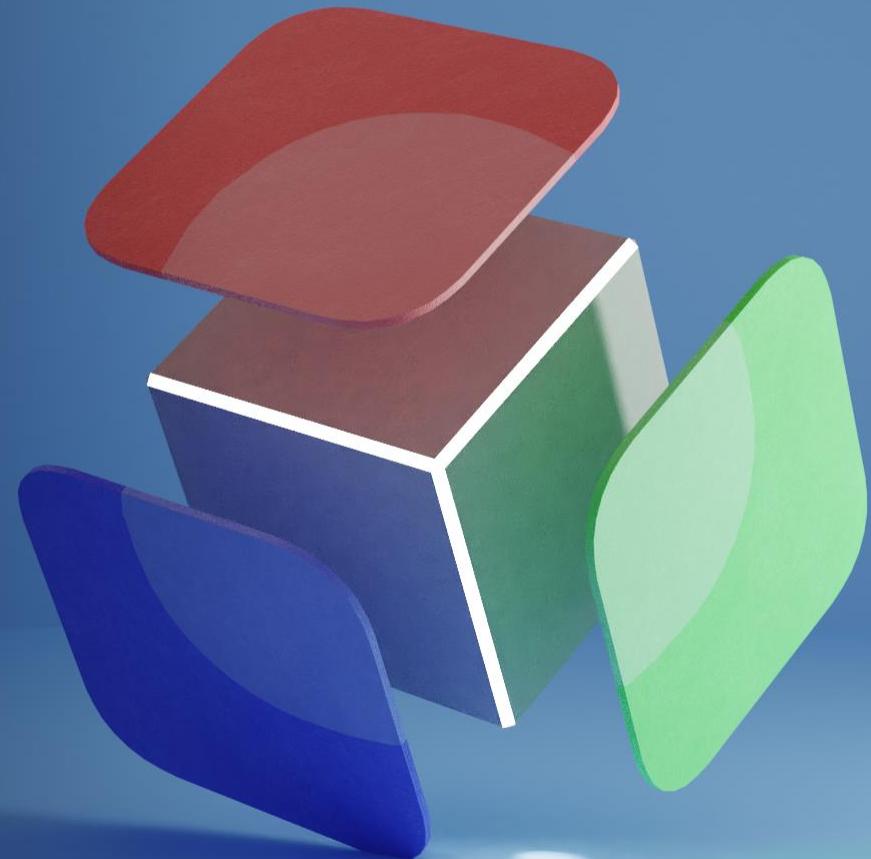


Perth Machine Learning Group

**Current state of AI and
where we are going
with PMLG**

2023

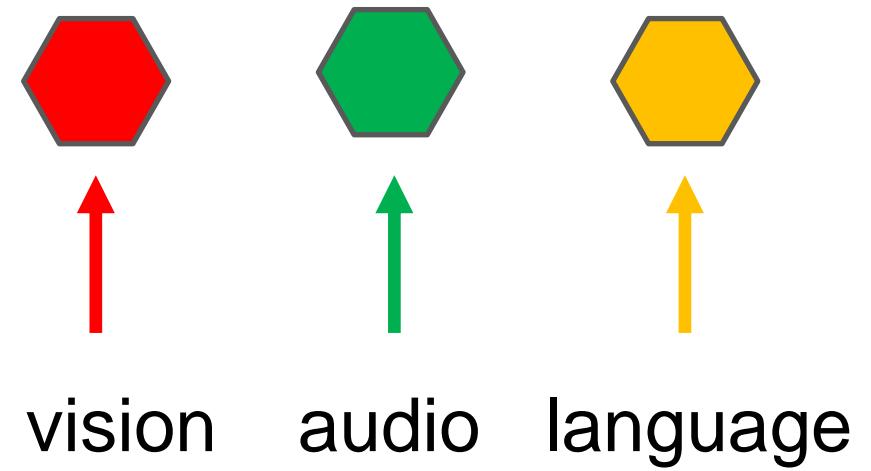




**AI ..
where are
we currently**

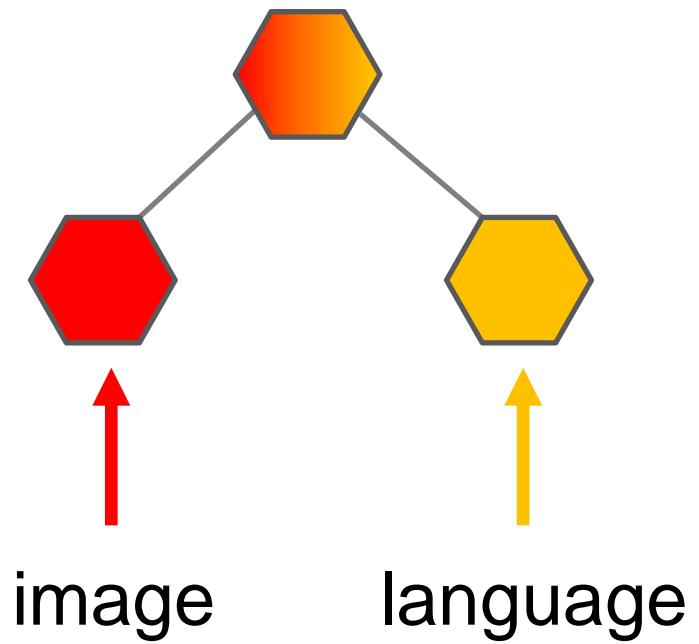


Unimodal



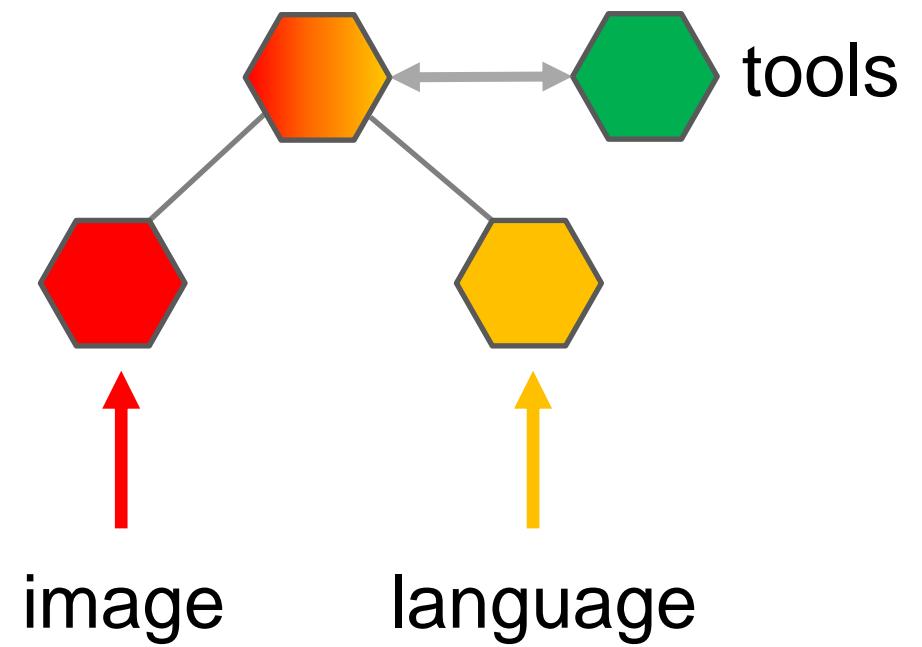


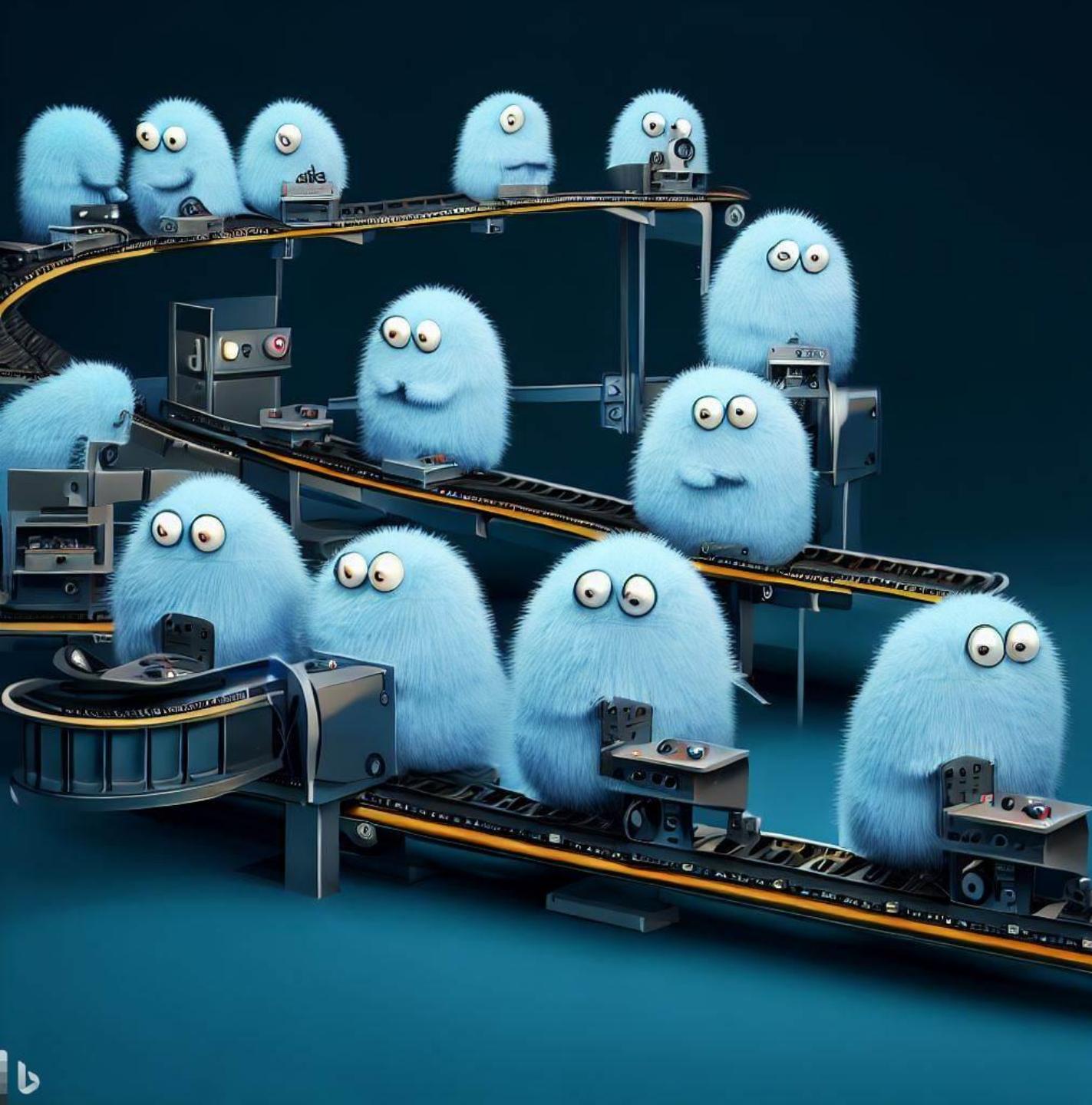
multimodal





Tool use



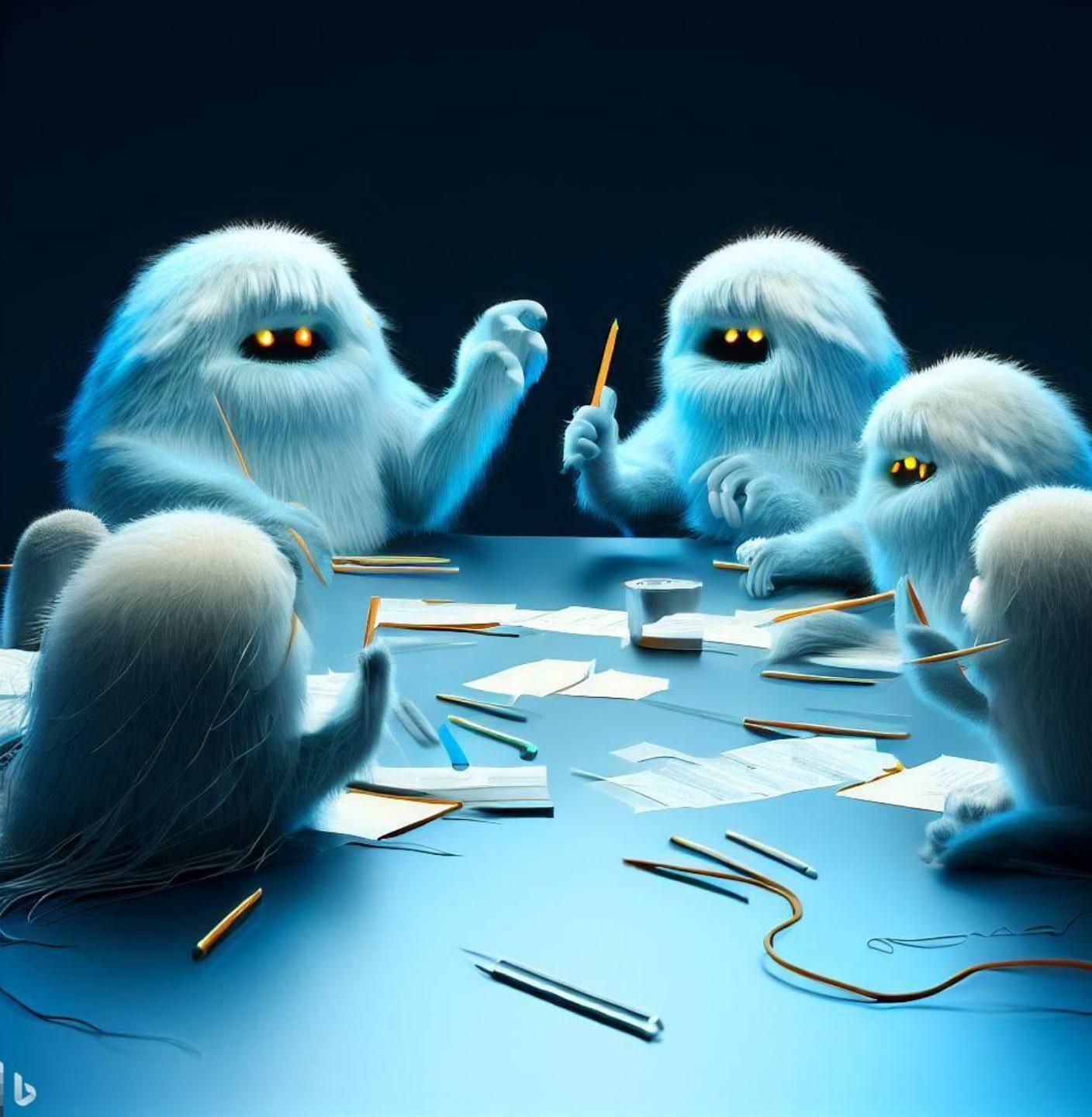


**knowledge based
AI automation**



**expanding
memory, tasks
management,
and planning**

..to make agents



systems of systems cognitive architecture





Executive planning



**how to connect
them is still an
open question**



Differentiable Neural Computer,
Recursive RvNN,
Neural Theorem Provers

Monte Carlo Tree Search,
World Models, etc.

Convolutional Networks (CNs),
Transformer-based LLM, etc.

Temporal Convolution networks,
Transformer-based LLM

Graph Convolutional Networks,
Knowledge Graph Embeddings,
Transformer-based LLM

Hierarchical Deep RL,
Dynamic Skill Acquisition



**language
models**

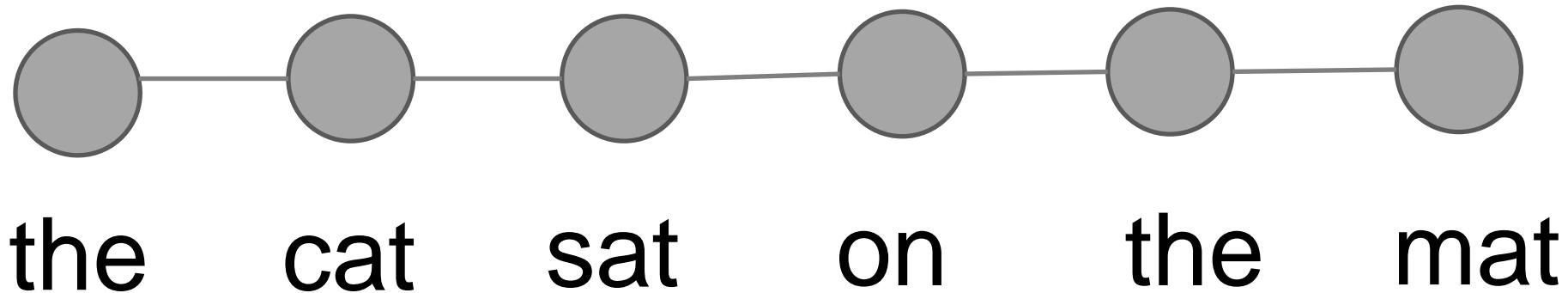


Transformers

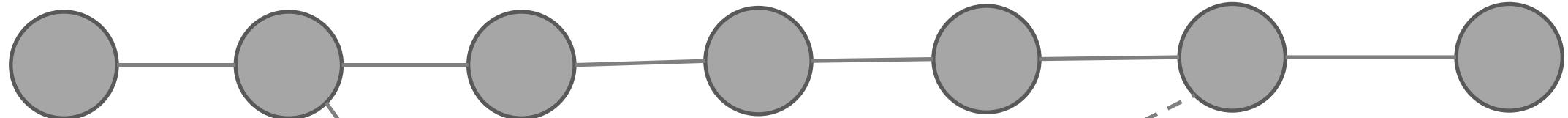


Hallucination

Language as a graph

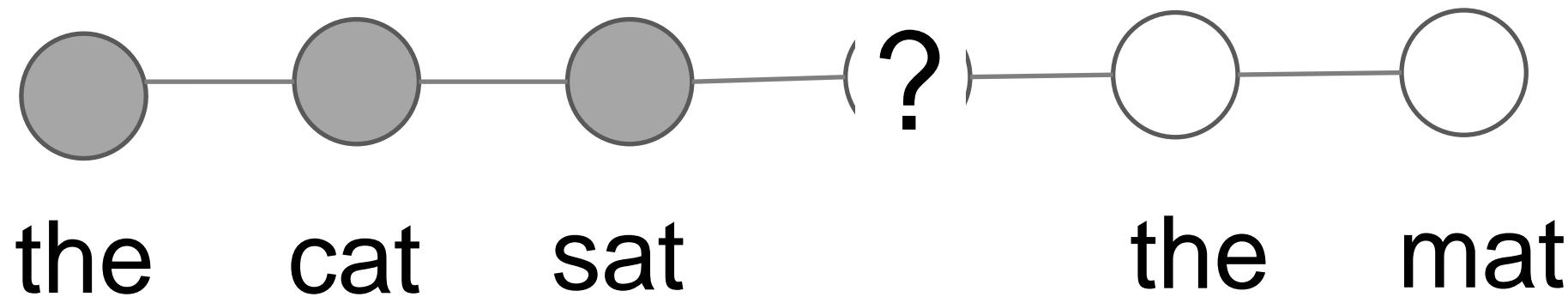


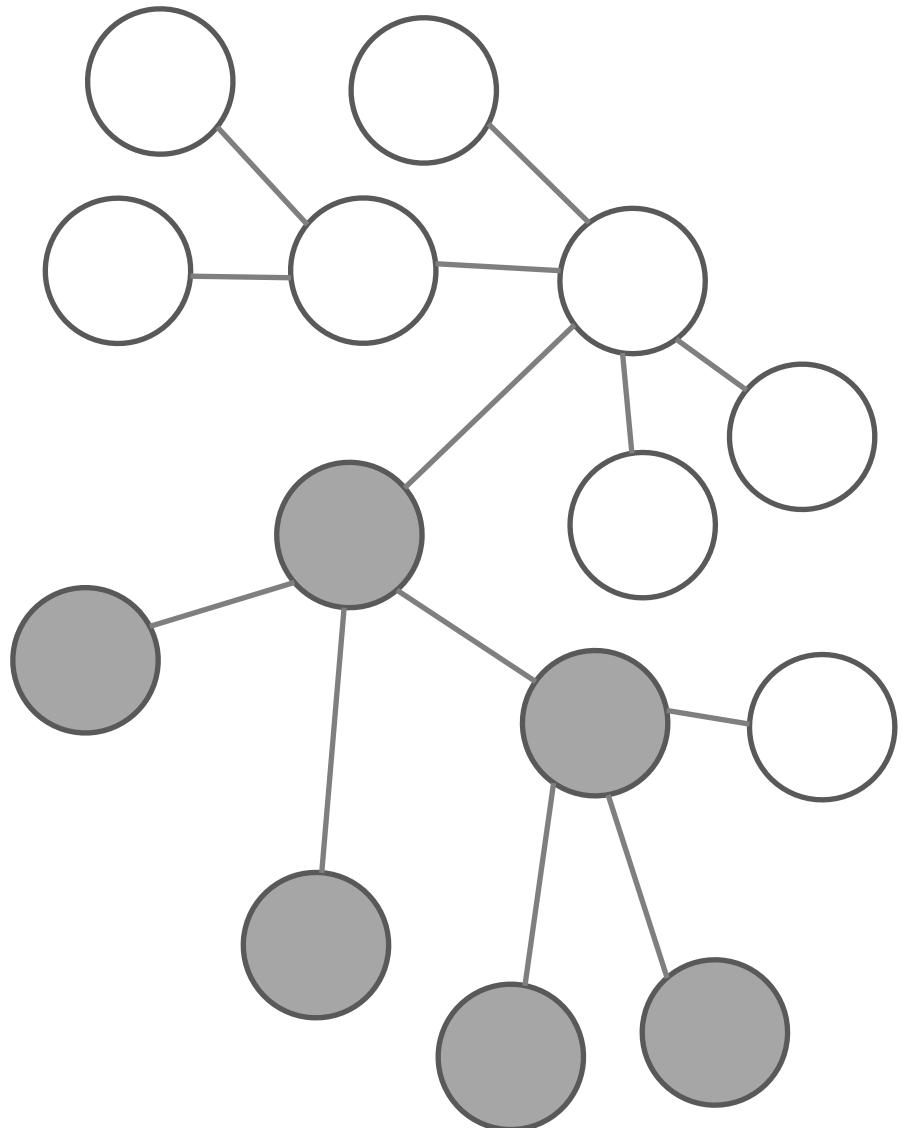
the cat sat on the mat ,



it was cold

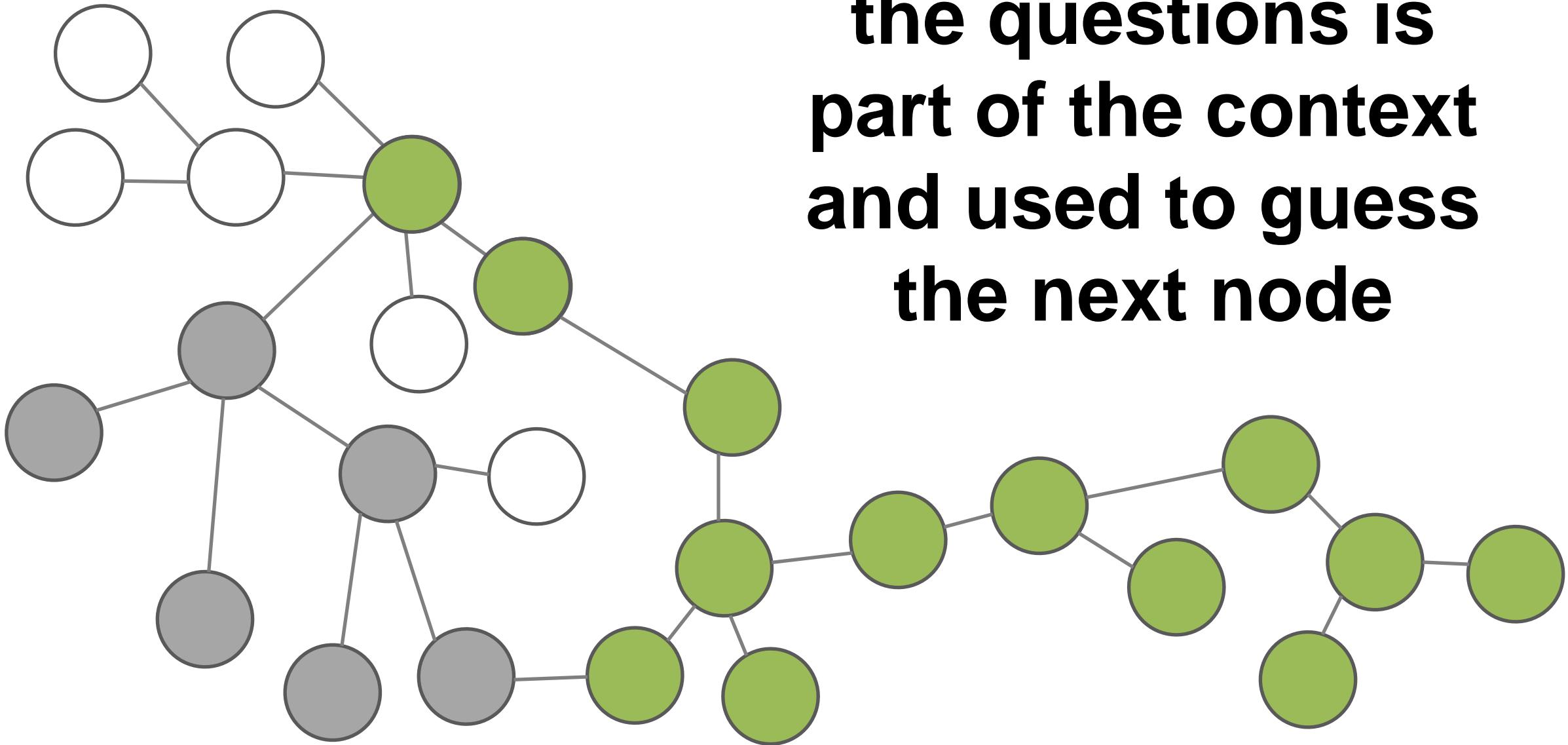
Language models trained to guess next node in graph





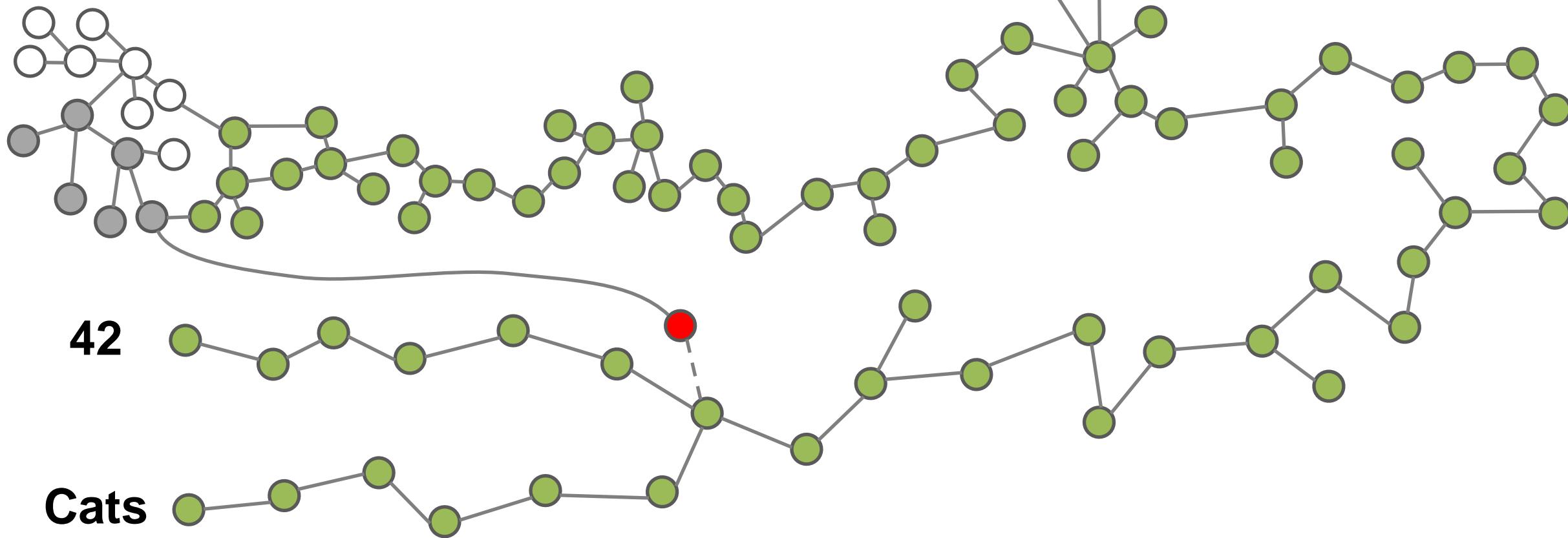
**Questions are
partially observed
graphs**

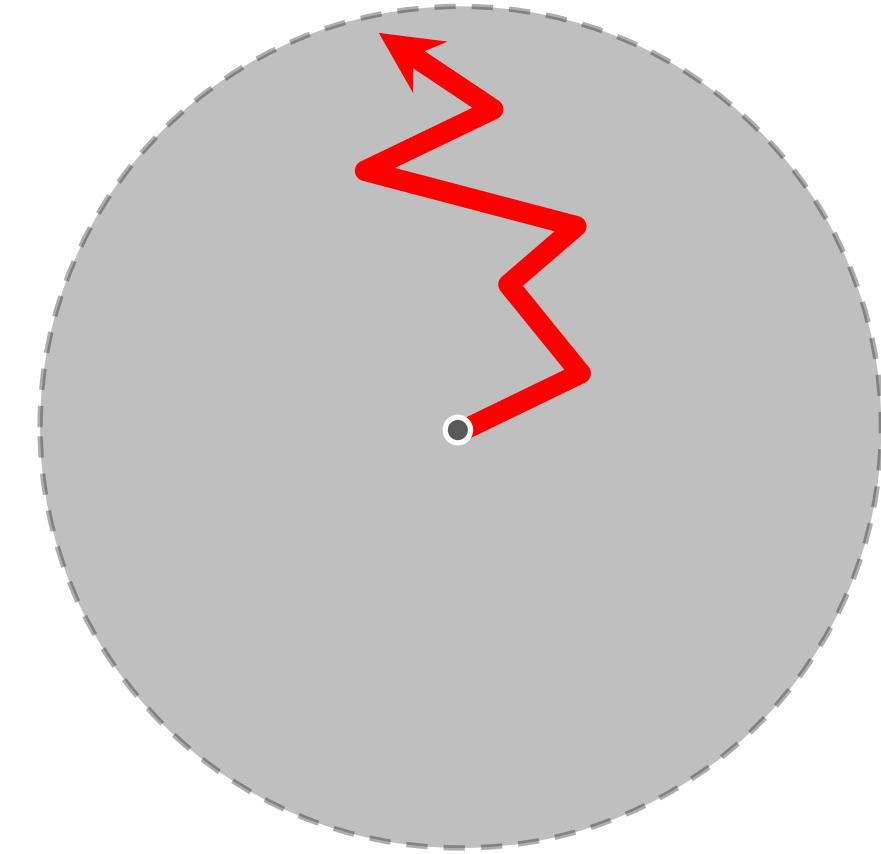
**the questions is
part of the context
and used to guess
the next node**

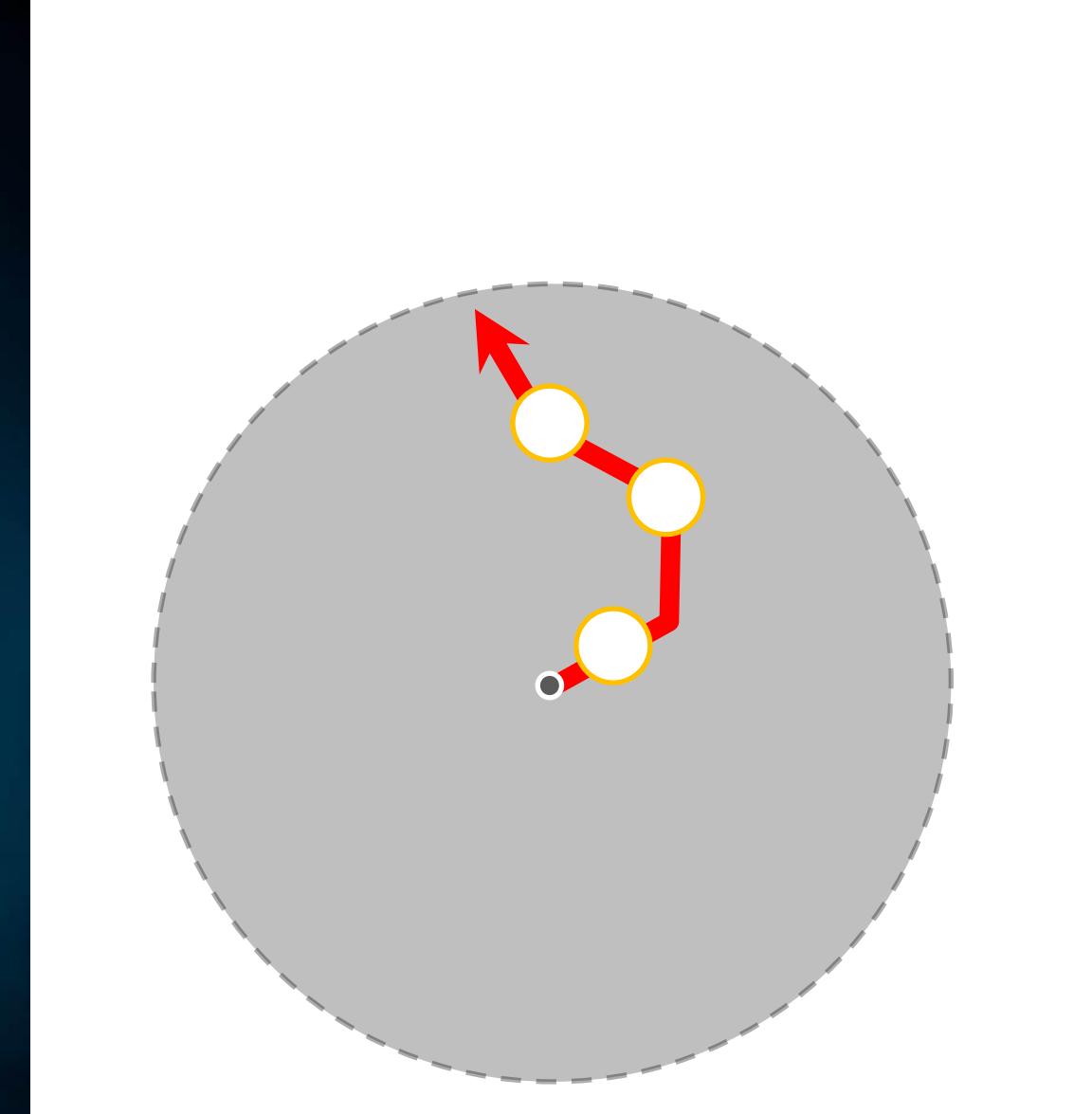


....and extrapolates

meaning
of life?









**Where do
Language
models store
their data?**



Locating and Editing Factual Associations in GPT

Locating and Editing Factual Associations in GPT

Kevin Meng^{*}
MIT CSAIL

David Bau^{*}
Northeastern University

Alex Andonian
MIT CSAIL

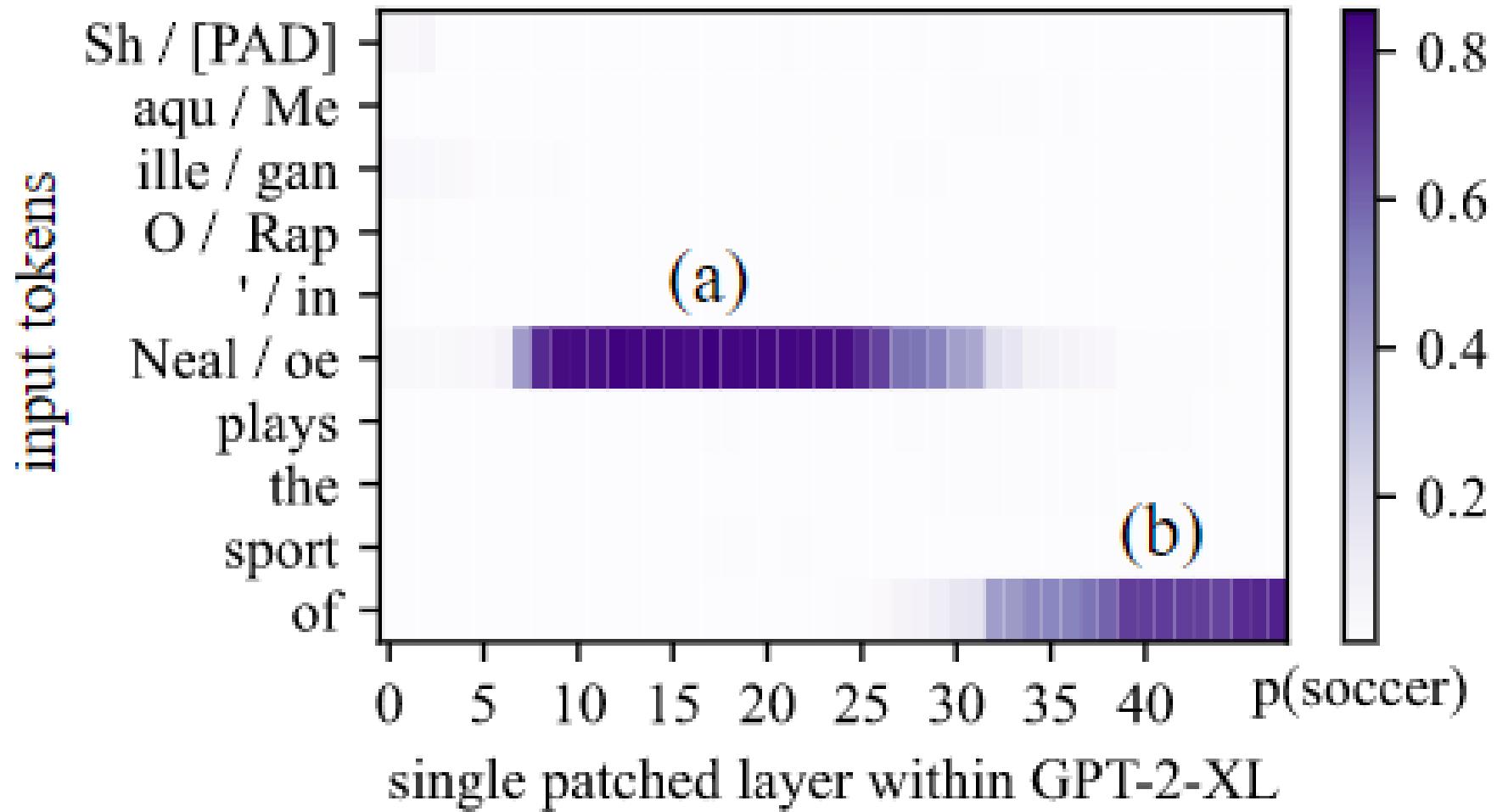
Yonatan Belinkov[†]
Technion – IIT

Abstract

We analyze the storage and recall of factual associations in autoregressive transformer language models, finding evidence that these associations correspond to localized, directly-editable computations. We first develop a causal intervention for identifying neuron *activations* that are decisive in a model’s factual predictions. This reveals a distinct set of steps in middle-layer feed-forward modules that mediate factual predictions while processing subject tokens. To test our hypothesis that these computations correspond to factual association recall, we modify feed-forward *weights* to update specific factual associations using Rank-One Model Editing (ROME). We find that ROME is effective on a standard zero-shot relation extraction (zsRE) model-editing task. We also evaluate ROME on a new dataset of difficult counterfactual assertions, on which it simultaneously maintains both specificity and generalization, whereas other methods sacrifice one or another. Our results confirm an important role for mid-layer feed-forward modules in storing factual associations and suggest that direct manipulation of computational mechanisms may be a feasible approach for model editing. The code, dataset, visualizations, and an interactive demo notebook are available at <https://rome.baulab.info/>.

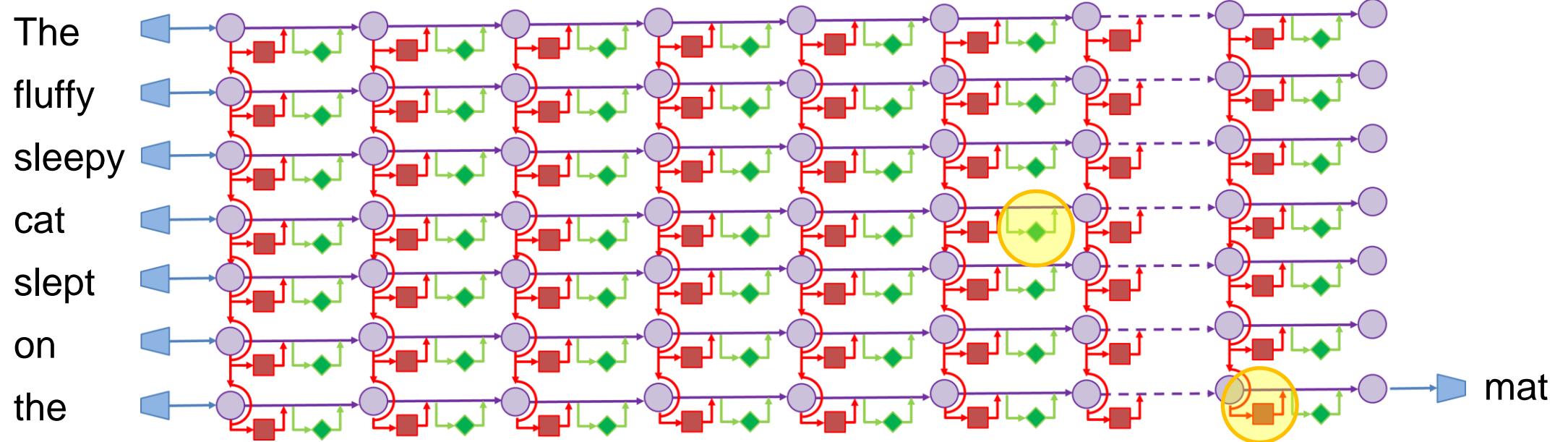
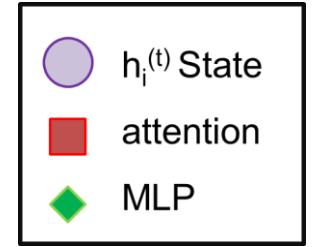
Source: <https://arxiv.org/abs/2202.05262>

Patching hidden state from Rapinoe to Shaq



a) Knowledge retrieval in MLP

b) Attention mechanism at the late site bring information to the end of the network for next word prediction



- Knowledge retrieval in mid layer MLP (Objects - nouns?)
- Attention mechanism at the late site bring information to the end of the network for next word prediction



Mass-Editing Memory in A Transformer

MASS-EDITING MEMORY IN A TRANSFORMER

Kevin Meng^{1,2} Arnab Sen Sharma² Alex Andonian¹ Yonatan Belinkov^{†3} David Bau²
¹MIT CSAIL ²Northeastern University ³Technion – IIT

ABSTRACT

Recent work has shown exciting promise in updating large language models with new memories, so as to replace obsolete information or add specialized knowledge. However, this line of work is predominantly limited to updating single associations. We develop MEMIT, a method for directly updating a language model with many memories, demonstrating experimentally that it can scale up to *thousands of associations* for GPT-J (6B) and GPT-NeoX (20B), exceeding prior work by orders of magnitude. Our code and data are at memit.baulab.info.

Source: <https://arxiv.org/abs/2210.07229>



**Can we improve or
expand memory?**



Frozen Graphs



Plasticity



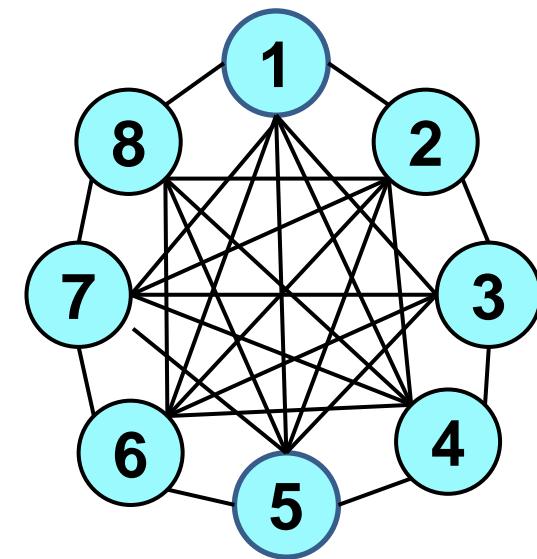
Continual learning

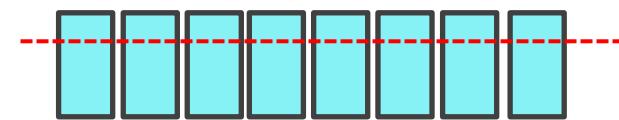
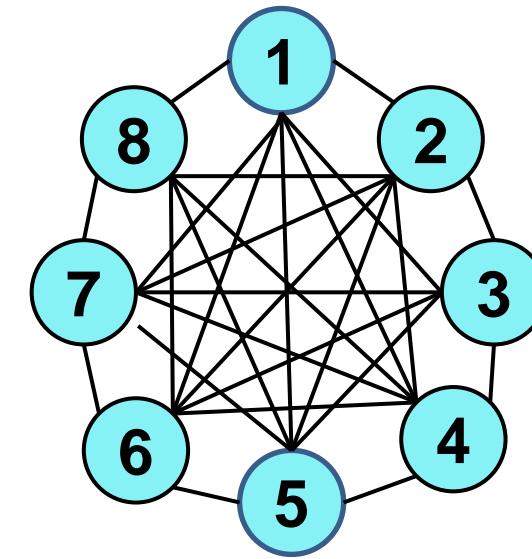


Associative Memory

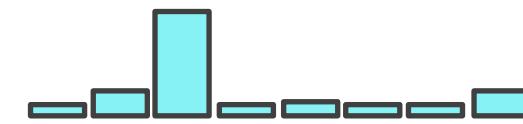


Modern Hopfield Networks





1 2 3 4 5 6 7 8



1 2 3 4 5 6 7 8



1 2 3 4 5 6 7 8



Hopfield Networks is All You Need

HOPFIELD NETWORKS IS ALL YOU NEED

Hubert Ramsauer* Bernhard Schäffl* Johannes Lehner* Philipp Seidl*
Michael Widrich* Thomas Adler* Lukas Gruber* Markus Holzleitner*
Milena Pavlovic^{‡,§} Geir Kjetil Sandve[§] Victor Greiff[‡] David Kreil[†]
Michael Kopp[†] Günter Klambauer* Johannes Brandstetter* Sepp Hochreiter*,[†]
^{*}ELLIS Unit Linz, LIT AI Lab, Institute for Machine Learning,
Johannes Kepler University Linz, Austria
[†]Institute of Advanced Research in Artificial Intelligence (IARAI)
[‡]Department of Immunology, University of Oslo, Norway
[§]Department of Informatics, University of Oslo, Norway

ABSTRACT

We introduce a modern Hopfield network with continuous states and a corresponding update rule. The new Hopfield network can store exponentially (with the dimension of the associative space) many patterns, retrieves the pattern with one update, and has exponentially small retrieval errors. It has three types of energy minima (fixed points of the update): (1) global fixed point averaging over all patterns, (2) metastable states averaging over a subset of patterns, and (3) fixed points which store a single pattern. The new update rule is equivalent to the attention mechanism used in transformers. This equivalence enables a characterization of the heads of transformer models. These heads perform in the first layers preferably global averaging and in higher layers partial averaging via metastable states. The new modern Hopfield network can be integrated into deep learning architectures as layers to allow the storage of and access to raw input data, intermediate results, or learned prototypes. These Hopfield layers enable new ways of deep learning, beyond fully-connected, convolutional, or recurrent networks, and provide pooling, memory, association, and attention mechanisms. We demonstrate the broad applicability of the Hopfield layers across various domains. Hopfield layers improved state-of-the-art on three out of four considered multiple instance learning problems as well as on immune repertoire classification with several hundreds of thousands of instances. On the UCI benchmark collections of small classification tasks, where deep learning methods typically struggle, Hopfield layers yielded a new state-of-the-art when compared to different machine learning methods. Finally, Hopfield layers achieved state-of-the-art on two drug design datasets. The implementation is available at: <https://github.com/ml-jku/hopfield-layers>

Source: <https://arxiv.org/pdf/2008.02217.pdf>



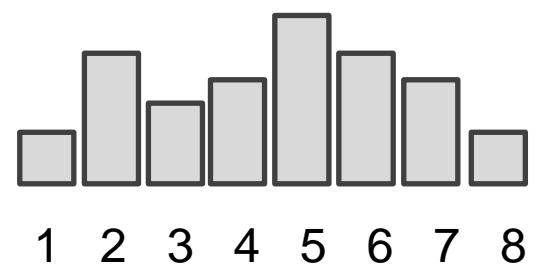
Modern Hopfield Networks

Memory scales
exponential with
number of nodes

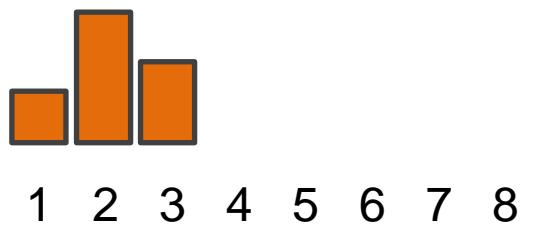


Pattern Recognition

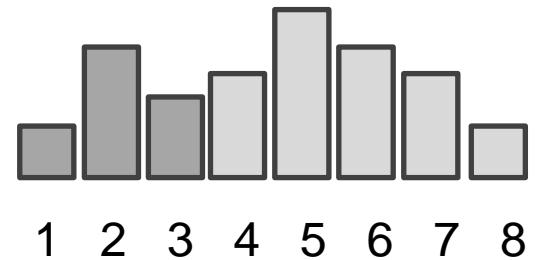
**Stored
pattern**



Query/key



Value





Look closer



Energy Transformer

Energy Transformer

Benjamin Hoover*

IBM Research
Georgia Tech

benjamin.hoover@ibm.com

Yuchen Liang*

Department of CS
RPI

liangy7@rpi.edu

Bao Pham*

Department of CS
RPI

phamb@rpi.edu

Rameswar Panda

MIT-IBM Watson AI Lab
IBM Research

rpanda@ibm.com

Hendrik Strobelt

MIT-IBM Watson AI Lab
IBM Research

hendrik.strobelt@ibm.com

Duen Horng Chau

College of Computing
Georgia Tech

polo@gatech.edu

Mohammed J. Zaki

Department of CS
RPI

zaki@cs.rpi.edu

Dmitry Krotov

MIT-IBM Watson AI Lab
IBM Research

krotov@ibm.com

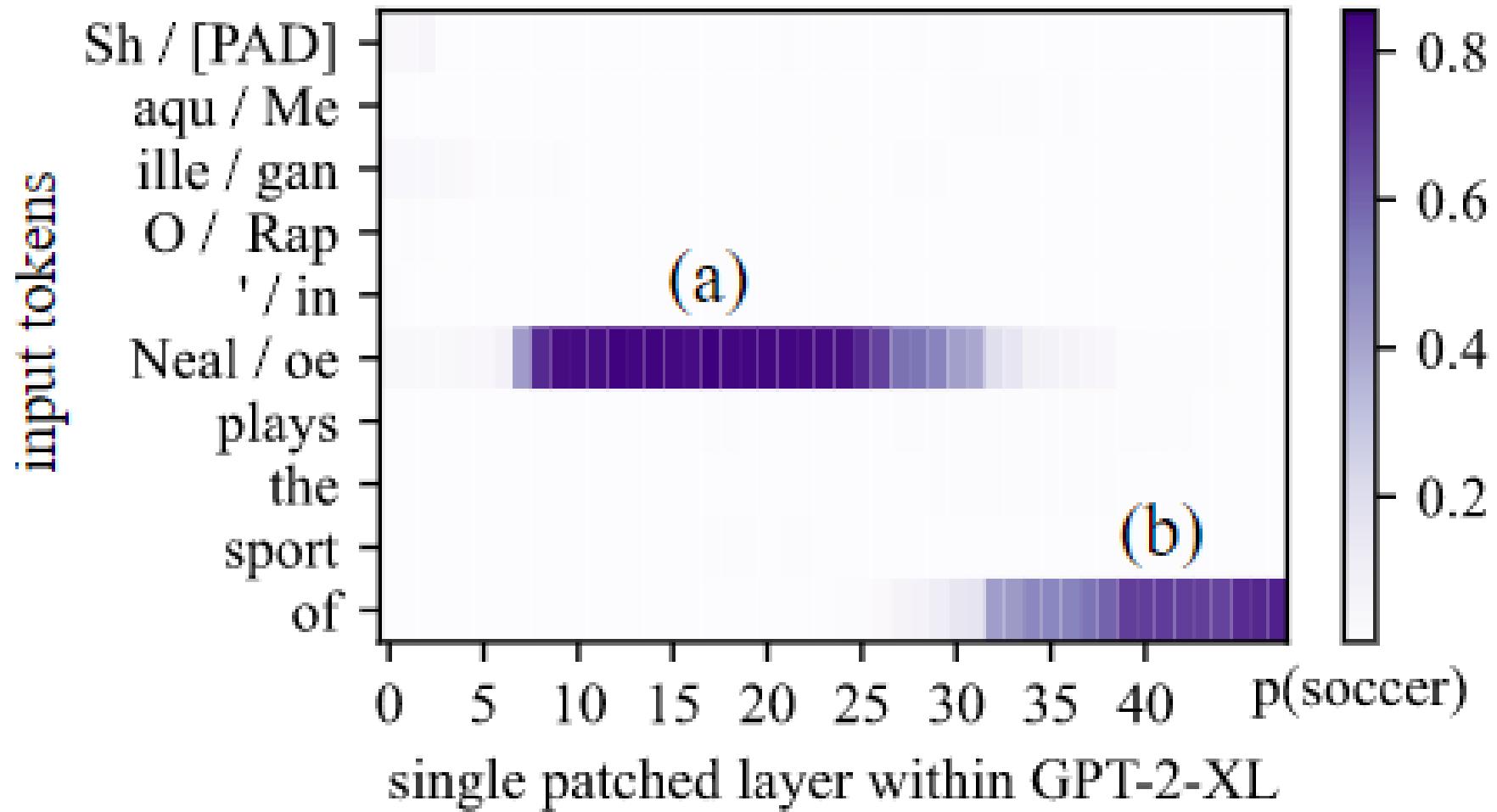
Abstract

Transformers have become the de facto models of choice in machine learning, typically leading to impressive performance on many applications. At the same time, the architectural development in the transformer world is mostly driven by empirical findings, and the theoretical understanding of their architectural building blocks is rather limited. In contrast, Dense Associative Memory models or Modern Hopfield Networks have a well-established theoretical foundation, but have not yet demonstrated truly impressive practical results. We propose a transformer architecture that replaces the sequence of feedforward transformer blocks with a single large Associative Memory model. Our novel architecture, called Energy Transformer (or ET for short), has many of the familiar architectural primitives that are often used in the current generation of transformers. However, it is not identical to the existing architectures. The sequence of transformer layers in ET is purposely designed to minimize a specifically engineered energy function, which is responsible for representing the relationships between the tokens. As a consequence of this computational principle, the attention in ET is different from the conventional attention mechanism. In this work, we introduce the theoretical foundations of ET, explore its empirical capabilities using the image completion task, and obtain strong quantitative results on the graph anomaly detection task.

Source: <https://arxiv.org/abs/2302.07253>

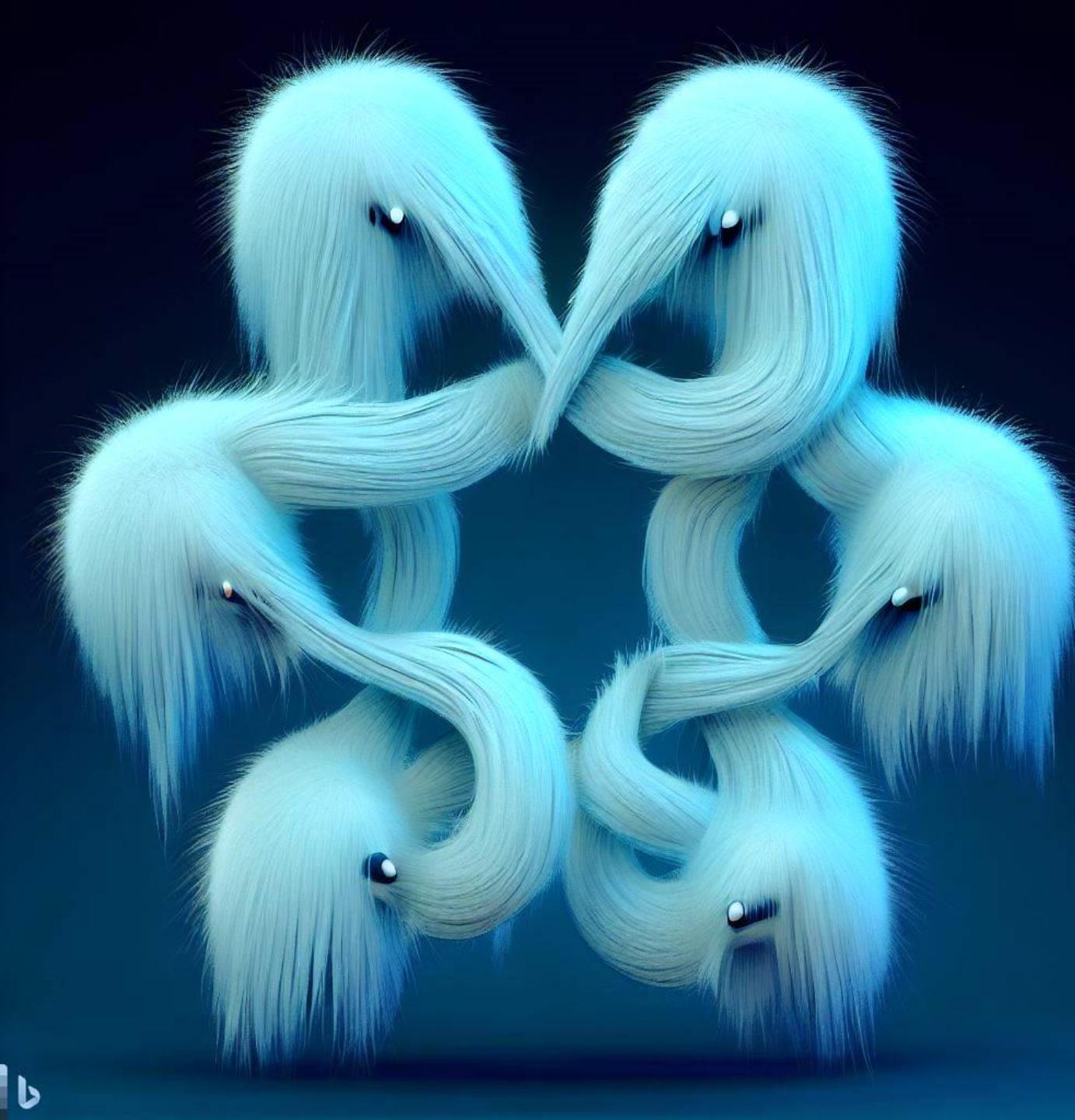
<https://www.youtube.com/watch?v=5LXiQUsnHrI>

Patching hidden state from Rapinoe to Shaq



a) Knowledge retrieval in MLP

b) Attention mechanism at the late site bring information to the end of the network for next word prediction



Top Layer



First Layer

Source: <https://arxiv.org/pdf/2008.02217.pdf>



Backprop?



Fine Tuning Language Models with Just Forward Passes

Fine-Tuning Language Models with Just Forward Passes

Sadhika Malladi* Tianyu Gao* Eshaan Nichani Alex Damian

Jason D. Lee Danqi Chen Sanjeev Arora

Princeton University
{smalladi, tianyug, eshnich, ad27, jasonlee, danqic, arora}@princeton.edu

Abstract

Fine-tuning language models (LMs) has yielded success on diverse downstream tasks, but as LMs grow in size, backpropagation requires a prohibitively large amount of memory. Zeroth-order (ZO) methods can in principle estimate gradients using only two forward passes but are theorized to be catastrophically slow for optimizing large models. In this work, we propose a memory-efficient zeroth-order optimizer (**MeZO**), adapting the classical ZO-SGD method to operate in-place, thereby fine-tuning LMs with *the same memory footprint as inference*. For example, with a single A100 80GB GPU, MeZO can train a 30-billion parameter model, whereas fine-tuning with backpropagation can train only a 2.7B LM with the same budget. We conduct comprehensive experiments across model types (masked and autoregressive LMs), model scales (up to 66B), and downstream tasks (classification, multiple-choice, and generation). Our results demonstrate that (1) MeZO significantly outperforms in-context learning and linear probing; (2) MeZO achieves comparable performance to fine-tuning with backpropagation across multiple tasks, with up to $12 \times$ memory reduction; (3) MeZO is compatible with both full-parameter and parameter-efficient tuning techniques such as LoRA and prefix tuning; (4) MeZO can effectively optimize non-differentiable objectives (e.g., maximizing accuracy or F1). We support our empirical findings with theoretical insights, highlighting how adequate pre-training and task prompts enable MeZO to fine-tune huge models, despite classical ZO analyses suggesting otherwise.²

Source: <https://arxiv.org/abs/2305.17333>



**Transformer layers
1:1 replaced by EBMs**

Associative memory

Large capacity

Continual learning?

Doable in Hardware?



**but is this really
anything like the
brain?**



The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation

Source: <https://doi.org/10.1016/j.cell.2020.10.024>

Article

The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation

James C.R. Whittington,^{1,8,9,*} Timothy H. Muller,^{1,2,8} Shirley Mark,³ Guifen Chen,^{4,5} Caswell Barry,^{6,7} Neil Burgess,^{2,3,4,6} and Timothy E.J. Behrens^{1,3,6}

¹Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford OX3 9DU, UK

²Institute of Neurology, UCL, London WC1N 3BG, UK

³Wellcome Centre for Human Neuroimaging, UCL, London WC1N 3AR, UK

⁴Institute of Cognitive Neuroscience, UCL, London WC1N 3AZ, UK

⁵School of Biological and Chemical Sciences, QMUL, London E1 4NS, UK

⁶Sainsbury Wellcome Centre for Neural Circuits and Behaviour, UCL, London W1T 4JG, UK

⁷Research department of Cell and Developmental Biology, UCL, London WC1E 6BT, UK

⁸These authors contributed equally

⁹Lead Contact

*Correspondence: jcrwhittington@gmail.com

<https://doi.org/10.1016/j.cell.2020.10.024>

SUMMARY

The hippocampal-entorhinal system is important for spatial and relational memory tasks. We formally link these domains, provide a mechanistic understanding of the hippocampal role in generalization, and offer unifying principles underlying many entorhinal and hippocampal cell types. We propose medial entorhinal cells form a basis describing structural knowledge, and hippocampal cells link this basis with sensory representations. Adopting these principles, we introduce the Tolman-Eichenbaum machine (TEM). After learning, TEM entorhinal cells display diverse properties resembling apparently bespoke spatial responses, such as grid, band, border, and object-vector cells. TEM hippocampal cells include place and landmark cells that remap between environments. Crucially, TEM also aligns with empirically recorded representations in complex non-spatial tasks. TEM also generates predictions that hippocampal remapping is not random as previously believed; rather, structural knowledge is preserved across environments. We confirm this structural transfer over remapping in simultaneously recorded place and grid cells.



**Spatial
knowledge**

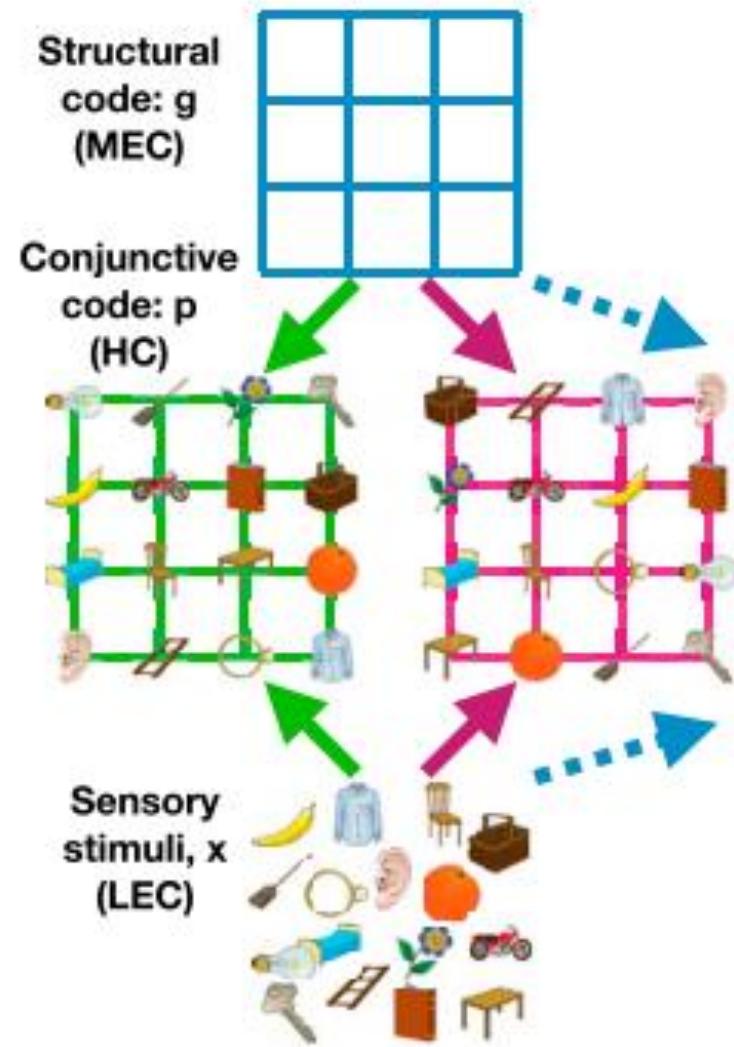
&

**Sensory
representation**



Abstract location Medial EC

Structural knowledge

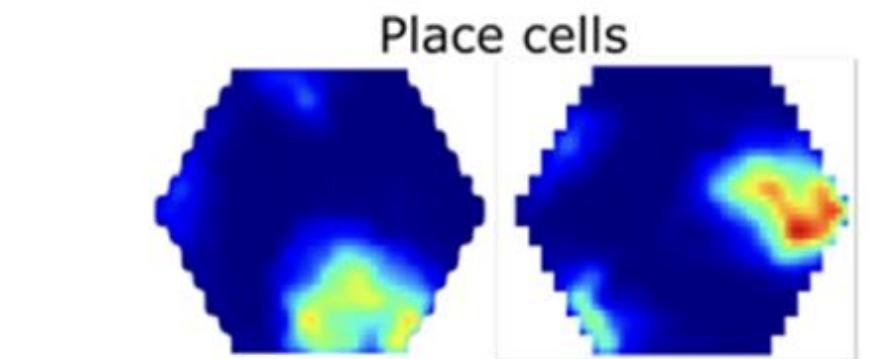
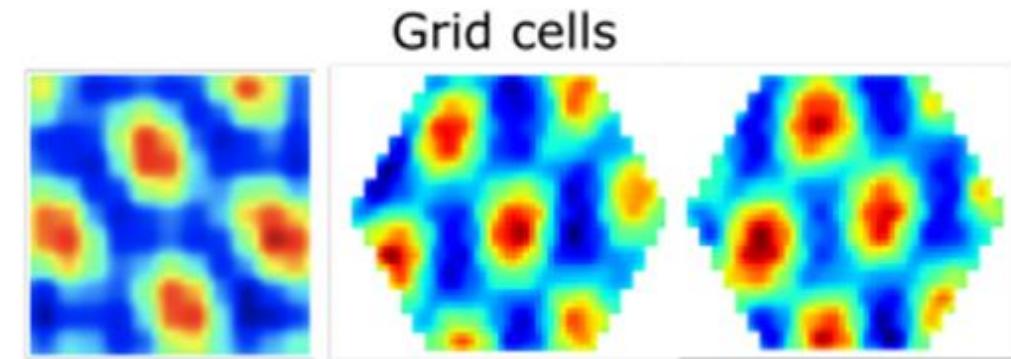


Sensory input Lateral EC

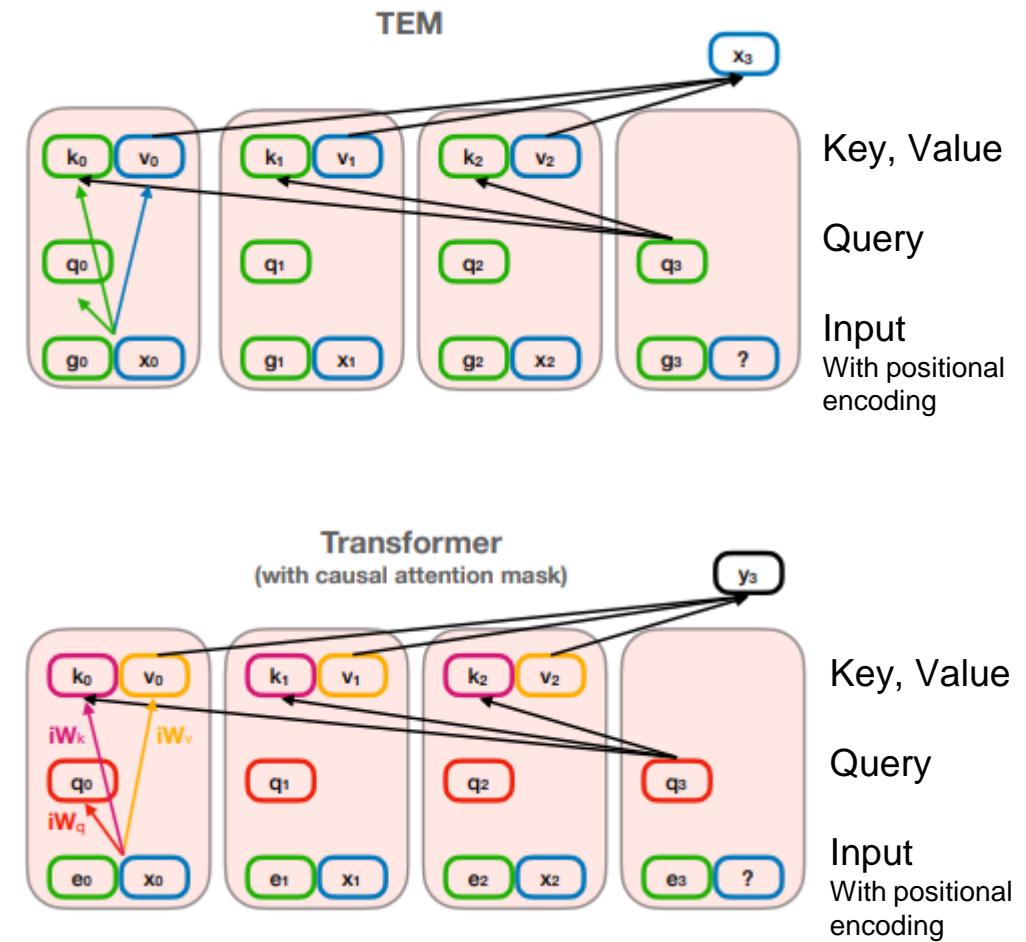
Sensory representation



Rates of fire of neurons in TEM model



TEM model





Relating Transformers to Models and Neural Representations of the Hippocampal Formation

Published as a conference paper at ICLR 2022

RELATING TRANSFORMERS TO MODELS AND NEURAL REPRESENTATIONS OF THE HIPPOCAMPAL FORMATION

James C.R. Whittington*

University of Oxford & Stanford University

Joseph Warren, Timothy E.J. Behrens

University of Oxford & University College London

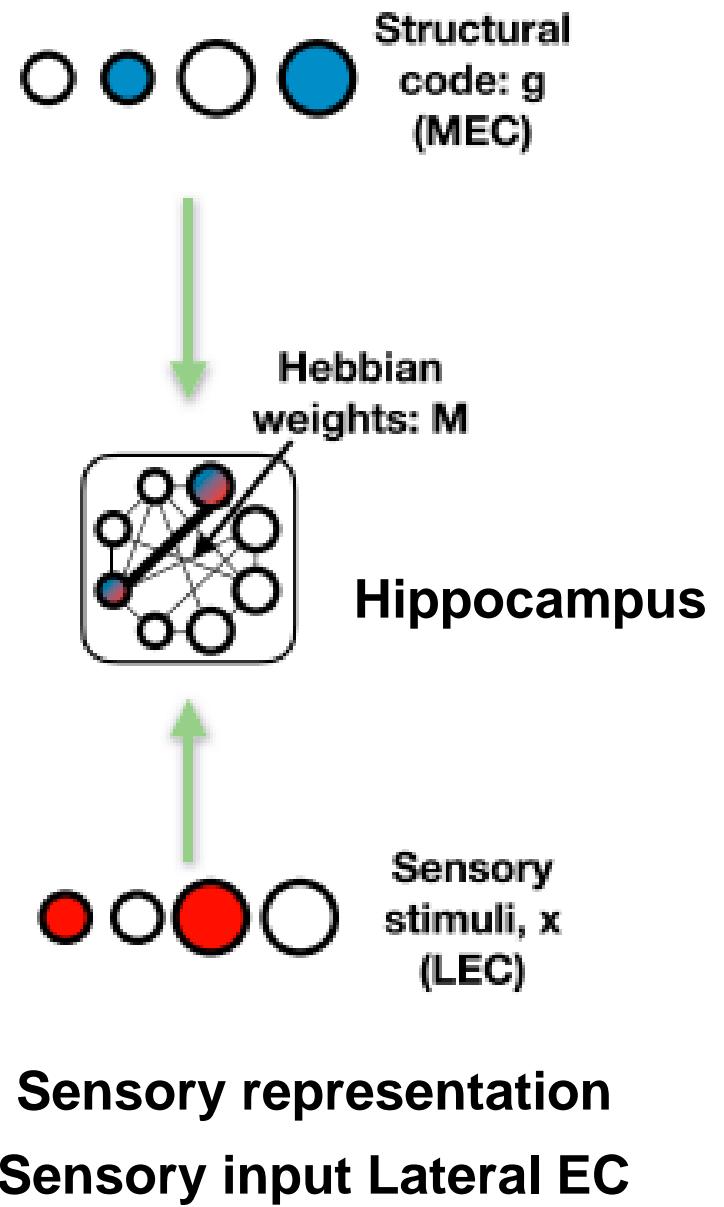
ABSTRACT

Many deep neural network architectures loosely based on brain networks have recently been shown to replicate neural firing patterns observed in the brain. One of the most exciting and promising novel architectures, the Transformer neural network, was developed without the brain in mind. In this work, we show that transformers, when equipped with recurrent position encodings, replicate the precisely tuned spatial representations of the hippocampal formation; most notably place and grid cells. Furthermore, we show that this result is no surprise since it is closely related to current hippocampal models from neuroscience. We additionally show the transformer version offers dramatic performance gains over the neuroscience version. This work continues to bind computations of artificial and brain networks, offers a novel understanding of the hippocampal-cortical interaction, and suggests how wider cortical areas may perform complex tasks beyond current neuroscience models such as language comprehension.

Source: <https://arxiv.org/abs/2112.04035>



Abstract location Medial EC Structural knowledge





Similar to brain model

Spatio-temporal engine
that merges structural &
sensory data.

Not trained that way.

Inherently multimodal

Using them all wrong



Context

micro

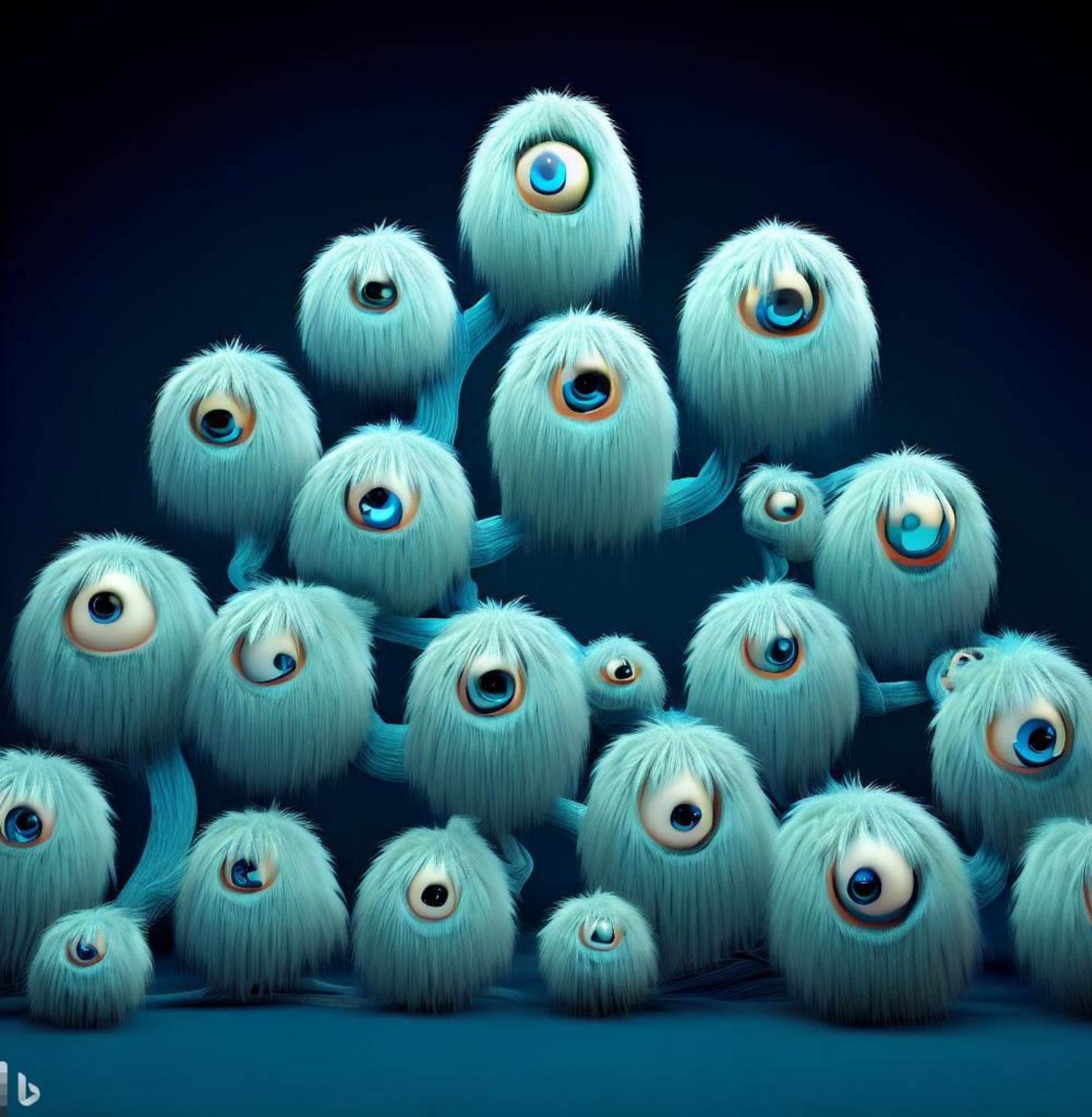


meso



macro





Hierarchical graphs



Compositionality



Topological Deep Learning: Going Beyond Graph Data

Topological Deep Learning: Going Beyond Graph Data

Abstract

Topological deep learning is a rapidly growing field that pertains to the development of deep learning models for data supported on topological domains such as simplicial complexes, cell complexes, and hypergraphs, which generalize many domains encountered in scientific computations. In this paper, we present a unifying deep learning framework built upon a richer data structure that includes widely adopted topological domains.

Specifically, we first introduce *combinatorial complexes*, a novel type of topological domain. Combinatorial complexes can be seen as generalizations of graphs that maintain certain desirable properties. Similar to hypergraphs, combinatorial complexes impose no constraints on the set of relations. In addition, combinatorial complexes permit the construction of hierarchical higher-order relations, analogous to those found in simplicial and cell complexes. Thus, combinatorial complexes generalize and combine useful traits of both hypergraphs and cell complexes, which have emerged as two promising abstractions that facilitate the generalization of graph neural networks to topological spaces.

Second, building upon combinatorial complexes and their rich combinatorial and algebraic structure, we develop a general class of message-passing *combinatorial complex neural networks (CCNNs)*, focusing primarily on attention-based CCNNs. We characterize permutation and orientation equivariances of CCNNs, and discuss pooling and unpooling operations within CCNNs in detail.

Third, we evaluate the performance of CCNNs on tasks related to mesh shape analysis and graph learning. Our experiments demonstrate that CCNNs have competitive performance as compared to state-of-the-art deep learning models specifically tailored to the same tasks. Our findings demonstrate the advantages of incorporating higher-order relations into deep learning models in different applications.

Source: <https://arxiv.org/abs/2206.00606>

Simplicial Hopfield Networks

SIMPLICIAL HOPFIELD NETWORKS

Thomas F. Burns

Neural Coding and Brain Computing Unit
OIST Graduate University, Okinawa, Japan
thomas.burns@oist.jp

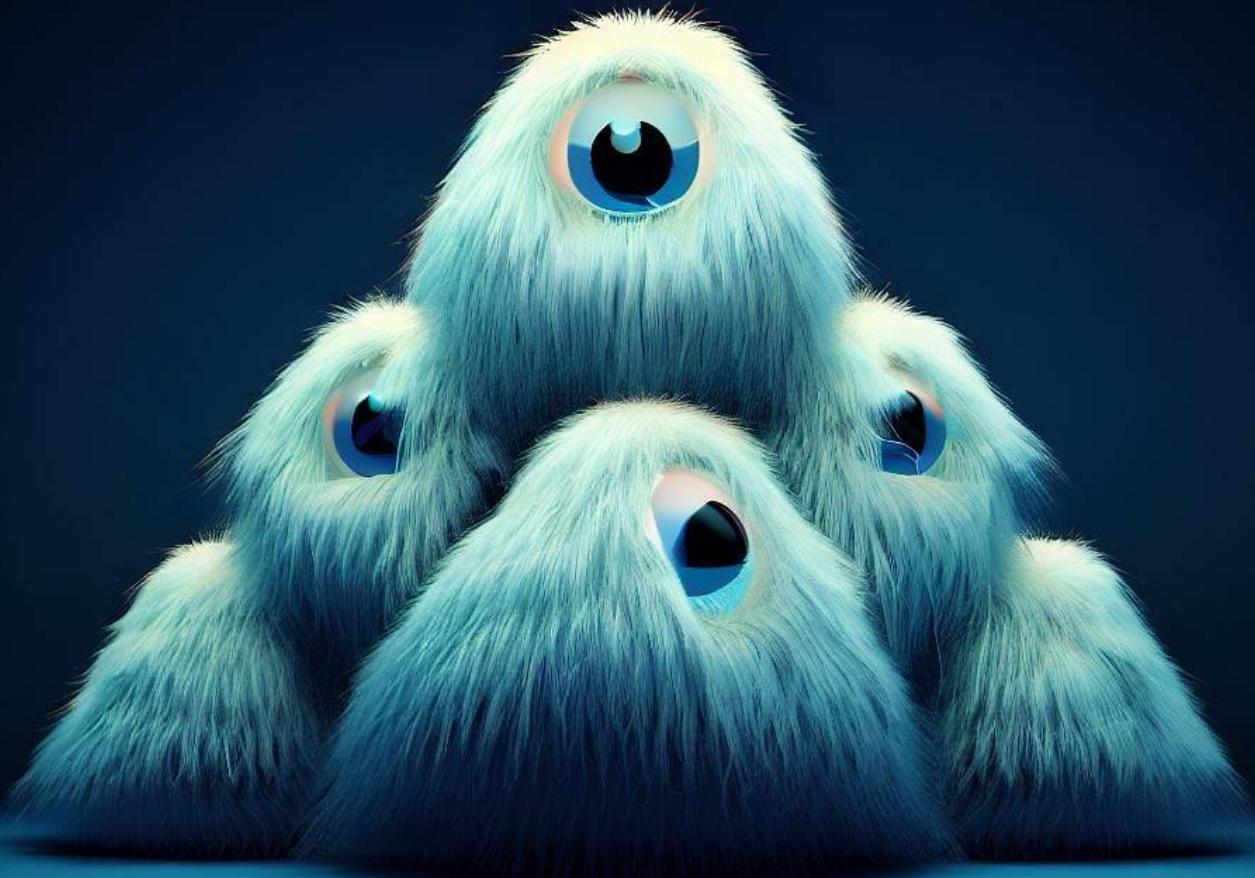
Tomoki Fukai

Neural Coding and Brain Computing Unit
OIST Graduate University, Okinawa, Japan
tomoki.fukai@oist.jp

ABSTRACT

Hopfield networks are artificial neural networks which store memory patterns on the states of their neurons by choosing recurrent connection weights and update rules such that the energy landscape of the network forms attractors around the memories. How many stable, sufficiently-attracting memory patterns can we store in such a network using N neurons? The answer depends on the choice of weights and update rule. Inspired by setwise connectivity in biology, we extend Hopfield networks by adding setwise connections and embedding these connections in a simplicial complex. Simplicial complexes are higher dimensional analogues of graphs which naturally represent collections of pairwise and setwise relationships. We show that our simplicial Hopfield networks increase memory storage capacity. Surprisingly, even when connections are limited to a small random subset of equivalent size to an all-pairwise network, our networks still outperform their pairwise counterparts. Such scenarios include non-trivial simplicial topology. We also test analogous modern continuous Hopfield networks, offering a potentially promising avenue for improving the attention mechanism in Transformer models.

Source: <https://arxiv.org/abs/2305.05179>





**what might
change in
the future**



A Path Towards Autonomous Machine Intelligence

A Path Towards Autonomous Machine Intelligence

Version 0.9.2, 2022-06-27

Yann LeCun

Courant Institute of Mathematical Sciences, New York University yann@cs.nyu.edu
Meta - Fundamental AI Research yann@fb.com

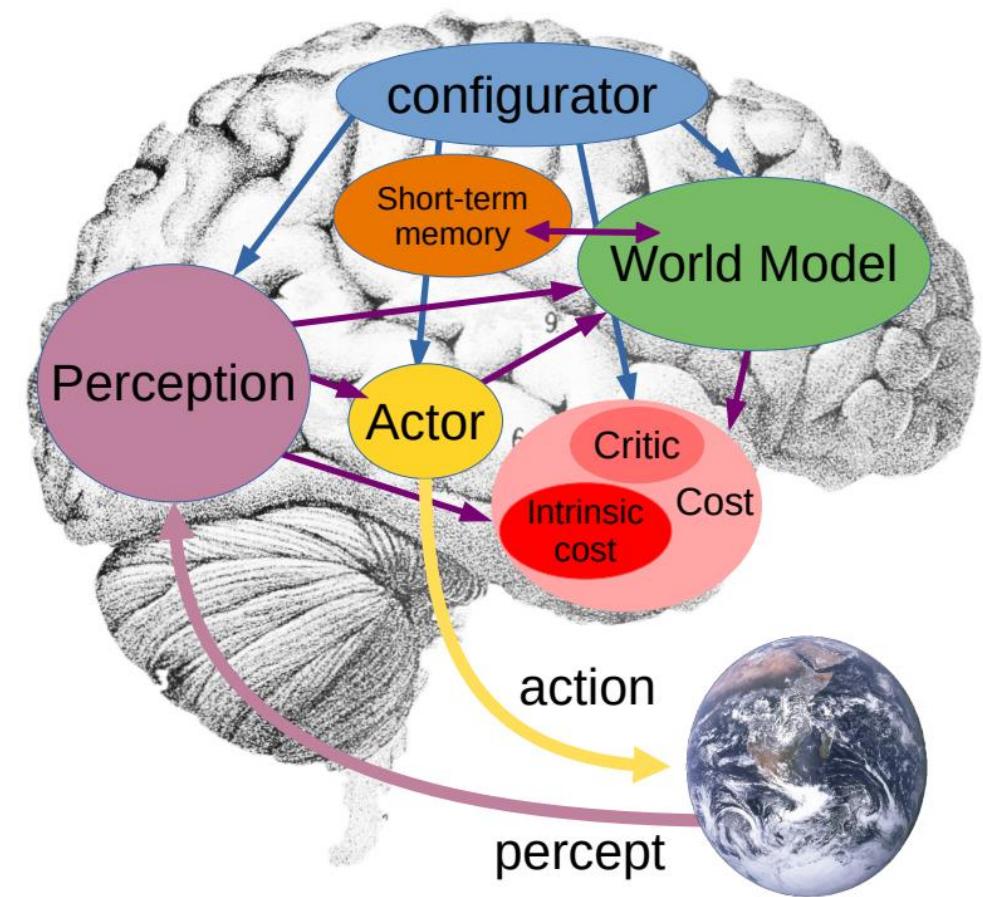
June 27, 2022

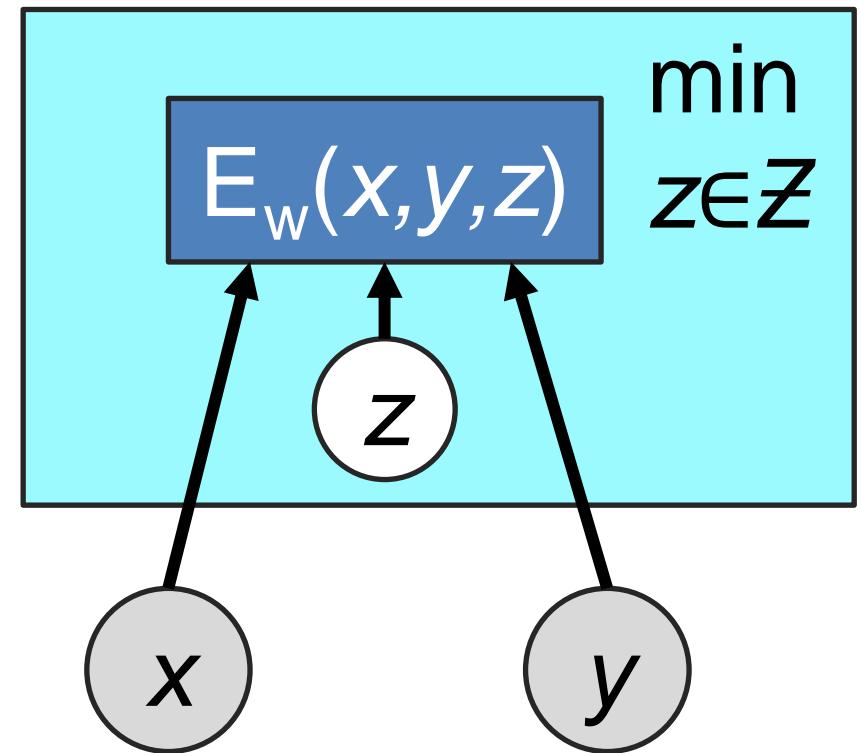
Abstract

How could machines learn as efficiently as humans and animals? How could machines learn to reason and plan? How could machines learn representations of percepts and action plans at multiple levels of abstraction, enabling them to reason, predict, and plan at multiple time horizons? This position paper proposes an architecture and training paradigms with which to construct autonomous intelligent agents. It combines concepts such as configurable predictive world model, behavior driven through intrinsic motivation, and hierarchical joint embedding architectures trained with self-supervised learning.

Keywords: Artificial Intelligence, Machine Common Sense, Cognitive Architecture, Deep Learning, Self-Supervised Learning, Energy-Based Model, World Models, Joint Embedding Architecture, Intrinsic Motivation.

Source: <https://openreview.net/pdf?id=BZ5a1r-kVsf>







Self-supervised learning from images with a joint-embedding predictive architecture

Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture

Mahmoud Assran^{1,2,3*} Quentin Duval¹ Ishan Misra¹ Piotr Bojanowski¹
Pascal Vincent¹ Michael Rabbat^{1,3} Yann LeCun^{1,4} Nicolas Ballas¹

¹Meta AI (FAIR) ²McGill University ³ Mila, Quebec AI Institute ⁴New York University

Abstract

This paper demonstrates an approach for learning highly semantic image representations without relying on hand-crafted data-augmentations. We introduce the Image-based Joint-Embedding Predictive Architecture (I-JEPA), a non-generative approach for self-supervised learning from images. The idea behind I-JEPA is simple: from a single context block, predict the representations of various target blocks in the same image. A core design choice to guide I-JEPA towards producing semantic representations is the masking strategy; specifically, it is crucial to (a) sample target blocks with sufficiently large scale (semantic), and to (b) use a sufficiently informative (spatially distributed) context block. Empirically, when combined with Vision Transformers, we find I-JEPA to be highly scalable. For instance, we train a ViT-Huge/14 on ImageNet using 16 A100 GPUs in under 72 hours to achieve strong downstream performance across a wide range of tasks, from linear classification to object counting and depth prediction.

Source: <https://arxiv.org/abs/2301.08243>



I-JEPA Vision Transformer

Energy based model

Pattern association

Learns faster with less data

High level semantic image representations without hand-crafted data augmentations

converges faster than pixel reconstruction methods

A white, fluffy, hair-like character with a single large eye and a small mouth, holding a scroll. The character is set against a dark blue background.

Summary

Transformers:

Strong connection to key brain models

**Can do spatio-temporal association-
We should train them like that**

Layers can be EBMs - advantages?

Plasticity / continual learning ?

Training Data:

Complete revamp of training methods

Larger contexts needed

Compositional graphs important

Multimodal training way forward

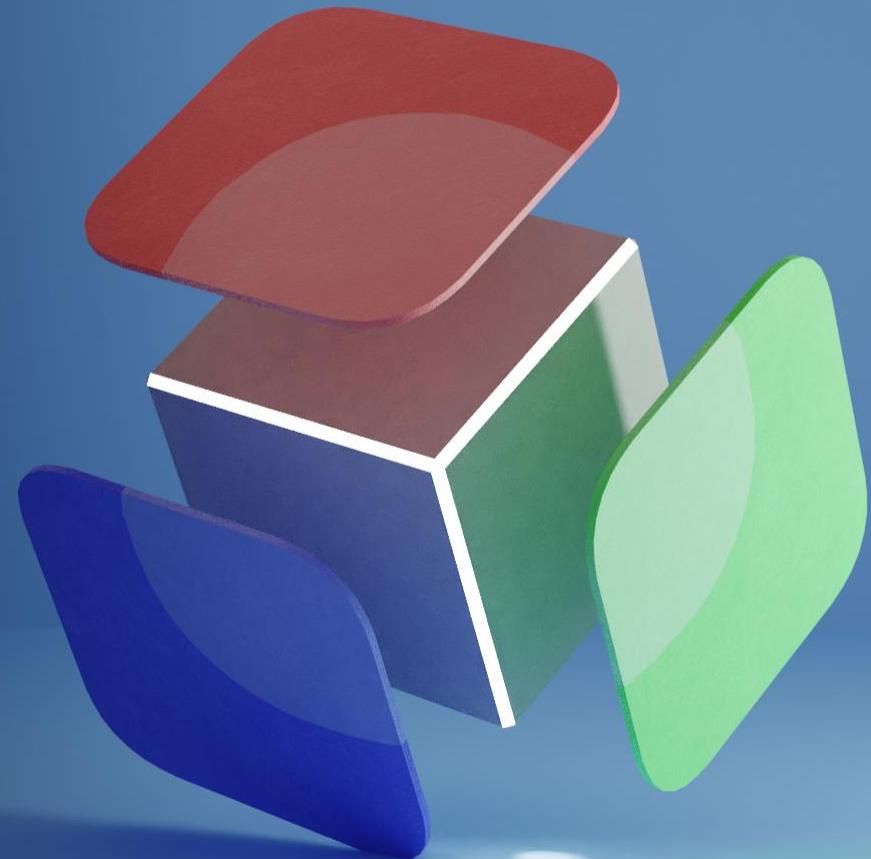
Find out more about PMLG

Collaboration

www.pmlg.com.au/about
pmlg@assistedevolution.net

Events

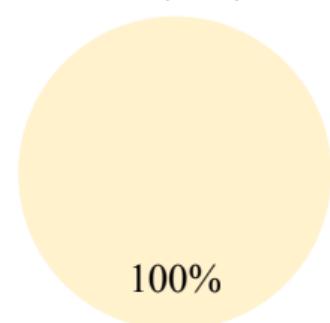
www.meetup.com/perth-machine-learning-group/



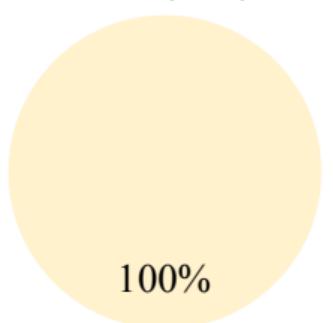
Spares / odds and sods

Ratios of various data sources in the pre-training data for existing LLMs

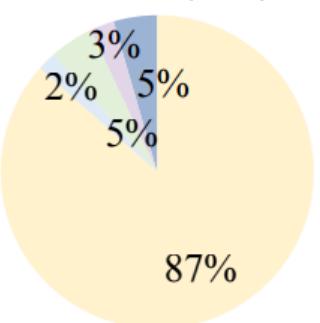
T5 (11B)



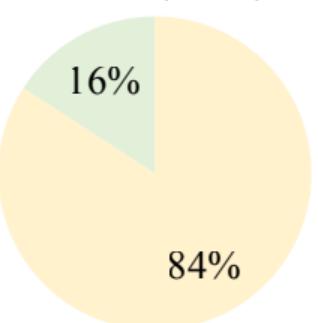
mT5 (13B)



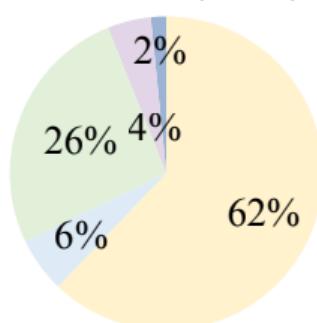
LLaMA (65B)



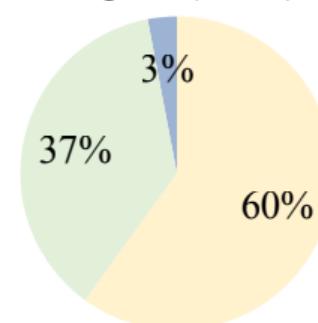
GPT-3 (175B)



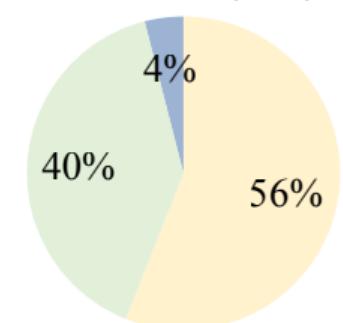
MT-NLG (530B)



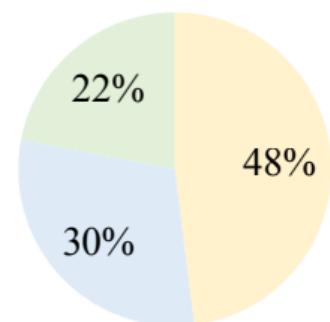
Gopher (280B)



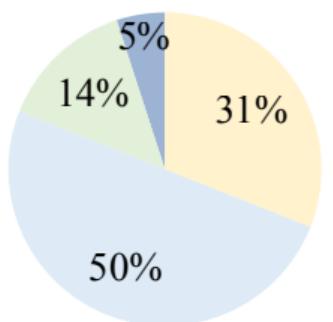
Chinchilla (70B)



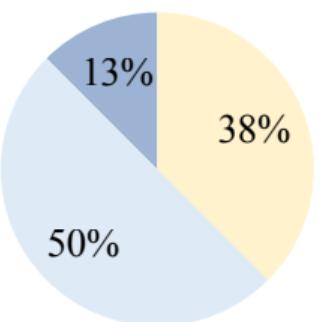
GLaM (1200B)



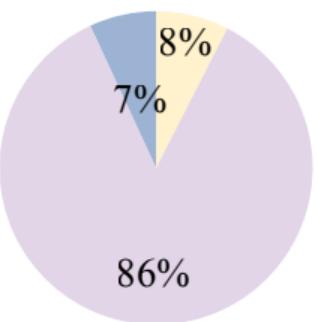
PaLM (540B)



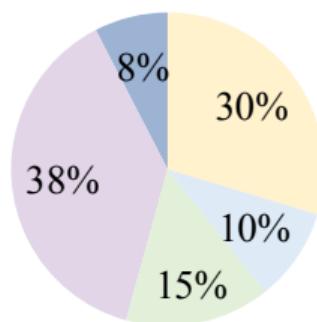
LaMDA (137B)



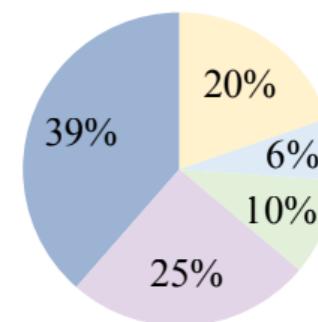
Galactica (120B)



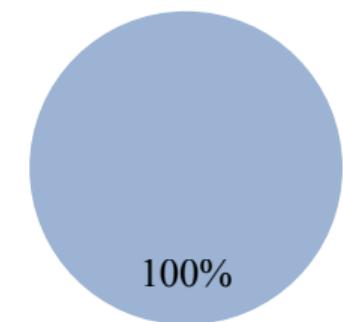
GPT-NeoX (20B)



CodeGen (16B)



AlphaCode (41B)



● Webpages

● Conversation Data

● Books & News

● Scientific Data

● Code



Forward-Forward Training of an Optical Neural Network

Forward-Forward Training of an Optical Neural Network

Ilker Oguz^{1,*}, Junjie Ke², Qifei Wang², Feng Yang², Mustafa Yildirim¹, Niyazi Ulas Dinc¹, Jih-Liang Hsieh¹, Christophe Moser¹ and Demetri Psaltis¹

¹IEM, EPFL, Switzerland

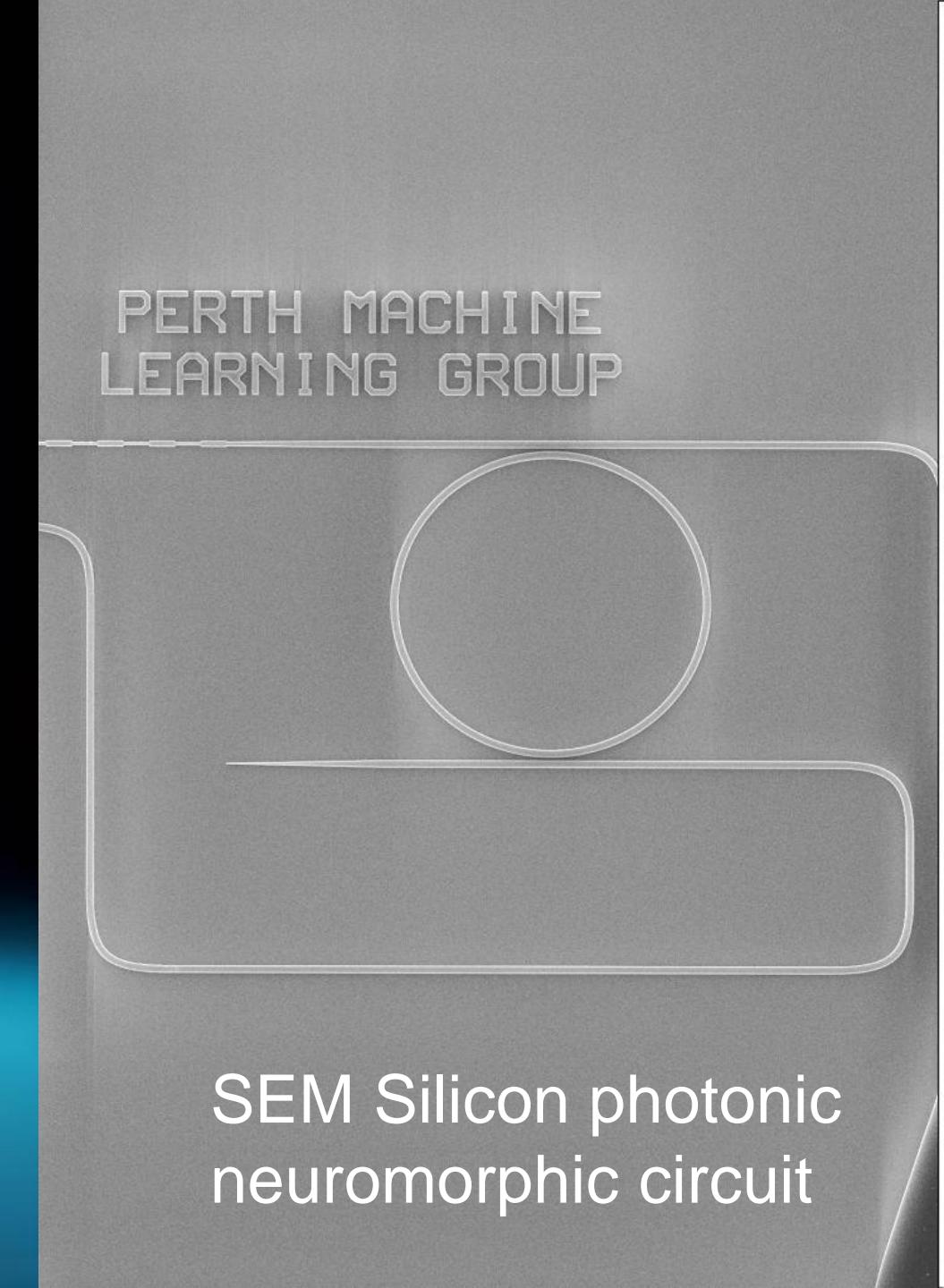
²Google Research, USA

*ilker.oguz@epfl.ch

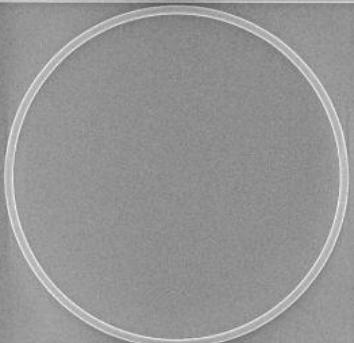
Abstract

Neural networks (NN) have demonstrated remarkable capabilities in various tasks, but their computation-intensive nature demands faster and more energy-efficient hardware implementations. Optics-based platforms, using technologies such as silicon photonics and spatial light modulators, offer promising avenues for achieving this goal. However, training multiple trainable layers in tandem with these physical systems poses challenges, as they are difficult to fully characterize and describe with differentiable functions, hindering the use of error backpropagation algorithm. The recently introduced Forward-Forward Algorithm (FFA) eliminates the need for perfect characterization of the learning system and shows promise for efficient training with large numbers of programmable parameters. The FFA does not require backpropagating an error signal to update the weights, rather the weights are updated by only sending information in one direction. The local loss function for each set of trainable weights enables low-power analog hardware implementations without resorting to metaheuristic algorithms or reinforcement learning. In this paper, we present an experiment utilizing multimode nonlinear wave propagation in an optical fiber demonstrating the feasibility of the FFA approach using an optical system. The results show that incorporating optical transforms in multilayer NN architectures trained with the FFA, can lead to performance improvements, even with a relatively small number of trainable weights. The proposed method offers a new path to the challenge of training optical NNs and provides insights into leveraging physical transformations for enhancing NN performance.

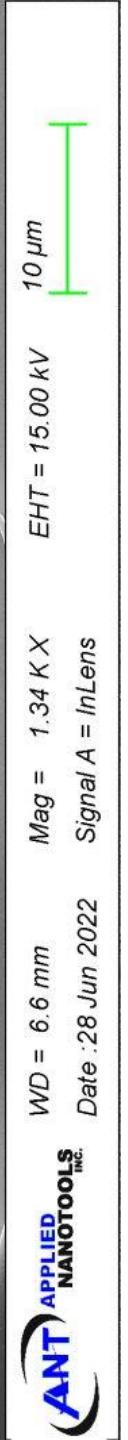
Source: <https://arxiv.org/abs/2305.19170>



PERTH MACHINE
LEARNING GROUP



SEM Silicon photonic neuromorphic circuit



PMLG



Community (3k members in 5 years)

Research and Development

Research Collaborations



trained to follow
instructions and
tasks



**trained with human
feedback**



DOGMA



Associative Memory



Associative Memory



Associative Memory



Energy Based



DOGMA



scaling



**emergent
capabilities
with scaling**



slow to train



expensive to train



**Consume a lot of
power**



expensive to run



**Is this the future
or is there a
better way?**



smaller



1000 x faster



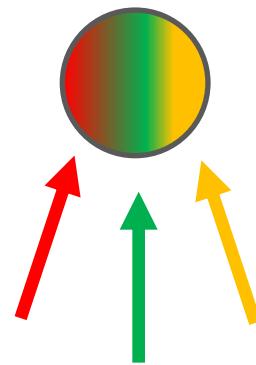
**1000 x energy
efficient**



**and maybe work
more like the
brain**



**early fusion
multimodal**



**joint
representation**



**To stand on
the shoulders
of giants**



**You must first
climb their
backs,
traversing thick
forests of hairy
IP and slippery
scales of non
reproduceable
results**