



BEST PAPERS 2023

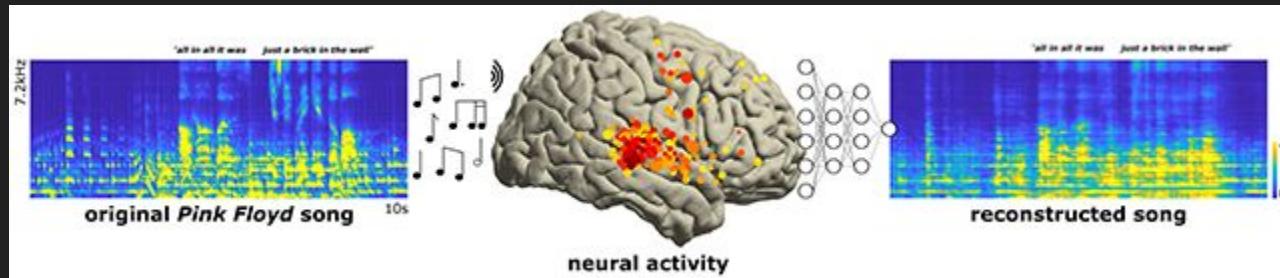
THE FUTURE OF INNOVATION

BESTENERT162925

Music can be reconstructed from human auditory cortex activity using nonlinear decoding models

Ludovic Bellier , Anaïs Llorens, Déborah Marciano, Aysegul Gunduz, Gerwin Schalk, Peter Brunner, Robert T. Knight 

Published: August 15, 2023 • <https://doi.org/10.1371/journal.pbio.3002176>



Exact learning dynamics of deep linear networks with prior knowledge*

Brad

Clémentine C J Dominé^{1,6}, Lukas Braun^{2,6},
James E Fitzgerald³ and Andrew M Saxe^{1,4,5,**}

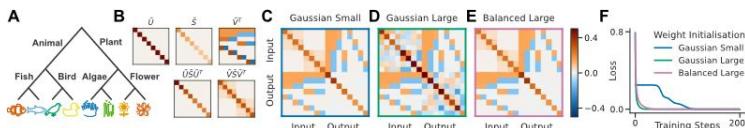


Figure 3. Rich and lazy learning. (A) Semantic learning task, (B) SVD of the input-output correlation of the task (top) and the respective RSMs (bottom). Rows and columns in the SVD and RSMs are identically ordered as the order of items in the hierarchical tree. (C) Final QQ^T matrices after training converged when initialised from random small weights, (D) random large weights (note how the upper left and lower right quadrant differ from the task's RSMs) and (E) large zero-balanced weights. (F) Learning curves for the three different initialisations as in (C) (green), (D) (pink) and (E) (blue). While both large weight initialisations lead to fast exponential learning curves, the small weight initialisation leads to a slow step-like decay of the loss.

Exact learning dynamics of deep linear networks with prior knowledge

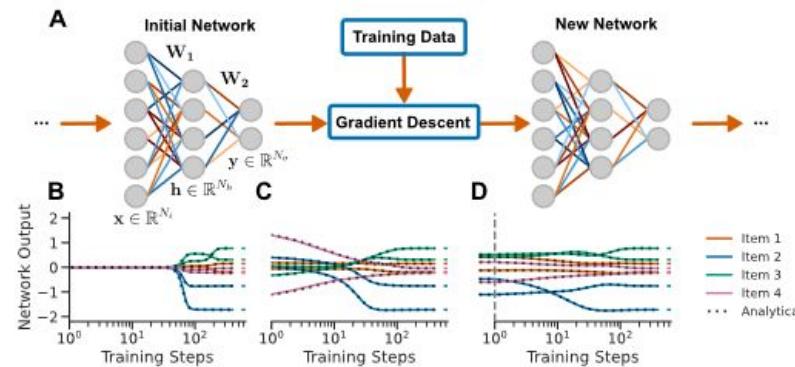


Figure 1. Learning with prior knowledge. (A) In our setting, a deep linear network with N_i input, N_h hidden and N_o output neurons is trained from a particular initialisation using gradient descent. (B)–(D) Network output for an example task over training time when starting from (B) small random weights, (C) large random weights, and (D) the weights of a previously learned task. The dynamics depend in detail on the initialisation. Solid lines indicate simulations, dotted lines indicate the analytical solutions we derive in this work.



Detecting abnormalities in resting-state dynamics: An unsupervised learning approach

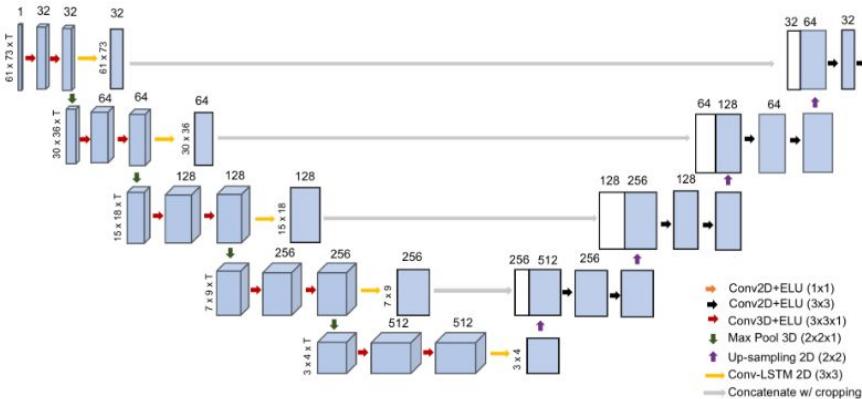


Fig. 1. Next frame prediction model. Each cuboid represents a 3D (2 spatial dimensions + time) feature map with number of features indicated on top. Flat boxes represent 2D feature maps, with number of channels on top. Input is an axial fMRI slice with T sequential frames. Conv-LSTM cell returns the last output of the output sequence.

Imputation models	Mean Squared Error	Pearson's Correlation
Last observation copy	0.01969	0.7558
Extrapolation	0.01203	0.8938
Interpolation*	0.00065	0.9939
Non-recurrent U-Net	0.00026	0.9967
Proposed recurrent U-Net	0.00007	0.9990

Table 1. Next frame prediction performance on healthy test subjects for different models. *Interpolation model had access to the frame after the predicted frame.

Recurrent autoencoder: sequence length	Mean squared error	Pearson's correlation
T=10 frames	0.0625	0.354
T=15 frames	0.0475	0.503
T=20 frames	0.0437	0.550

Table 2. Reconstruction performance of the proposed recurrent autoencoder on healthy test subjects for different input sequence lengths.



Thalamic deep brain stimulation in traumatic brain injury: a phase 1, randomized feasibility study

Received: 11 August 2023

Accepted: 10 October 2023

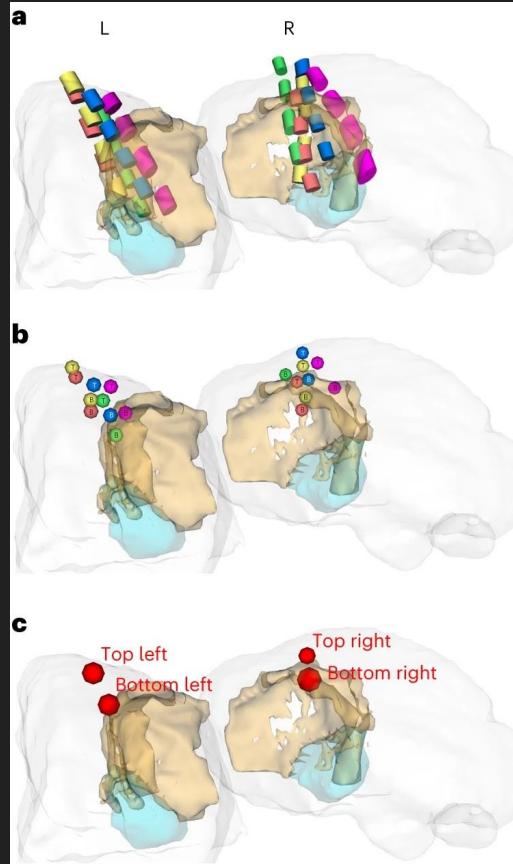
Published online: 04 December 2023

 Check for updates

Nicholas D. Schiff   ^{1,2}, Joseph T. Giacino ^{3,4}, Christopher R. Butson   ^{5,6}, Eun Young Choi  ⁷, Jonathan L. Baker  ¹, Kyle P. O'Sullivan ⁵, Andrew P. Janson ^{5,8}, Michael Bergin ³, Helen M. Bronte-Stewart ⁷, Jason Chua ⁹, Laurel DeGeorge ¹, Sureyya Dikmen ¹⁰, Adam Fogarty ⁷, Linda M. Gerber   ^{9,11}, Mark Krel ⁷, Jose Maldonado ¹², Matthew Radovan   ¹³, Sudhin A. Shah ¹⁴, Jason Su   ¹⁵, Nancy Temkin ¹⁶, Thomas Tourdias ¹⁷, Jonathan D. Victor   ^{1,2}, Abigail Waters ³, Stephanie A. Kolakowsky-Hayner ⁷, Joseph J. Fins   ¹¹, Andre G. Machado   ¹⁸, Brian K. Rutt   ^{19,20,21} & Jaimie M. Henderson   ^{7,20,21}

Converging evidence indicates that impairments in executive function and information-processing speed limit quality of life and social reentry after moderate-to-severe traumatic brain injury (msTBI). These deficits reflect dysfunction of frontostriatal networks for which the central lateral (CL) nucleus of the thalamus is a critical node. The primary objective of this feasibility study was to test the safety and efficacy of deep brain stimulation within the CL and the associated medial dorsal tegmental (CL/DTTm) tract. Six participants with msTBI, who were between 3 and 18 years post-injury, underwent surgery with electrode placement guided by imaging and subject-specific biophysical modeling to predict activation of the CL/DTTm tract. The primary efficacy measure was improvement in executive control indexed by processing speed on part B of the trail-making test. All six participants were safely implanted. Five participants completed the study and one was withdrawn for protocol non-compliance. Processing speed on part B of the trail-making test improved 15% to 52% from baseline, exceeding the 10% benchmark for improvement in all five cases. CL/DTTm deep brain stimulation can be safely applied and may improve executive control in patients with msTBI who are in the chronic phase of recovery. ClinicalTrials.gov identifier: [NCT02881151](https://clinicaltrials.gov/ct2/show/NCT02881151).

Isabelle



Are Transformers Effective for Time Series Forecasting?

Joanne

Transformers - Self attention leads to temporal information loss in long term time series forecasting (LTSF).

1-layer linear models outperform sophisticated transformers on LSTF tasks.

AAAI's #3 most influential paper of 2023 based on citations and granted patents

<https://ojs.aaai.org/index.php/AAAI/article/view/26317/26089>

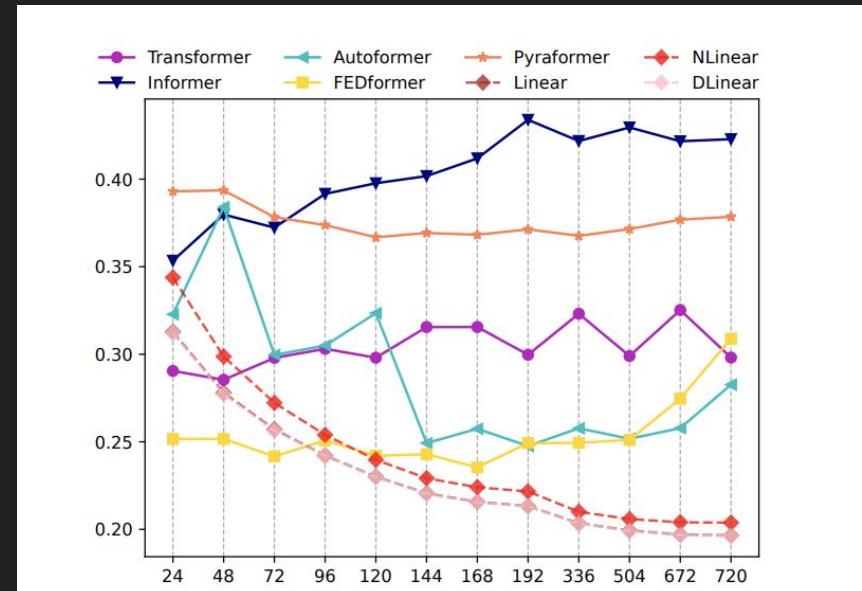
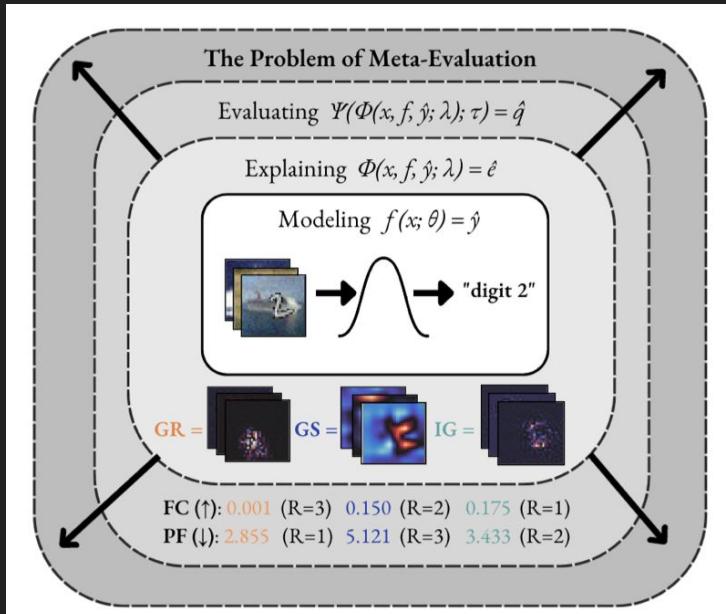


Figure 4: The MSE results (Y-axis) of models with different look-back window sizes (X-axis) of long-term forecasting ($T=720$) on Electricity.



The Meta-Evaluation Problem in Explainable AI: Identifying Reliable Estimators with MetaQuantus



To evaluate the quality of the explanations, we apply different estimators of faithfulness such as Faithfulness Correlation (FC) and Pixel-Flipping (PF), which return a correlation coefficient and an AUC score, respectively. However, since the scores vary depending on the estimator, both in range and direction, with lower or higher scores indicating more faithful explanations, interpreting the resulting faithfulness scores remains difficult for the practitioner.

[paper](#)

Transactions on Machine Learning
Research (06/2023)

VOYAGER: An Open-Ended Embodied Agent with Large Language Models

Guanzhi Wang^{1,2✉}, Yuqi Xie³, Yunfan Jiang^{4*}, Ajay Mandlekar^{1*},
 Chaowei Xiao^{1,5}, Yuke Zhu^{1,3}, Linxi “Jim” Fan^{1†✉}, Anima Anandkumar^{1,2†}

¹NVIDIA, ²Caltech, ³UT Austin, ⁴Stanford, ⁵UW Madison

*Equal contribution †Equal advising ✉ Corresponding authors

<https://voyager.minedojo.org>

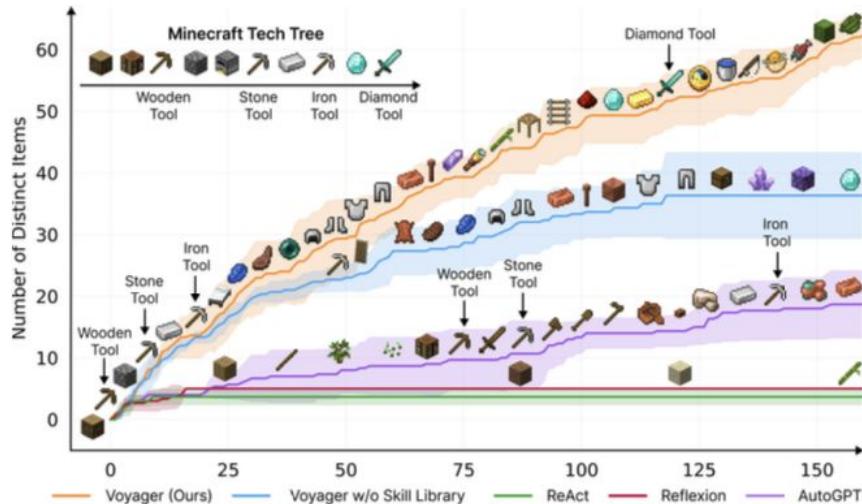


Figure 1: VOYAGER discovers new Minecraft items and skills continually by self-driven exploration, significantly outperforming the baselines. X-axis denotes the number of prompting iterations.



Sparks of Artificial General Intelligence: Early experiments with GPT-4

Sébastien Bubeck Varum Chandrasekaran Ronen Eldan Johannes Gehrke
Eric Horvitz Ece Kamar Peter Lee Yin Tat Lee Yuanzhi Li Scott Lundberg
Harsha Nori Hamid Palangi Marco Tulio Ribeiro Yi Zhang

Microsoft Research

Abstract

Artificial intelligence (AI) researchers have been developing and refining large language models (LLMs) that exhibit remarkable capabilities across a variety of domains and tasks, challenging our understanding of learning and cognition. The latest model developed by OpenAI, GPT-4 [Ope23], was trained using an unprecedented scale of compute and data. In this paper, we report on our investigation of an early version of GPT-4, when it was still in active development by OpenAI. We contend that (this early version of) GPT-4 is part of a new cohort of LLMs (along with ChatGPT and Google's PaLM for example) that exhibit more general intelligence than previous AI models. We discuss the rising capabilities and implications of these models. We demonstrate that, beyond its mastery of language, GPT-4 can solve novel and difficult tasks that span mathematics, coding, vision, medicine, law, psychology and more, without needing any special prompting. Moreover, in all of these tasks, GPT-4's performance is strikingly close to human-level performance, and often vastly surpasses prior models such as ChatGPT. Given the breadth and depth of GPT-4's capabilities, we believe that it could reasonably be viewed as an early (yet still incomplete) version of an artificial general intelligence (AGI) system. In our exploration of GPT-4, we put special emphasis on discovering its limitations, and we discuss the challenges ahead for advancing towards deeper and more comprehensive versions of AGI, including the possible need for pursuing a new paradigm that moves beyond next-word prediction. We conclude with reflections on societal influences of the recent technological leap and future research directions.

<https://arxiv.org/abs/2303.12712>



Semi-Supervised Domain Adaptation with Source Label Adaptation

Yu-Chu Yu Hsuan-Tien Lin
 National Taiwan University

{r09922104, htlin}@csie.ntu.edu.tw

Abstract

Semi-Supervised Domain Adaptation (SSDA) involves learning to classify unseen target data with a few labeled and lots of unlabeled target data, along with many labeled source data from a related domain. Current SSDA approaches usually aim at aligning the target data to the labeled source data with feature space mapping and pseudo-label assignments. Nevertheless, such a source-oriented model can sometimes align the target data to source data of the wrong classes, degrading the classification performance. This paper presents a novel source-adaptive paradigm that adapts the source data to match the target data. Our key idea is to view the source data as a noisily-labeled version of the ideal target data. Then, we propose an SSDA model that cleans up the label noise dynamically with the help of a robust cleaner component designed from the target perspective. Since the paradigm is very different from the core ideas behind existing SSDA ap-

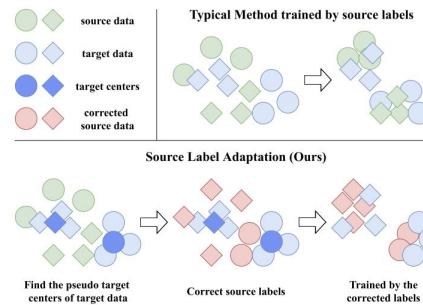


Figure 1. **Top.** Training the model with the original source labels might lead to the misalignment of the target data. **Bottom.** After cleaning up the noisy source labels with our SLA framework, the target data can be aligned with the correct classes.



METHODOLOGY ARTICLE

Open Access



Single-cell spatial proteomic imaging for human neuropathology

Kausalia Vijayaragavan^{1†}, Bryan J. Cannon^{1†}, Dmitry Tebaykin¹, Marc Bossé¹, Alex Baranski¹, J. P. Oliveria¹, Syed A. Bukhari¹, Dunja Mrdjen¹, M. Ryan Corces³, Erin F. McCaffrey¹, Noah F. Greenwald¹, Yari Sigal², Diana Marquez¹, Zumana Khair¹, Trevor Bruce¹, Mako Goldston¹, Anusha Bharadwaj¹, Kathleen S. Montine¹, R. Michael Angelo¹, Thomas J. Montine¹ and Sean C. Bendall^{1*}

Abstract:

Neurodegenerative disorders are characterized by phenotypic changes and hallmark proteopathies. Quantifying these in archival human brain tissues remains indispensable for validating animal models and understanding disease mechanisms. We present a framework for nanometer-scale, spatial proteomics with multiplex ion beam imaging (MIBI) for capturing neuropathological features. MIBI facilitated simultaneous, quantitative imaging of 36 proteins on archival human hippocampus from individuals spanning cognitively normal to dementia. Customized analysis strategies identified cell types and proteopathies in the hippocampus across stages of Alzheimer's disease (AD) neuropathologic change. We show microglia-pathologic tau interactions in hippocampal CA1 subfield in AD dementia. Data driven, sample independent creation of spatial proteomic regions identified persistent neurons in pathologic tau neighborhoods expressing mitochondrial protein MFN2, regardless of cognitive status, suggesting a survival advantage. Our study revealed unique insights from multiplexed imaging and data-driven approaches for neuropathologic analysis and serves broadly as a methodology for spatial proteomic analysis of archival human neuropathology.

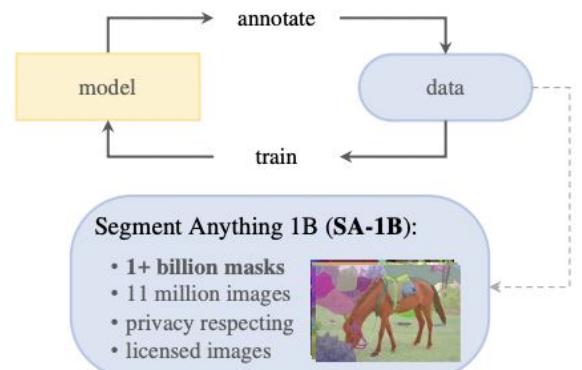
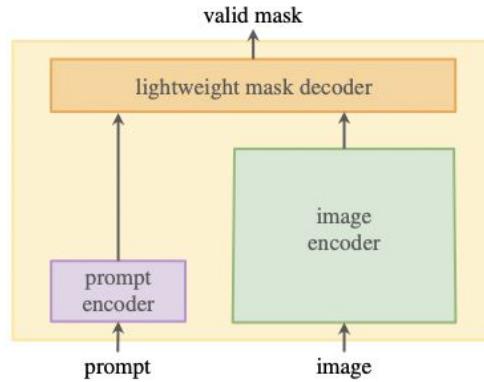
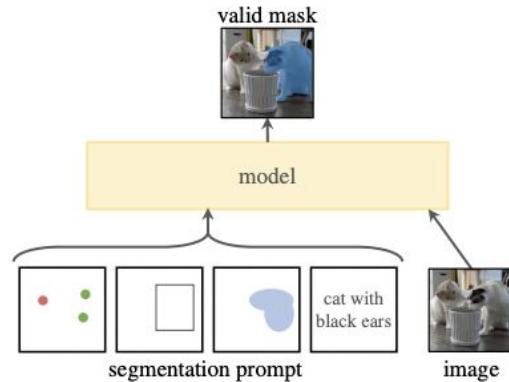
Teaser: Multiplex Ion beam Imaging enables deep spatial phenotyping of human neuropathology-associated cellular and disease features.



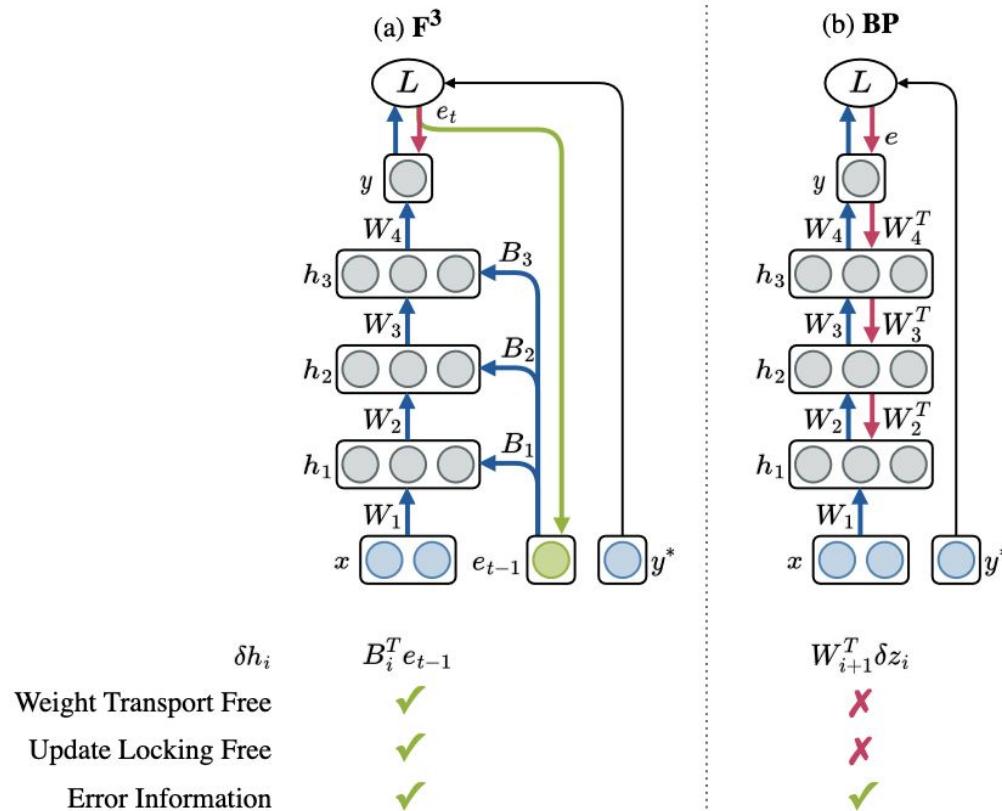
Segment Anything

Alexander Kirillov^{1,2,4} Eric Mintun² Nikhila Ravi^{1,2} Hanzi Mao² Chloe Rolland³ Laura Gustafson³
 Tete Xiao³ Spencer Whitehead Alexander C. Berg Wan-Yen Lo Piotr Dollár⁴ Ross Girshick⁴
¹project lead ²joint first author ³equal contribution ⁴directional lead

Meta AI Research, FAIR

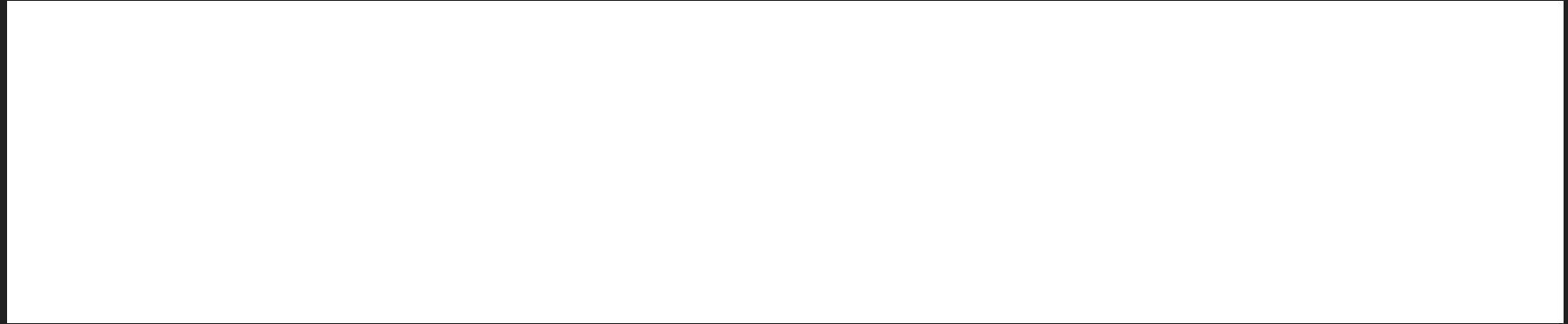


Feed-Forward Optimization With Delayed Feedback for Neural Networks



Riyasat

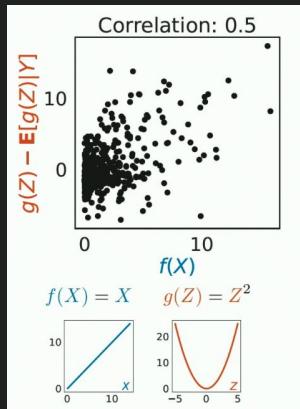
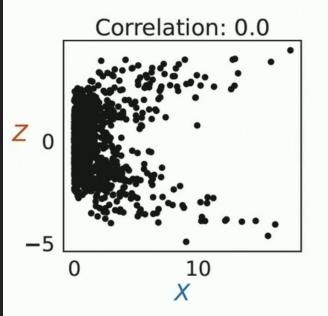
Sparse Mixture of Experts (MOE)



Efficient Conditionally Invariant Representation Learning

Sajad

$X \perp\!\!\!\perp Z | Y$ if and only if (per Daudin, 1980)
all $f(X)$ and $g(Z) - \mathbb{E}[g(Z) | Y]$ are uncorrelated

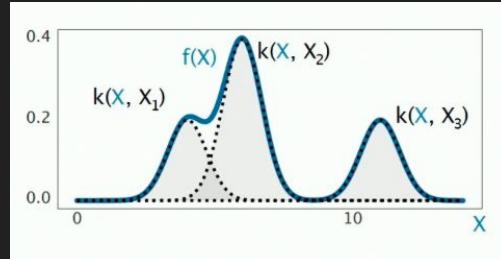


①

②

Given observations $\{\mathbf{X}_i, \mathbf{Z}_i, Y_i\}_{i=1}^B$,

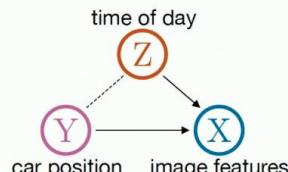
$$\widehat{\text{CIRCE}} = \frac{1}{B^2} \text{Tr}(KL)$$



CIRCE: as conditional independence regularizer

- Learn features \mathbf{X}_θ for predicting Y s.t. $\mathbf{X}_\theta \perp\!\!\!\perp Z | Y$:

$$\min_{\theta} L(\mathbf{X}_\theta, Y) + \lambda \underbrace{\frac{1}{B^2} \text{Tr}(K_\theta \mathbf{L})}_{\text{CIRCE}}$$



$$C_{QK}^{h \in H_2} = \left(\text{Id} \otimes \text{Id} \otimes W_E^T + \sum_{h_q \in H_1} A^{h_q} \otimes \text{Id} \otimes (W_{OV}^{h_q} W_E)^T \right)$$



The “query side” residual stream at the start of the second layer contains both the layer 1

$$\cdot \text{Id} \otimes \text{Id} \otimes W_{QK}^h \cdot$$



W_{QK} of the second layer

$$\left(\text{Id} \otimes \text{Id} \otimes W_E + \sum_{h_k \in H_1} \text{Id} \otimes A^{h_k} \otimes W_{OV}^{h_k} W_E \right)$$



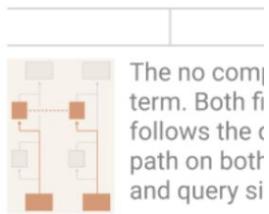
The “key side” residual stream at the start of the second layer contains both the layer 1

Progress measures for grokking via mechanistic interpretability

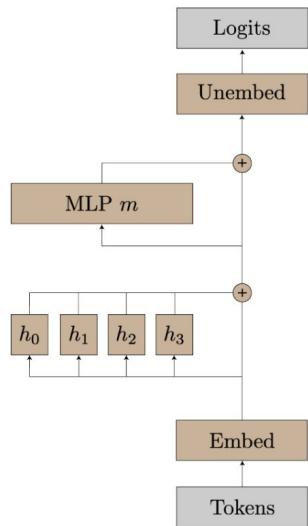
Neel Nanda, Lawrence Chan Tom Lieberum, Jess Smith, Jacob Steinhardt, ICLR 2023

Published as a conference paper at ICLR 2023

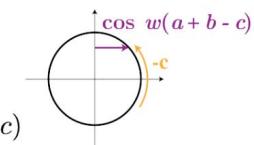
$$= \text{Id} \otimes \text{Id} \otimes (W_E^T W_C)$$



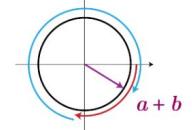
$$+ \sum_{h_k \in H_1} \text{Id} \otimes A^{h_k} \otimes$$



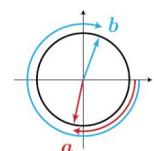
Computes logits using further trig identities:
 $\text{Logit}(c) \propto \cos(w(a+b-c))$
 $= \cos(w(a+b)) \cos(wc) + \sin(w(a+b)) \sin(wc)$



Calculates sine and cosine of $a+b$ using trig identities:
 $\sin(w(a+b)) = \sin(wa) \cos(wb) + \cos(wa) \sin(wb)$
 $\cos(w(a+b)) = \cos(wa) \cos(wb) - \sin(wa) \sin(wb)$



Translates one-hot a, b to Fourier basis:
 $a \rightarrow \sin(wa), \cos(wa)$
 $b \rightarrow \sin(wb), \cos(wb)$



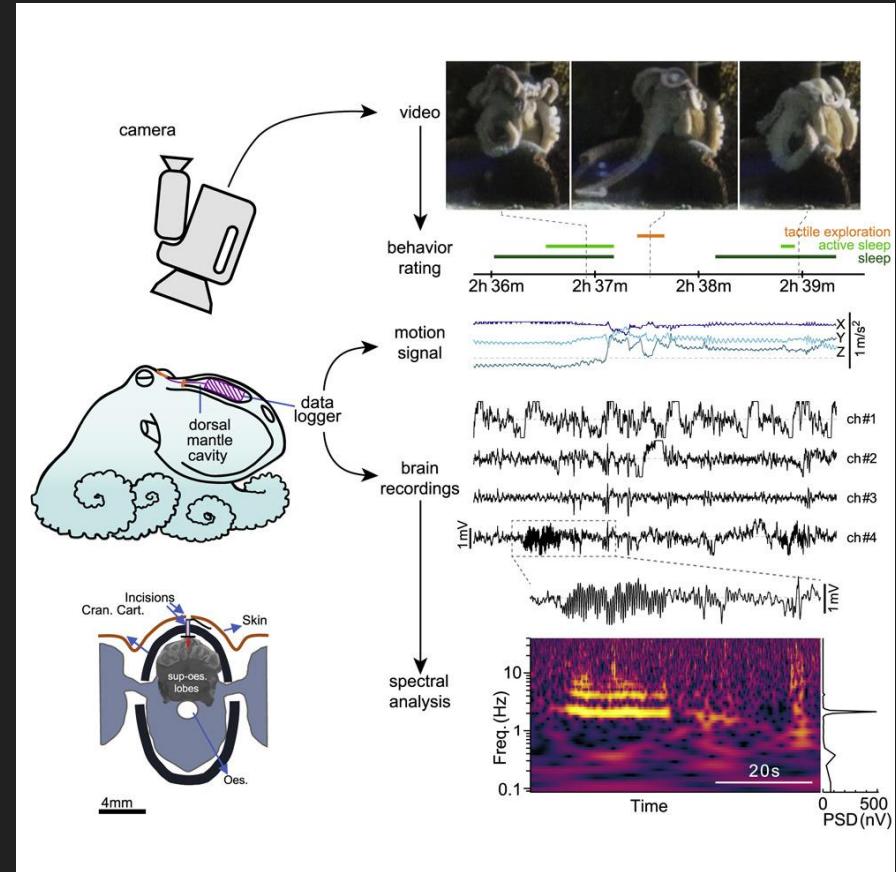
Thu

Recording electrical activity from the brain of behaving octopus

Gutnick, T. et al. *Curr. Biol.*

<https://doi.org/10.1016/j.cub.2023.02.006> (2023)

Observing the brain waves of octopus
(the most intelligent invertebrate with
too many limbs) for the first time

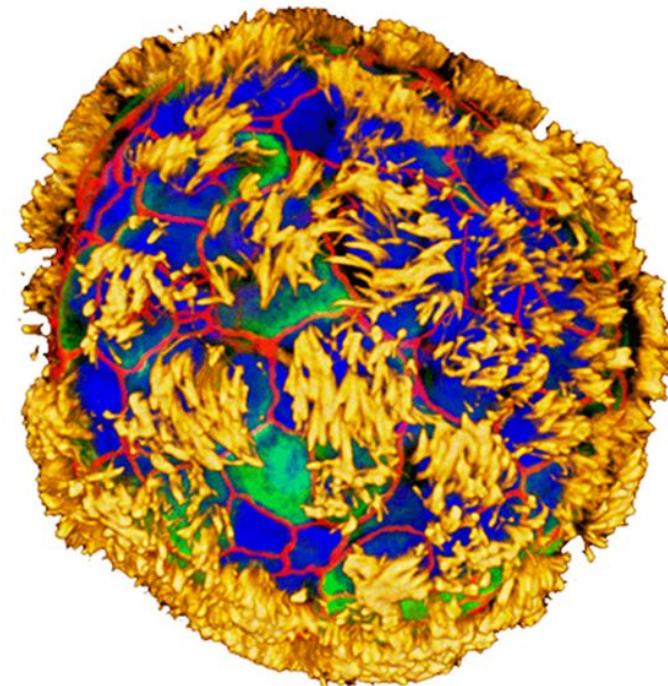


Tiny 'anthrobots' built from human cells could help heal the body

Self-propelled organoids repair nerve tissue in lab, could one day ferret out disease or deliver drugs

30 NOV 2023 • 3:00 PM ET • BY ELIZABETH PENNISI

In the medicine of the future, molecular physicians built from a patient's own cells might ferret out cancer, repair injured tissue, and even remove plaque from blood vessels. Researchers have now taken a step toward that vision: They've coaxed tracheal cells to form coordinated groups called organoids that can propel themselves with tiny appendages. When added to wounded neurons in the lab, [these](#)



[Ads in Science](#)

[The paper](#)

Scaling up GANs for Text-to-Image Synthesis

Minguk Kang^{1,3} Jun-Yan Zhu² Richard Zhang³
 Jaesik Park¹ Eli Shechtman³ Sylvain Paris³ Taesung Park³

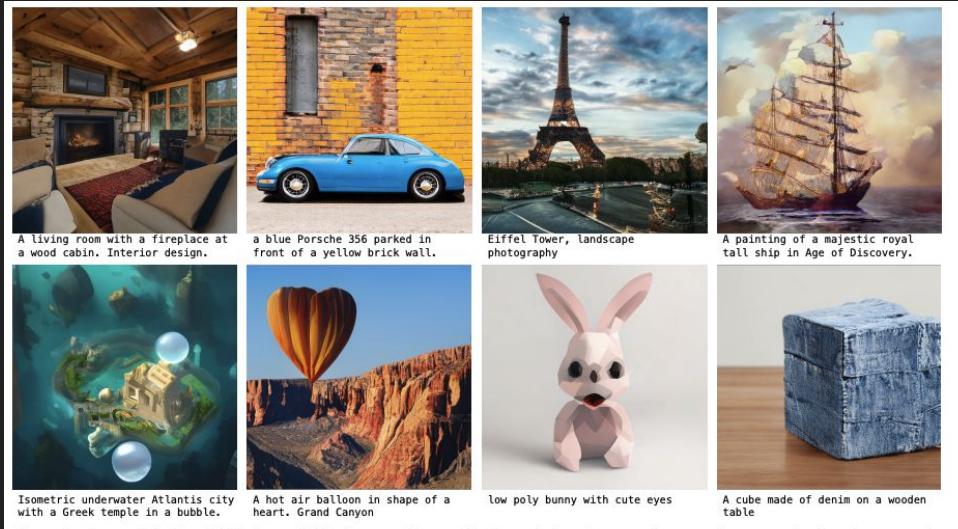


Figure 1. Our model, GigaGAN, shows GAN frameworks can also be scaled up for general text-to-image synthesis tasks, generating a 512px output at an interactive speed of 0.13s, and 4096px at 3.7s. Selected examples at 2K or 4K resolutions are shown. Please zoom in for more details. See Appendix C and our [website](#) for more uncurated comparisons.



How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection.

ChatGPT has changed the world so much. It's safe to say that the trend would go upward from here as the public is already in favor of using ChatGPT. However, how is the ChatGPT current result compared to the Human Experts? It's exactly a question that Guo et al. (2023) try to answer. The team tried to collect data from experts and ChatGPT prompt results, which they compared. The result shows that implicit differences between ChatGPT and experts were there. The research is something that I feel would be kept asked in the future as the generative AI model would keep growing over time.



On over-squashing in MPNNs: The Impact of Width, Depth, and Topology - William

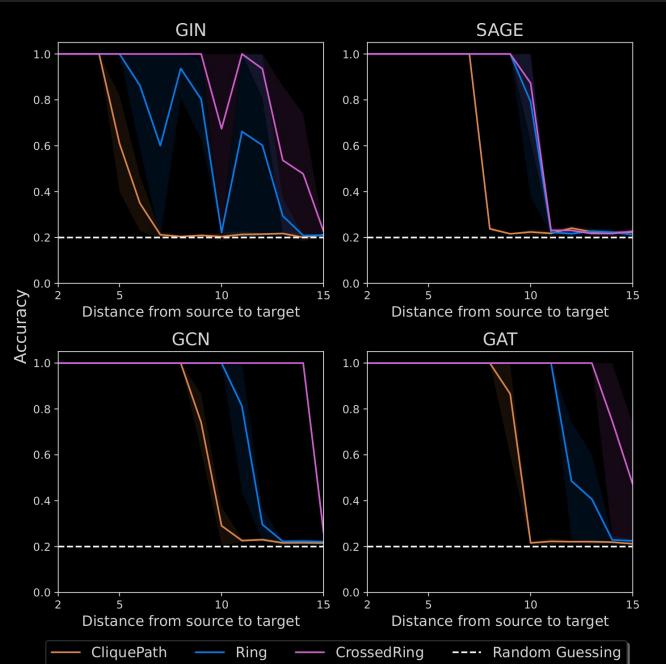
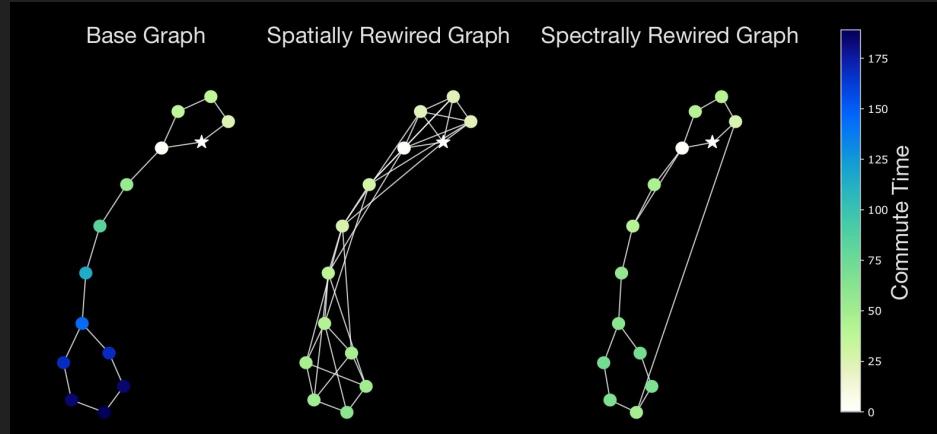
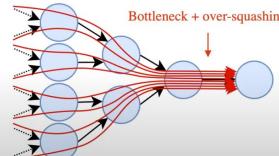


Figure 3. Performance of GIN, SAGE, GCN, and GAT on the CliquePath, Ring, and CrossedRing tasks. In the case where depth and distance are comparable, over-squashing highly depends on the topology of the graph as we increase the distance.



Over-squashing

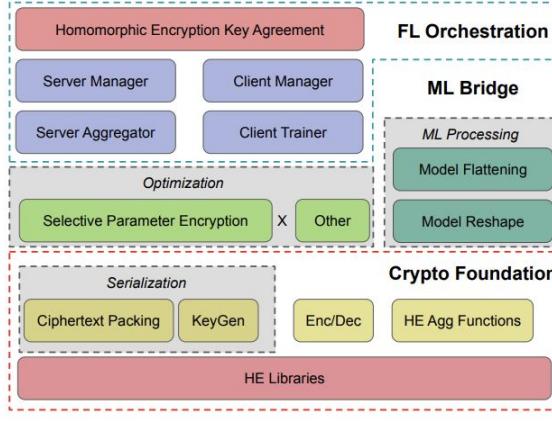
To flow a message to a distance of 4, we need to squash $O(\text{degree}^4)$ messages into a single node representation (the representation of the target node).



An exponential amount of information is squashed into a fixed-size vector.



Takeaway: Decrease commute time without inducing oversquashing



FEDML-HE: AN EFFICIENT HOMOMORPHIC-ENCRYPTION-BASED PRIVACY-PRESERVING FEDERATED LEARNING SYSTEM

Weizhao Jin ^{*1} Yuhang Yao ^{*2} Shanshan Han³ Carlee Joe-Wong² Srivatsan Ravi¹ Salman Avestimehr⁴
Chaoyang He⁴

	Model Degradation	Overheads	Client Dropout	Interactive Sync	Aggregated Model Visible To Server
Differential Privacy	With noise	Light	Robust	No	Yes
Secure Aggregation	Exact	Medium	Susceptible	Yes	Yes
Homomorphic Encryption	Exact	Large	Robust	No	No

Table 1: Comparison of Differential Privacy, Secure Aggregation, and Homomorphic Encryption

Features	IBMFL	Nvidia FLARE	Ours
Homomorphic Encryption	✓	✓	✓
Threshold Key Management	✗	✗	✓
Selective Parameter Encryption	✗	○	✓
Encrypted Foundation Model Training	○	○	✓

Table 2: Comparison with Existing HE-Based FL Systems. ○ implies limited support: for Selective Parameter Encryption, FLARE offers the (random) partial encryption option which does not have clear indications on privacy impacts; for Encrypted Foundation Model Training, the other two platforms require massive resources to train foundation models in encrypted federated learning.

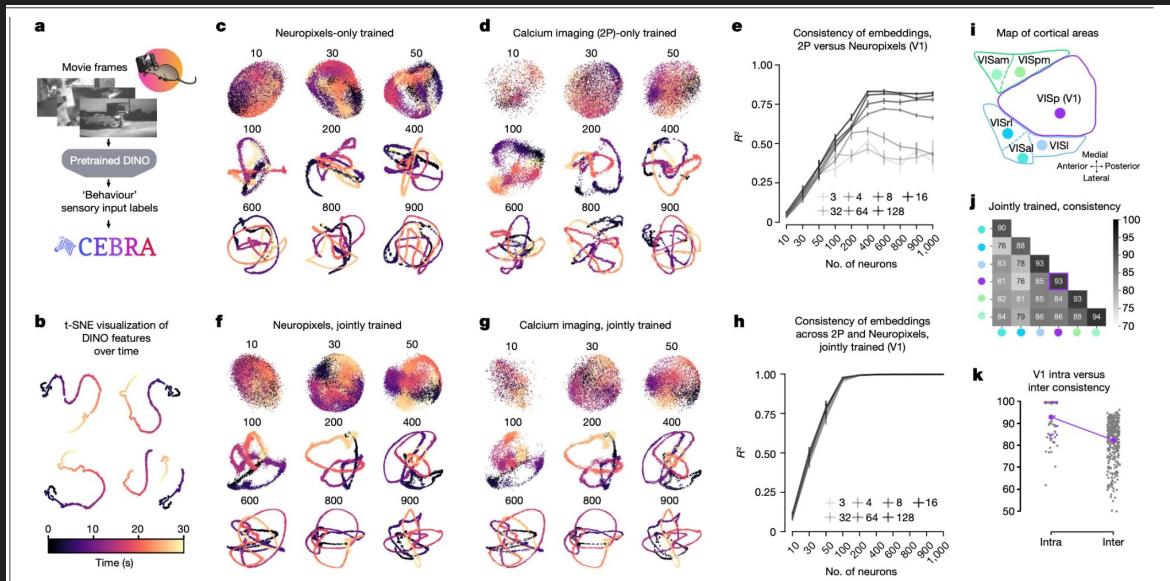
The paper presents FedML-HE, a practical system for secure Federated Learning (FL) using homomorphic encryption (HE). By encrypting only sensitive parts of the model, it significantly cuts down the computational burden, especially for big models like ResNet-50 and BERT. This approach reduces overhead by up to 10x for ResNet-50 and up to 40x for BERT, making FL with strong privacy protection more feasible. FedML-HE marks a crucial step toward using FL widely across various fields while safeguarding user privacy.

Article

Learnable latent embeddings for joint behavioural and neural analysis

<https://doi.org/10.1038/s41586-023-06031-6>

Steffen Schneider^{1,2}, Jin Hwa Lee^{1,2} & Mackenzie Weygandt Mathis¹✉



Donghyun

