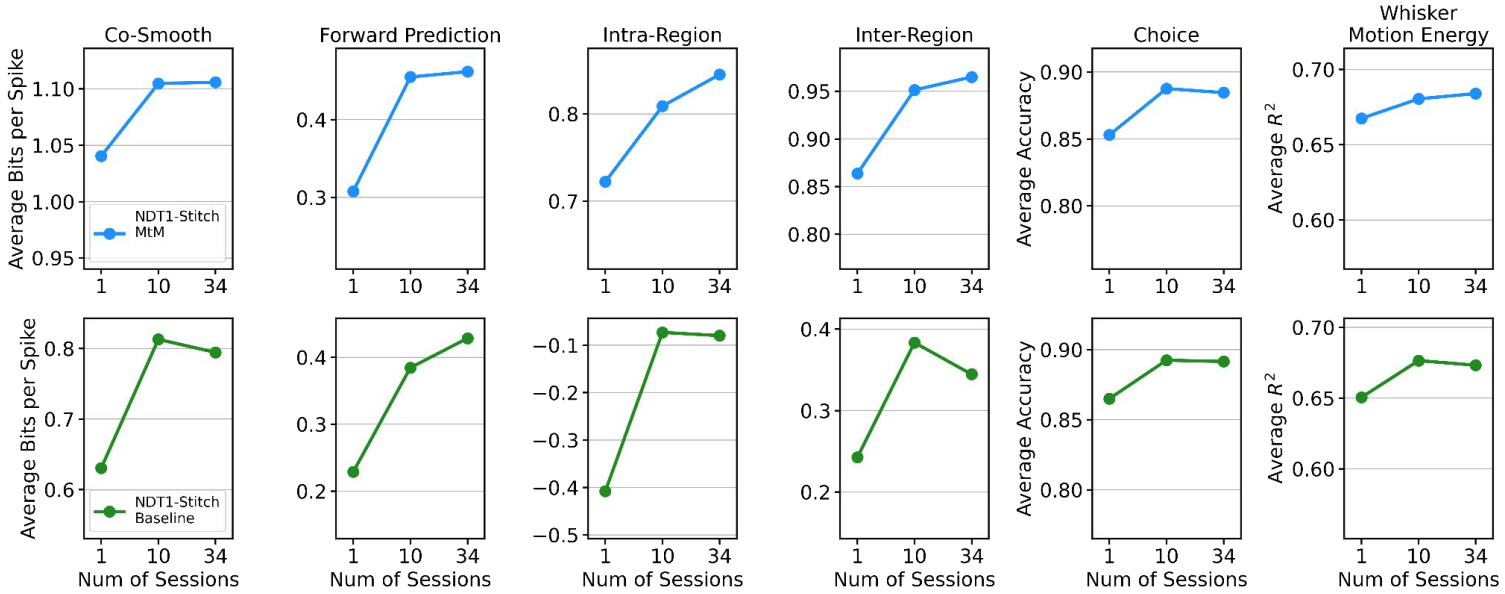


Generalization of pre-trained MtM to unseen animals

- We pre-train MtM and the temporal masking baseline on 34 animals.
- We fine-tune the final models on the training data from 5 unseen animals.



Scaling curves - MtM shows improvement after pretraining on more animals



COLUMBIA | Zuckerman Institute
MORTIMER B. ZUCKERMAN MIND BRAIN BEHAVIOR INSTITUTE

INTERNATIONAL BRAIN LABORATORY

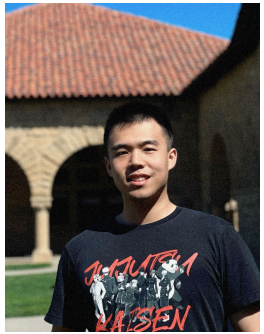
Artificial and Natural Intelligence

Special thanks to:

Cole Hurwitz

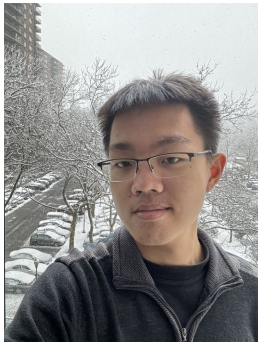
Liam Paninski

Yanchen Wang



Donato Jiménez-Benetó

Zixuan Wang



Mehdi Azabou

Eva Dyer

Blake Richards

Renee Tung

Olivier Winter

Thank you!