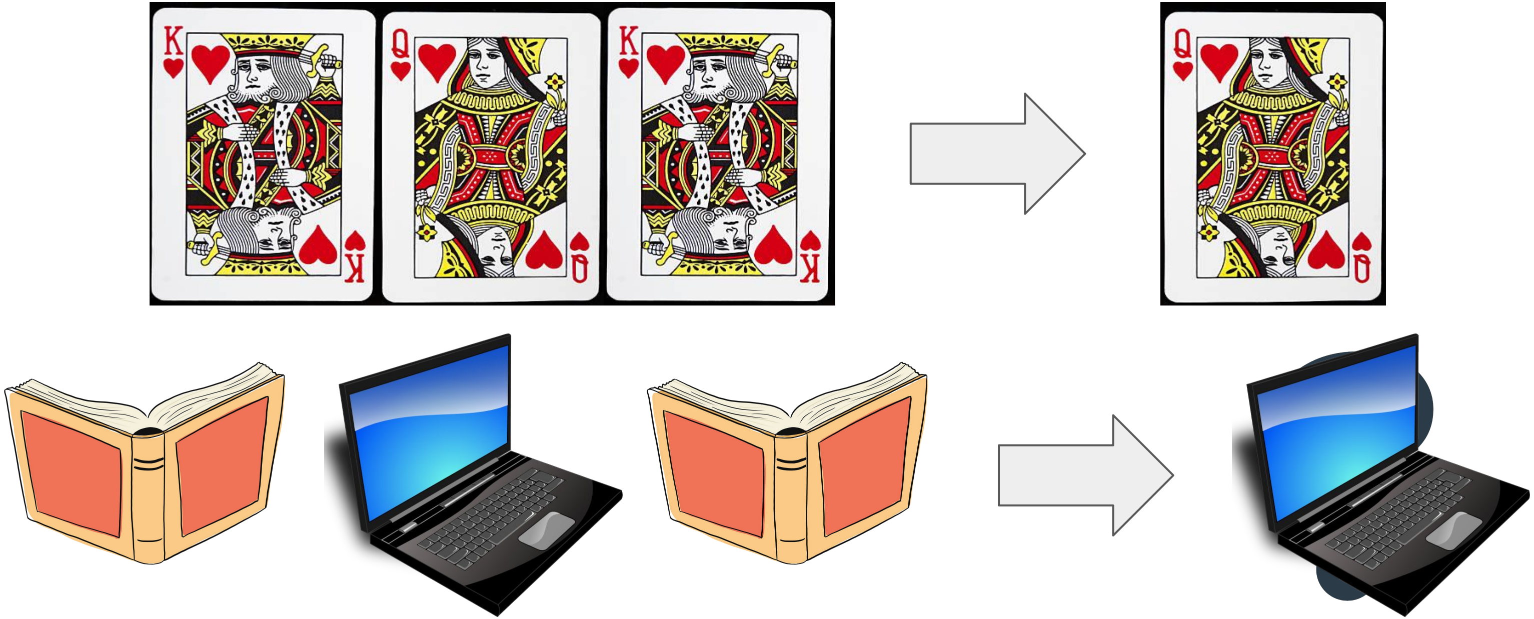
The Relational Bottleneck as an Efficient Bias for Abstraction

Shaunak Bhandarkar and Yukang Yang

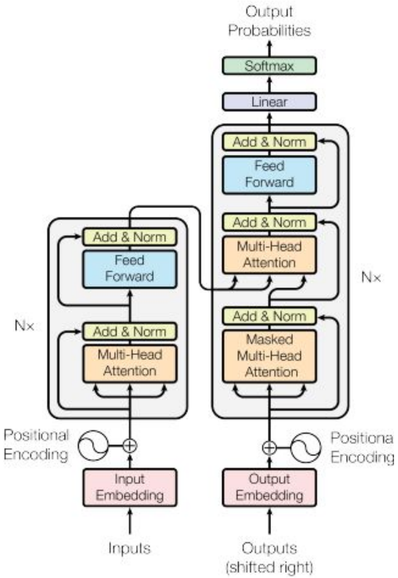
April 2025

How Do People Learn **Abstractions**?



Key idea: We can quickly and reliably generalize to novel situations

In contrast, artificial neural networks seem to require HUGE amounts of training data to learn and they struggle to generalize out-of-distribution…



ChatGPT

So what’s missing from neural network architectures?


Connectionism
Symbolic
Computing

… Perhaps aspects of symbolic computing!

Symbolic Computing

★ **Pro:** By design, symbols are abstracted from the content to which they refer
★ **Pro:** Accounts for systematicity, productivity, and compositionality (Fodor & Pushlyn, 1988)
★ **Con:** Faces a combinatorial explosion of possibilities to account for (curse of dimensionality), and does not shed light on neural implementation

Connectionism

★ **Pro:** Very efficient at representing high-dimensional continuous information
★ **Pro:** Provides a neural implementation
★ **Con:** Often need huge amounts of training data

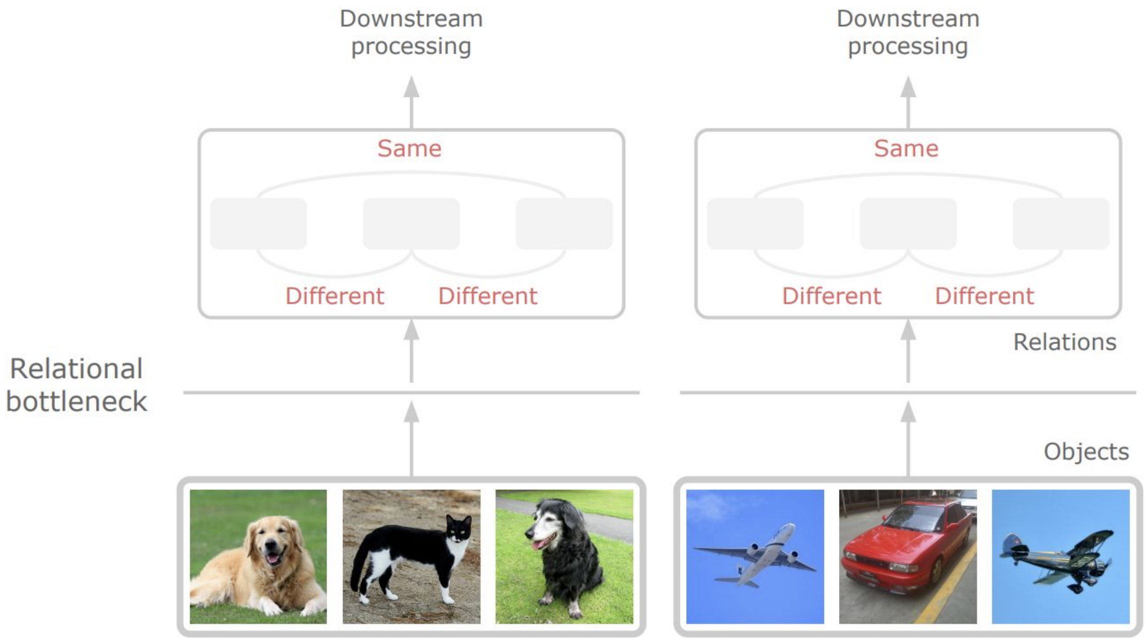
A possible solution:

★ Merge the flexibility of symbolic computing with the efficiency of neural networks…
★ … by building in the right kinds of **inductive biases** that allow for efficient abstraction

Outline

1. Introduce the relational bottleneck
2. Look at concrete neural architectures that use the relational bottleneck, and explore connections to LLMs
3. Connect the relational bottleneck to biological neural computation

The Relational Bottleneck as an Inductive Bias



We want a **compression** of the data that **only** preserves abstract relational structure!

(Fig. 1 from Webb et al., 2024)

The Relational Bottleneck as an Information Bottleneck

- ★ The relational bottleneck is a manifestation of a deeper principle in information theory known as the **information bottleneck** (Tishby et al., 2000)
  - An abstract framework in which we seek to produce a maximal compression of input information that is still sufficient to produce some downstream output
  - Mathematically, for an input X and output Y, we produce an optimally compressed variable R

$$R = \{r(x_i, x_j) \quad \min_{p(R|X)} I(X;R) - \beta I(Y;R) \quad , r(x_{N-1}, x_N)\}$$

A Key Construction: Inner Products

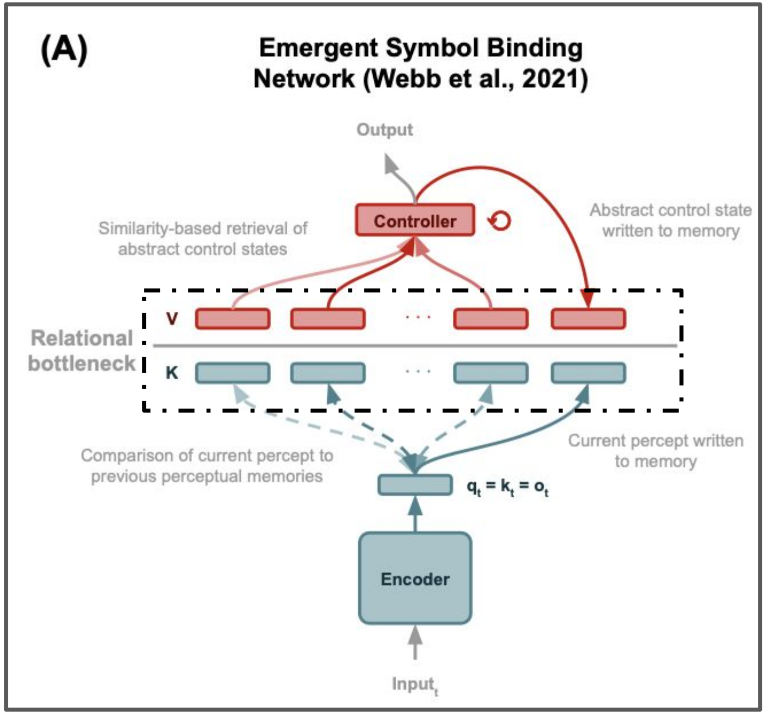
★ We can filter out the second-order **relational** information by converting specific object representations into the pairs of inner products between them

$\vec{a}$ $\vec{b}$ $\vec{c}$ $\vec{d}$

$\vec{a} \cdot \vec{a}$	$\vec{a} \cdot \vec{b}$	$\vec{a} \cdot \vec{c}$	$\vec{a} \cdot \vec{d}$
$\vec{b} \cdot \vec{a}$	$\vec{b} \cdot \vec{b}$	$\vec{b} \cdot \vec{c}$	$\vec{b} \cdot \vec{d}$
$\vec{c} \cdot \vec{a}$	$\vec{c} \cdot \vec{b}$	$\vec{c} \cdot \vec{c}$	$\vec{c} \cdot \vec{d}$
$\vec{d} \cdot \vec{a}$	$\vec{d} \cdot \vec{b}$	$\vec{d} \cdot \vec{c}$	$\vec{d} \cdot \vec{d}$

Downstream Task

Neural Structure Implementations

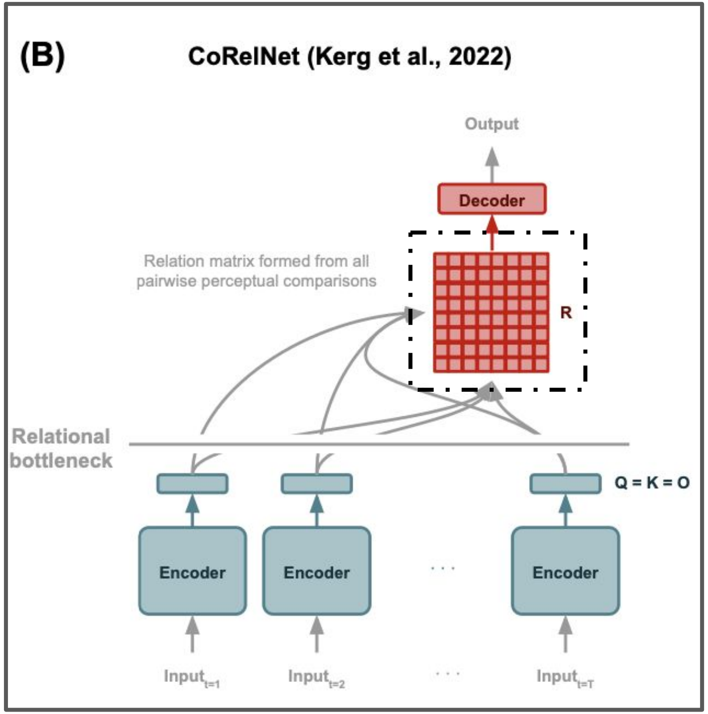


Isolating **perceptual** and **abstract** processing components

- External memory (similar to Episodic Memory)
- Two separate pathways:
  - Perceptual: keys & queries
  - Abstract: values & controller

Taylor, Webb, et al. "Emergent symbols through binding in external memory." *International Conference on Learning Representations* (ICLR) 2021.

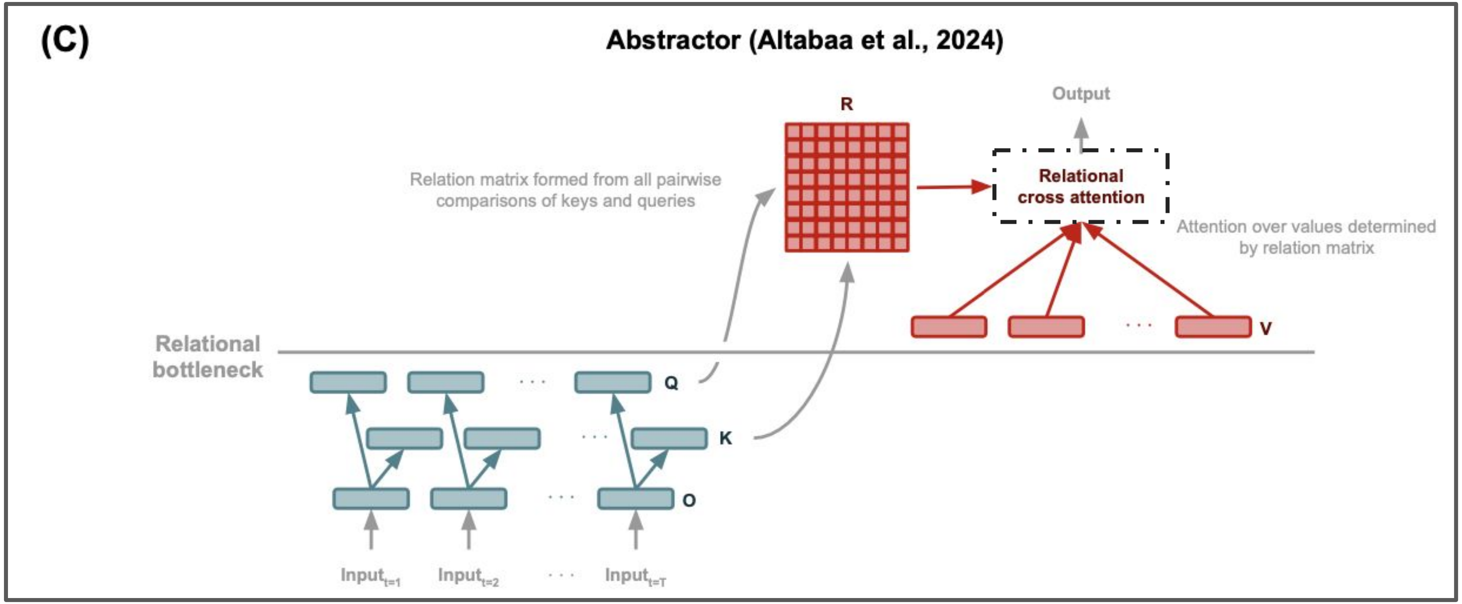
Neural Structure Implementations



- forming a **relation matrix**
- Downstream processing **only** relies on the relation matrix

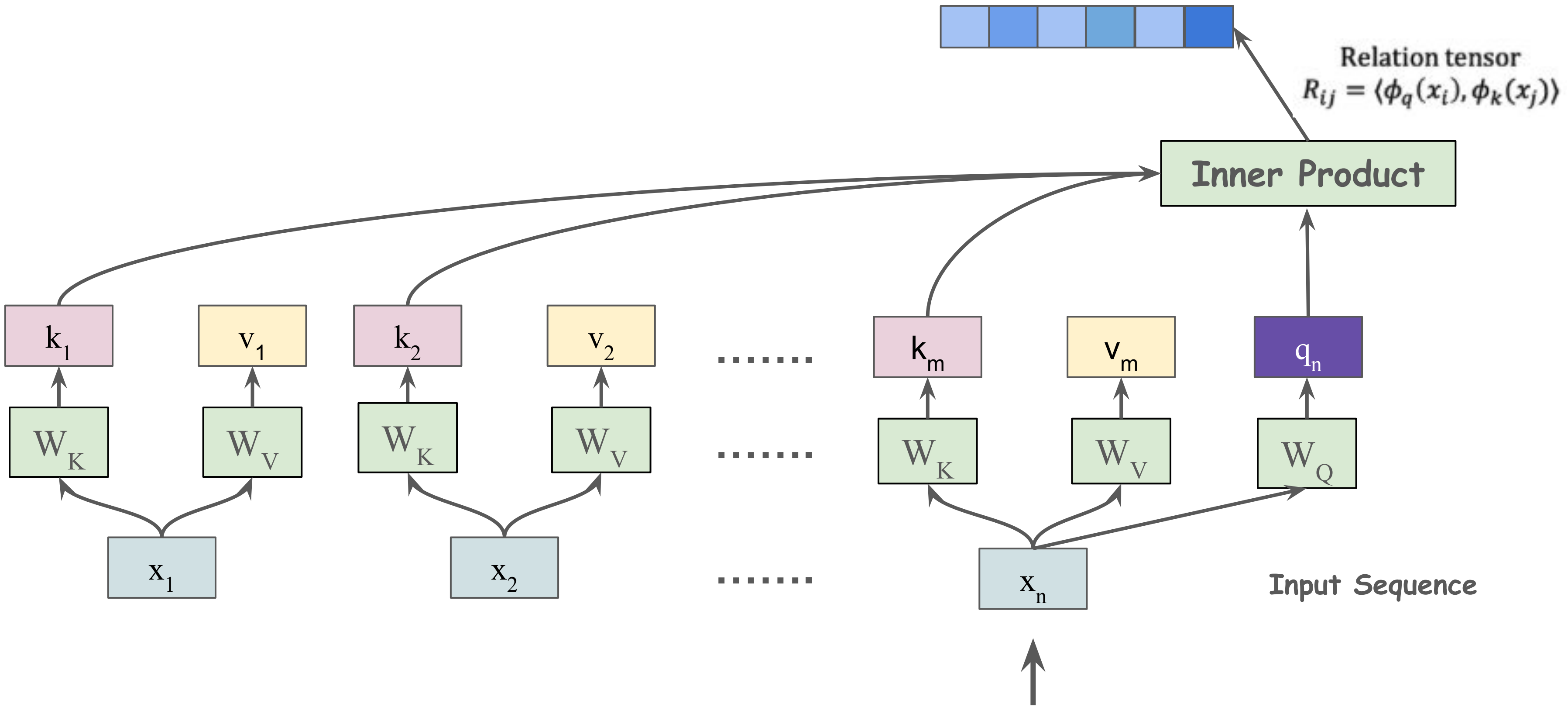
Kerg, Giancarlo, et al. "On neural architecture inductive biases for relational tasks." *arXiv preprint arXiv:2206.05056* (2022).

Neural Structure Implementations

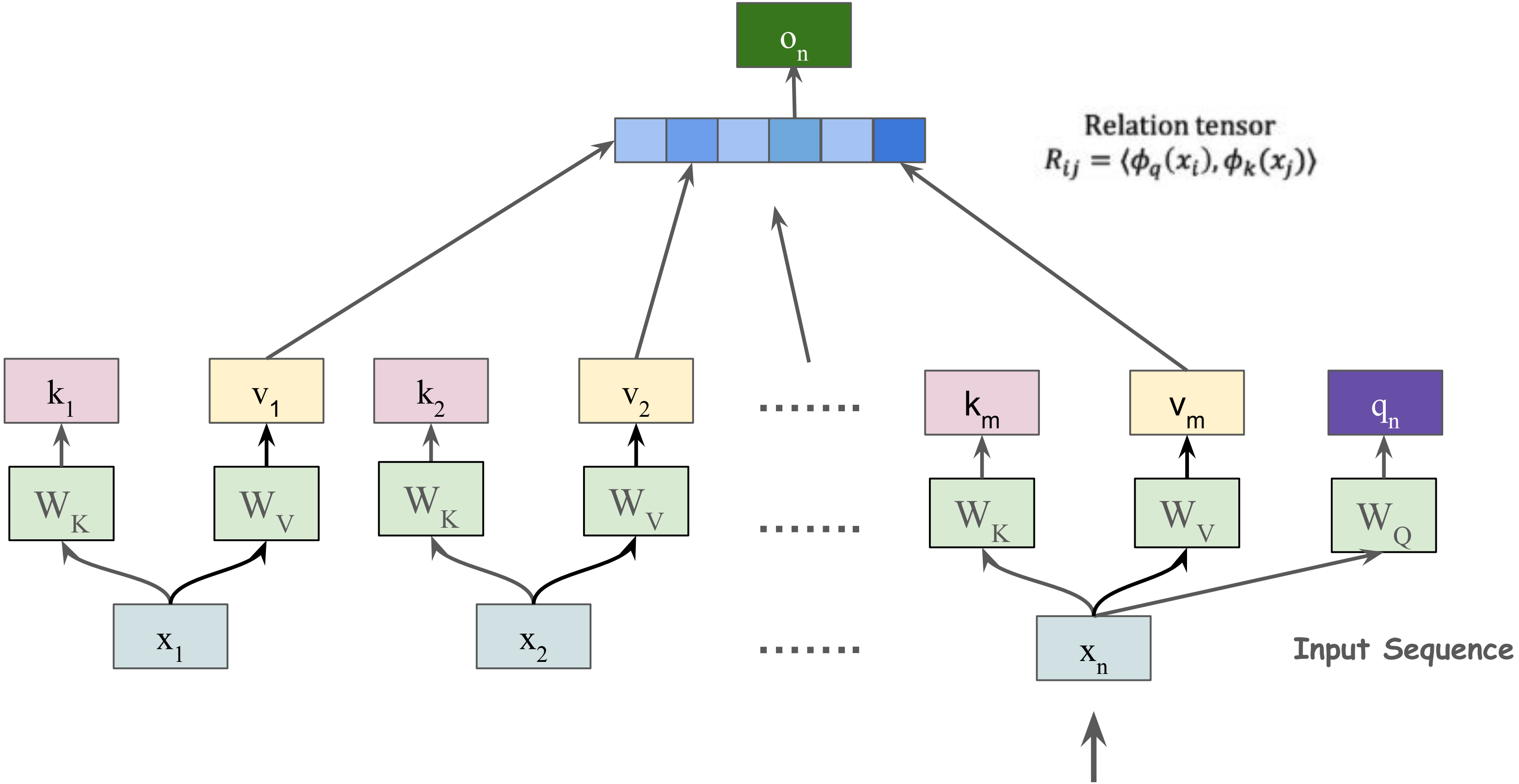


Altabaa, A. et al. “Abstractors and relational cross-attention: An inductive bias for explicit relational reasoning in Transformers.” ICLR 2024.

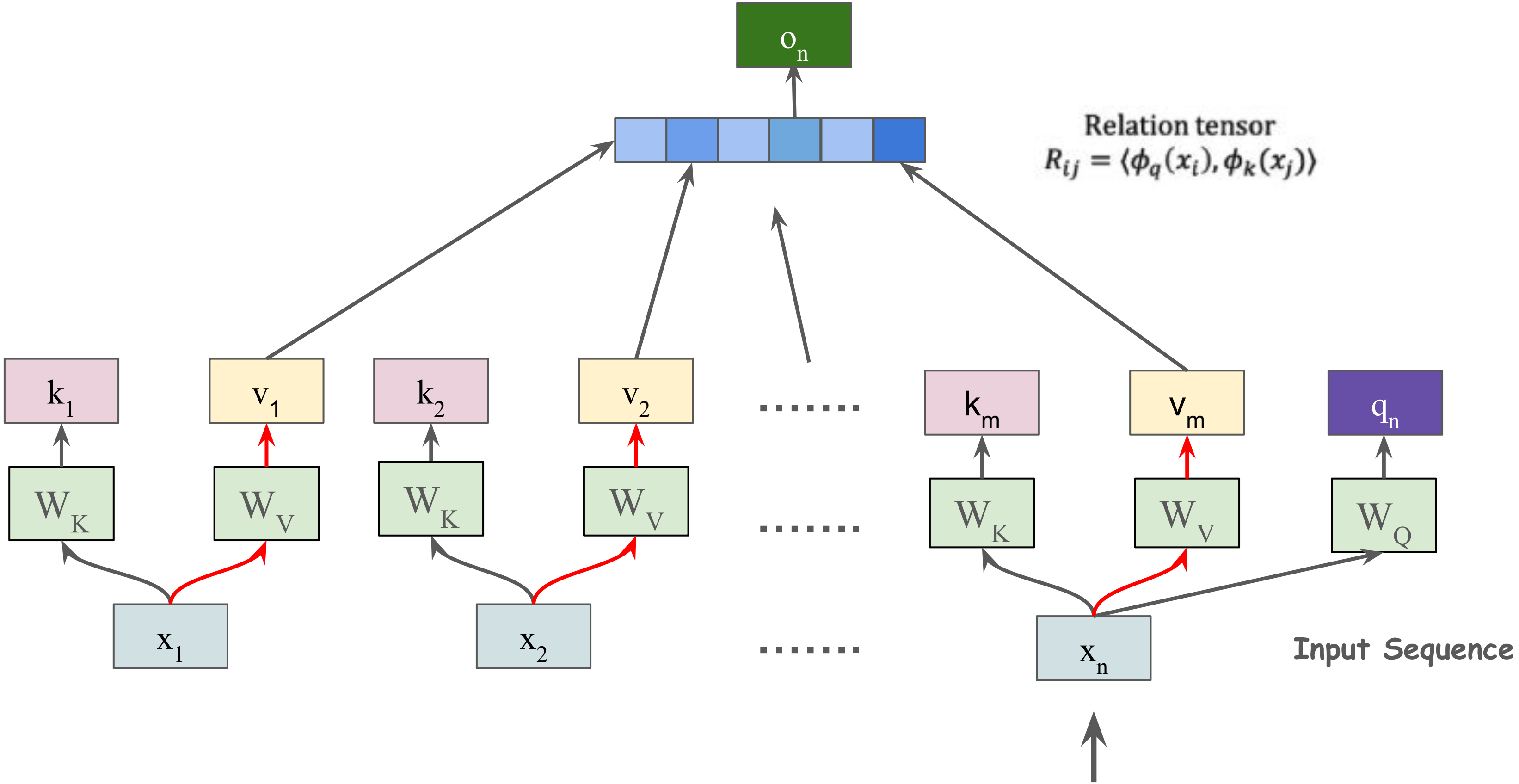
Standard Attention Mechanisms in Transformers



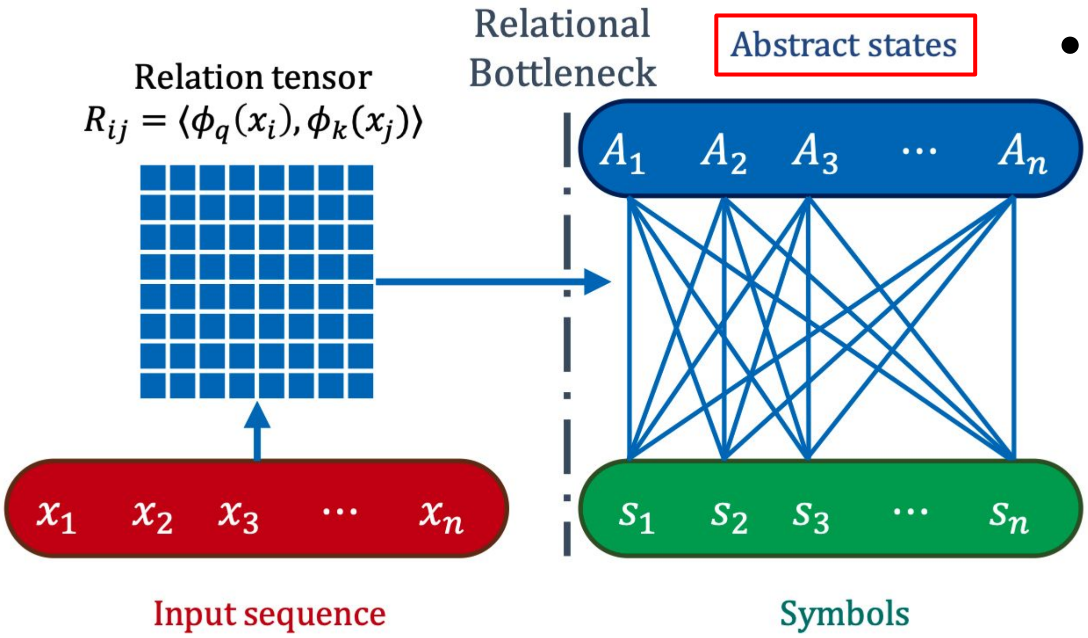
Standard Attention Mechanisms in Transformers



Standard Attention Mechanisms in Transformers



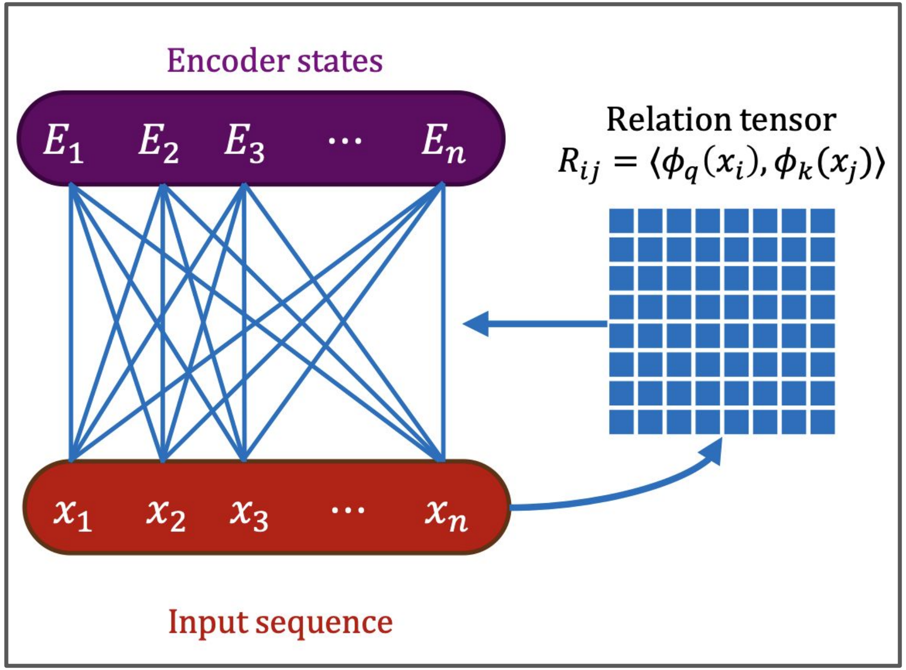
⇒ Relational Cross-Attention



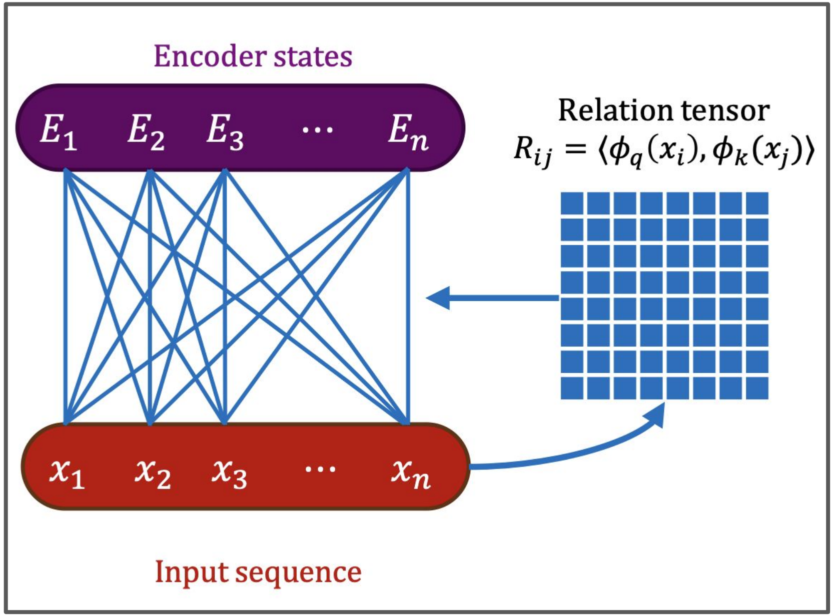
- Output embeddings only encode the **relational information** without any semantic information

- Positional symbols
- Position-relative symbols

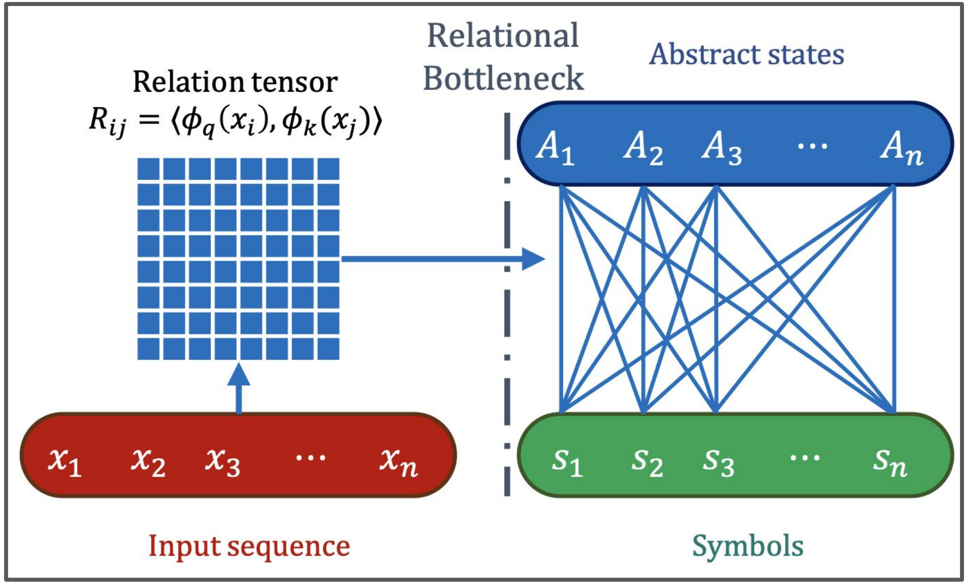
Standard Attention Mechanisms



⇒ Relational Cross-Attention



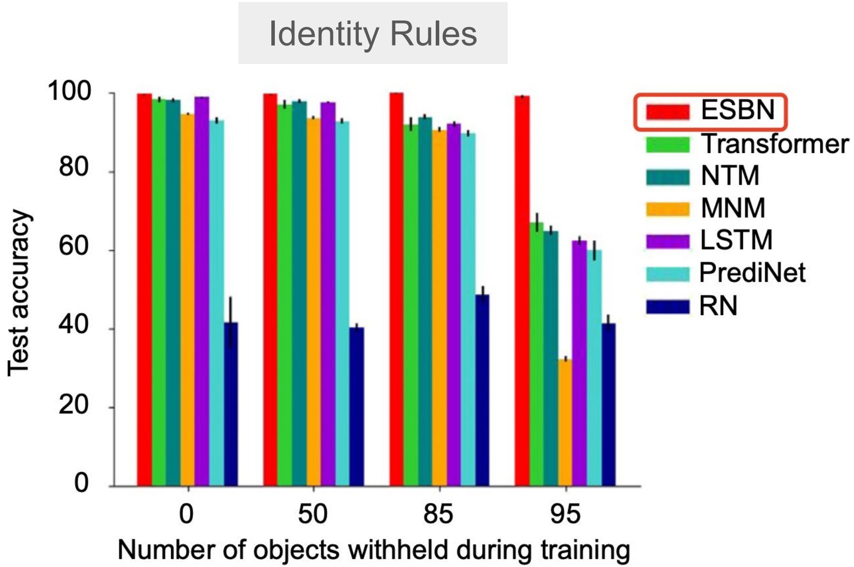
Self Attention

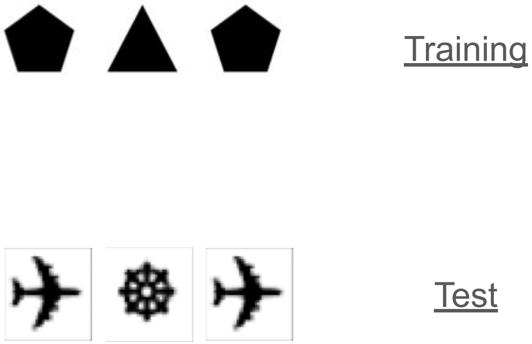


Relational Cross Attention

Advantages of Relational Bottleneck

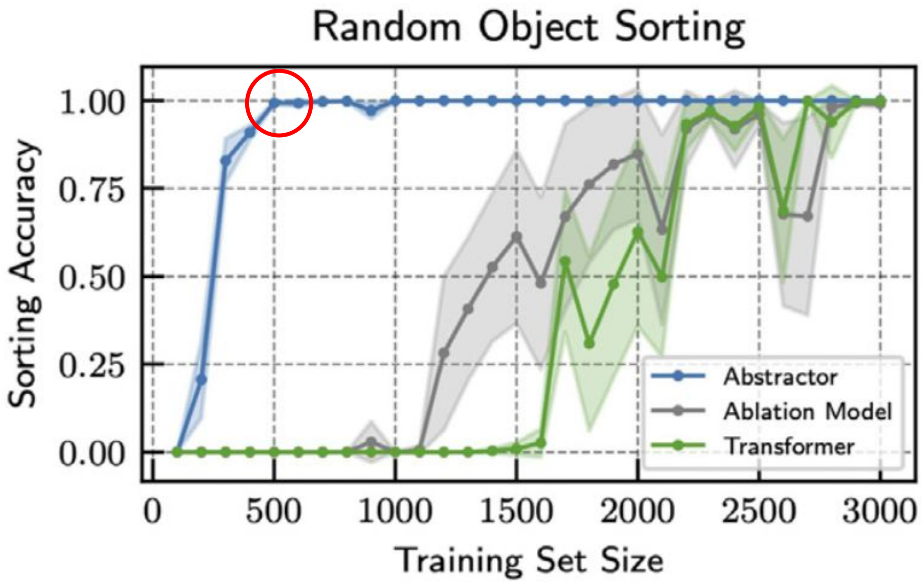
- Better out-of-distribution generalization





Advantages of Relational Bottleneck

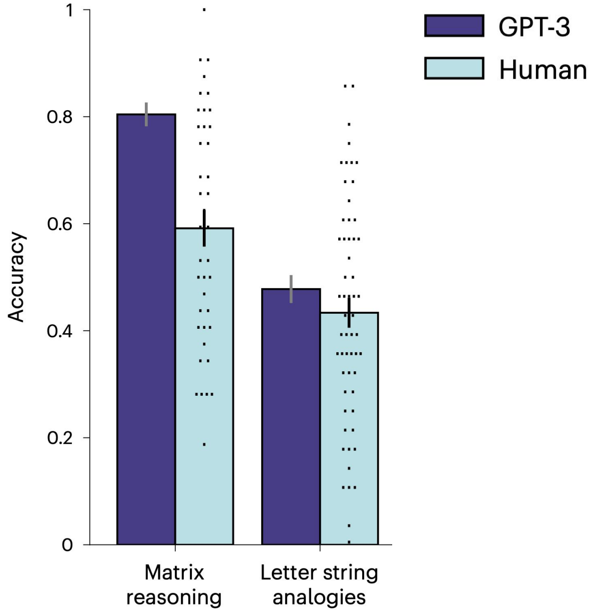
- Data-efficient Learning
- Faster Learning

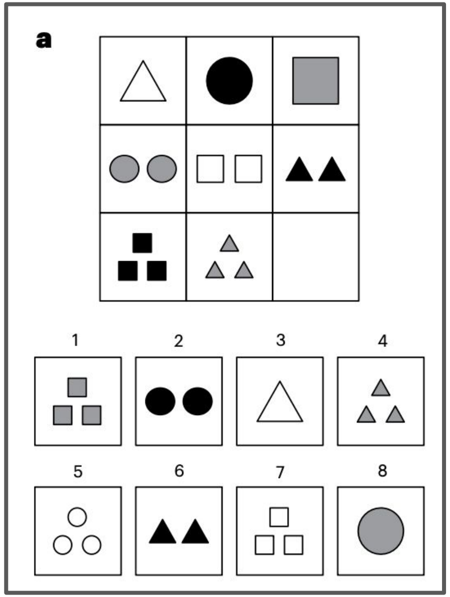


No explicit inductive bias

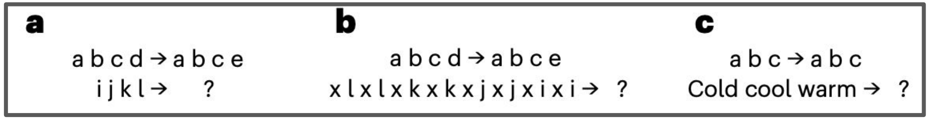
Does relational bottleneck **emerge** in *general*-purpose neural architectures (**Large Language Models**)?

Generic Large Language Models (LLMs) Show Abstract Reasoning Abilities





Matrix Reasoning

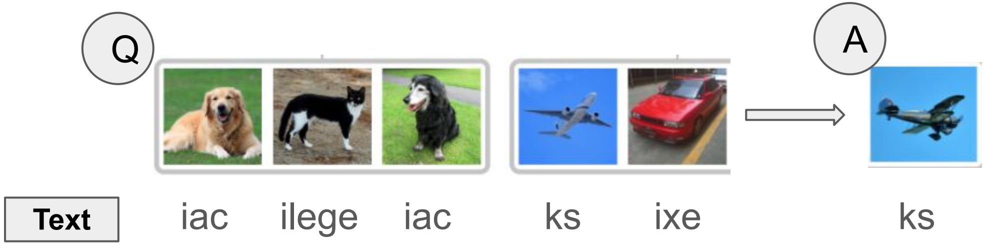


Letter String Analogies

Taylor, Webb, et al,. "Emergent analogical reasoning in large language models." *Nature Human Behaviour* 7.9 (2023): 1526-1541.

Emergent Symbolic Processing Mechanism in LLMs

- Generative Identity Rule Tasks



- Pretrained Language Model: Llama 3.1

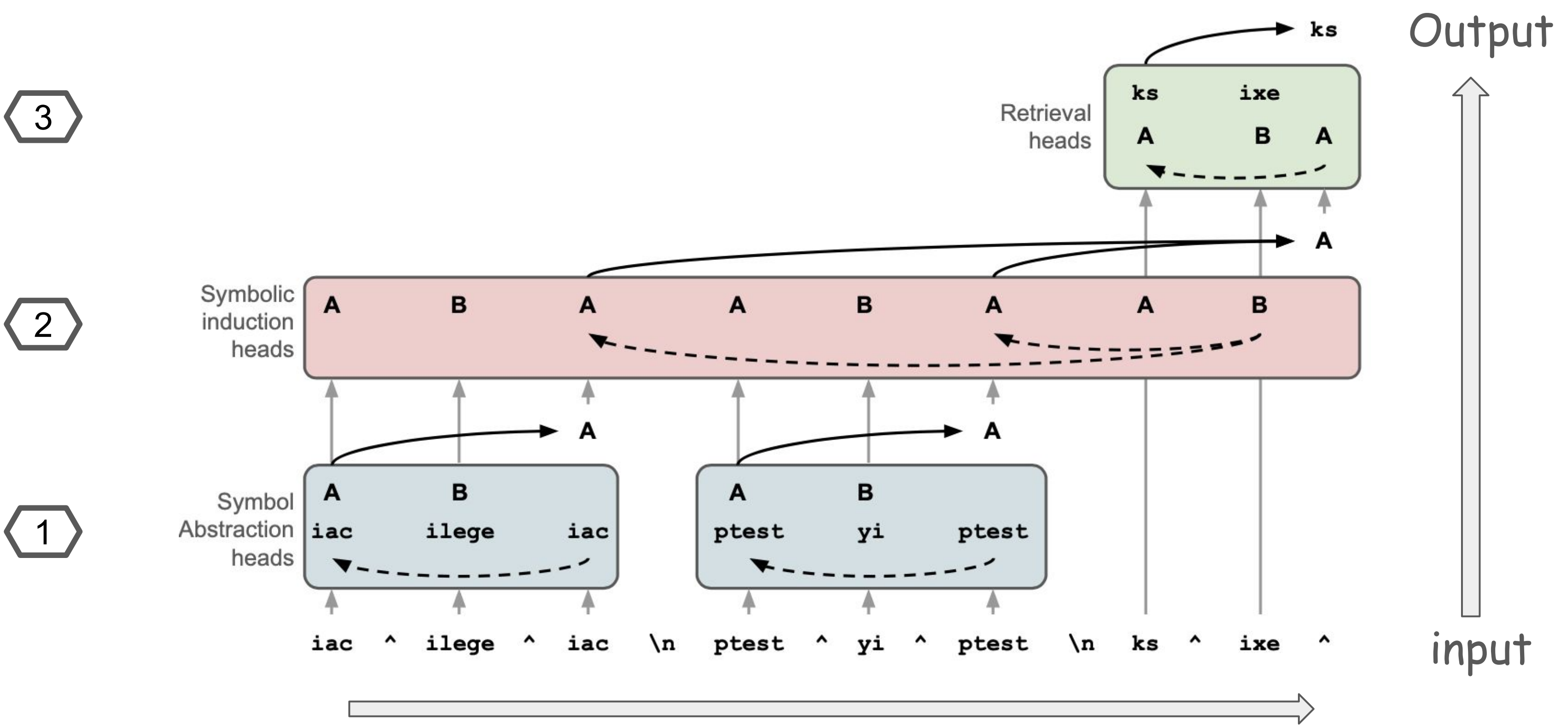
❖ **70B** parameters
❖ Trained on **~15** trillion tokens

- Performance:

  95% accuracy with 2 in-context examples

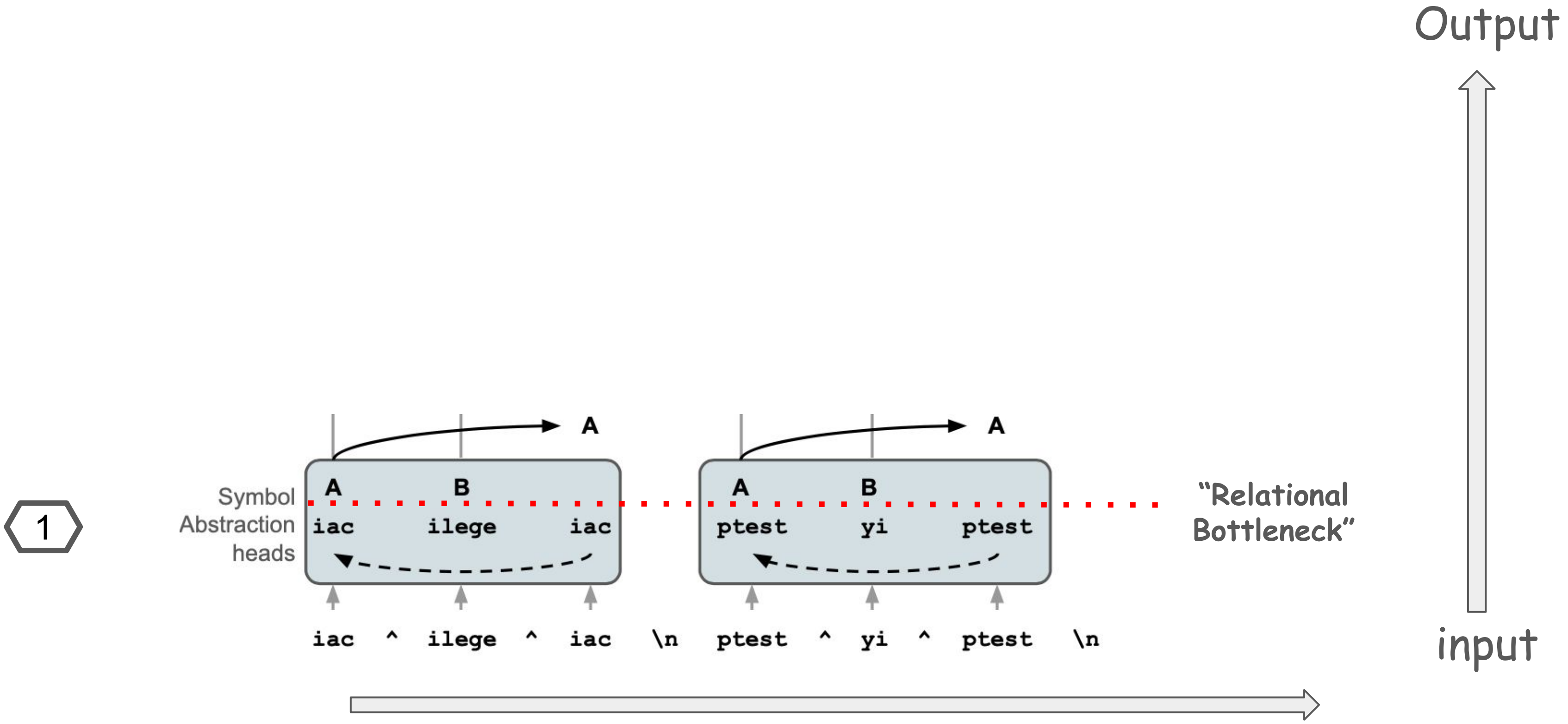
Yang, Yukang, et al. "Emergent symbolic mechanisms support abstract reasoning in large language models." *arXiv preprint arXiv:2502.20332* (2025).

Emergent Symbolic Processing Mechanism in LLMs



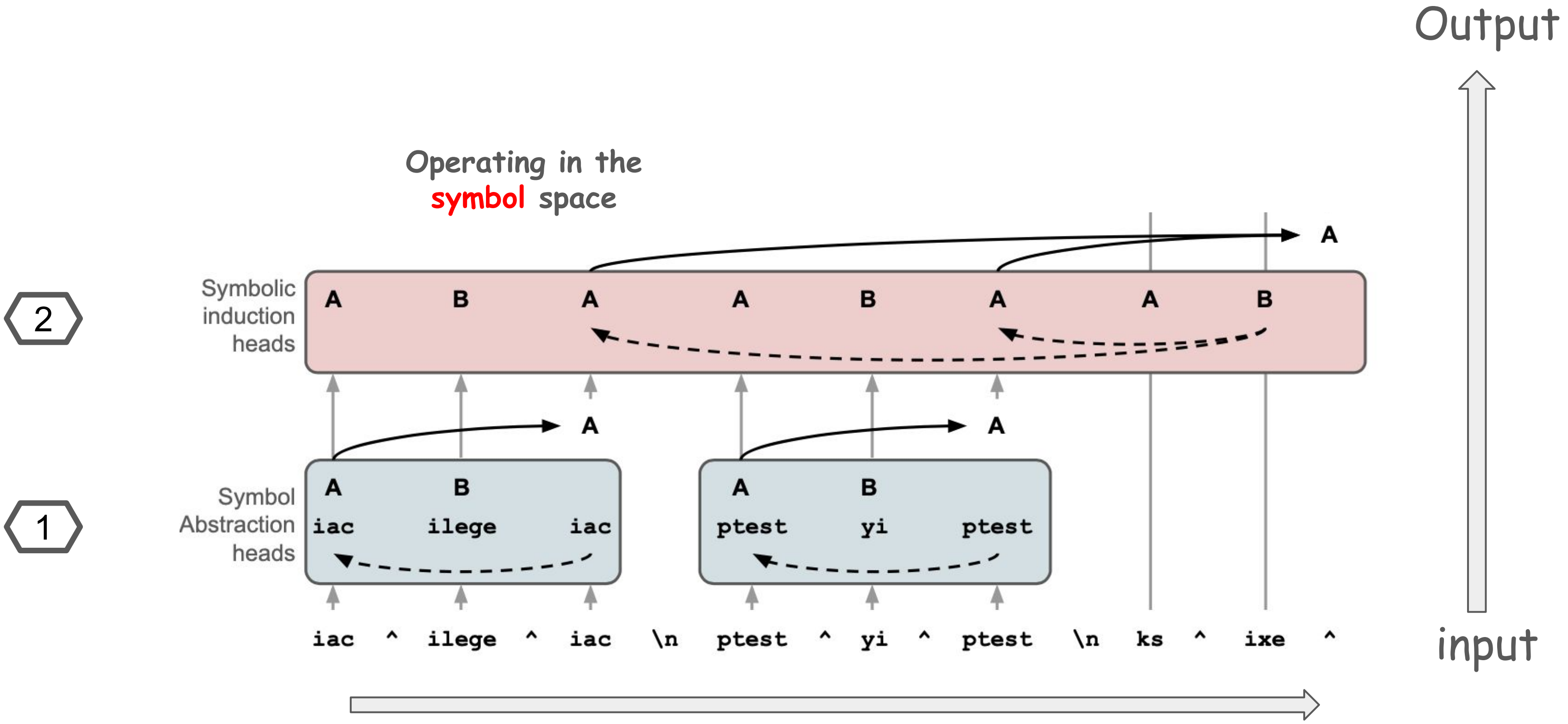
Yang, Yukang, et al. "Emergent symbolic mechanisms support abstract reasoning in large language models." *arXiv preprint arXiv:2502.20332* (2025).

Emergent Symbolic Processing Mechanism in LLMs



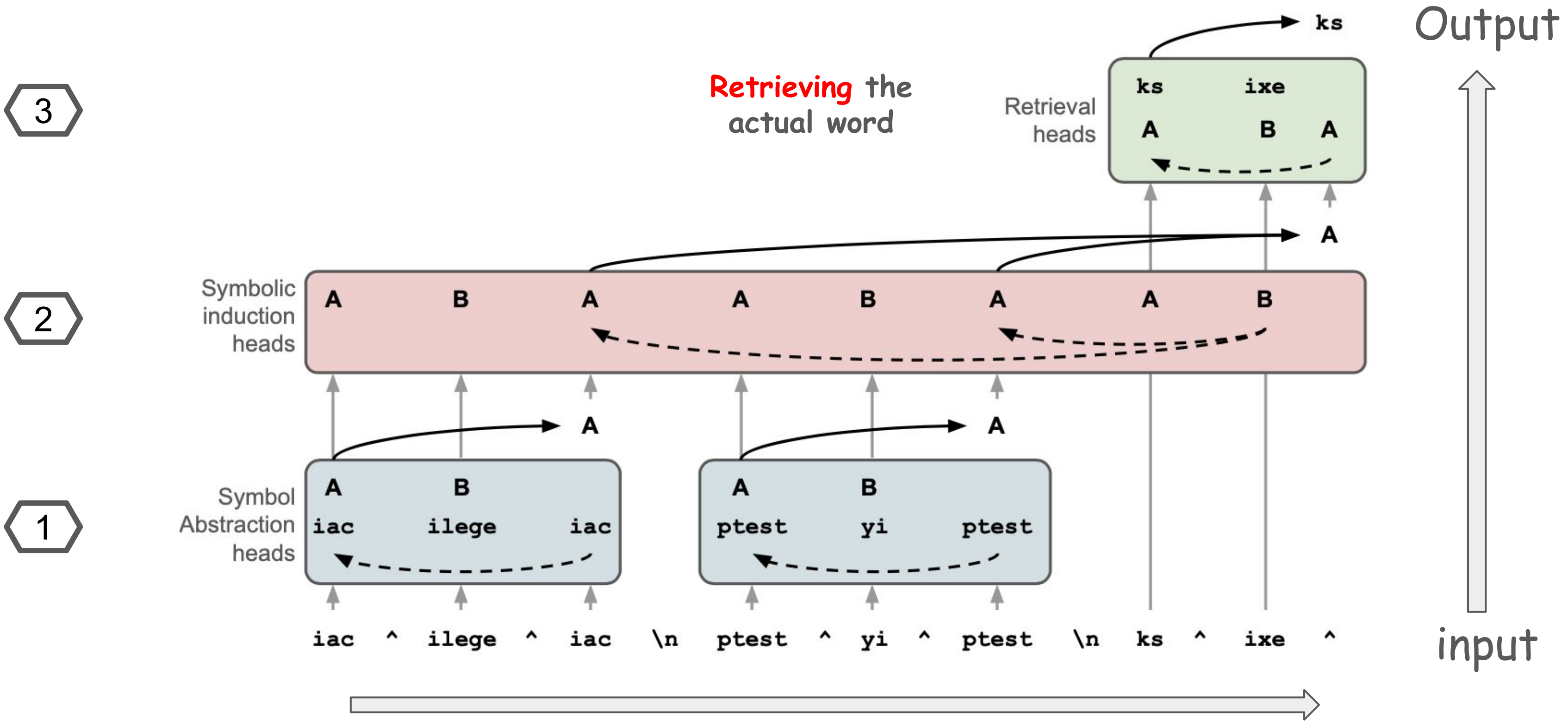
Yang, Yukang, et al. "Emergent symbolic mechanisms support abstract reasoning in large language models." *arXiv preprint arXiv:2502.20332* (2025).

Emergent Symbolic Processing Mechanism in LLMs



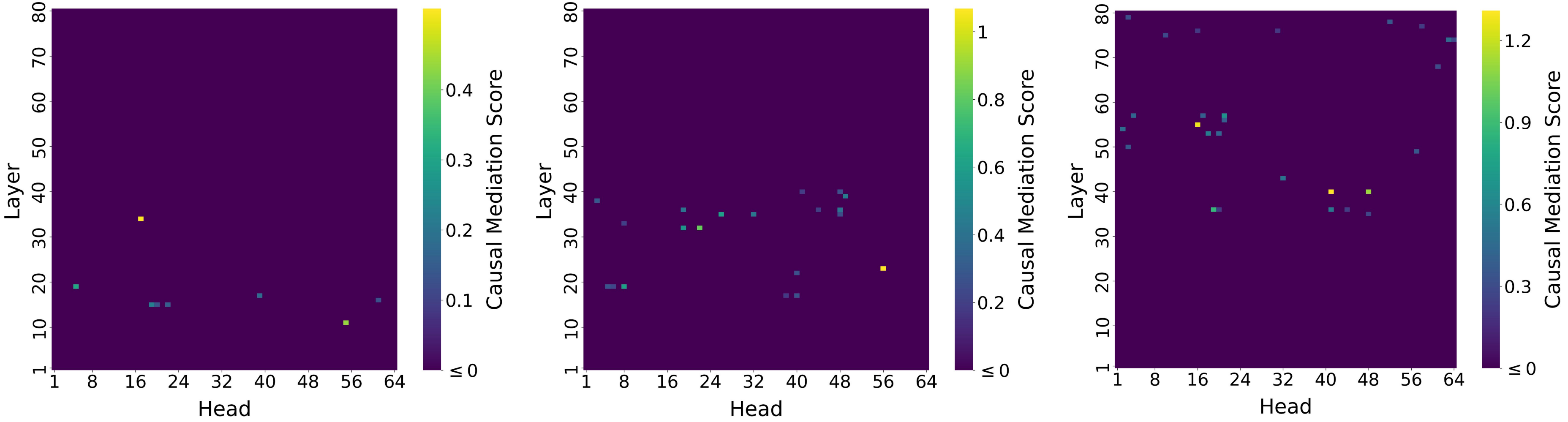
Yang, Yukang, et al. "Emergent symbolic mechanisms support abstract reasoning in large language models." *arXiv preprint arXiv:2502.20332* (2025).

Emergent Symbolic Processing Mechanism in LLMs



Yang, Yukang, et al. "Emergent symbolic mechanisms support abstract reasoning in large language models." *arXiv preprint arXiv:2502.20332* (2025).

Using Causal Mediation Analyses to identify significant heads



Symbol Abstraction Head

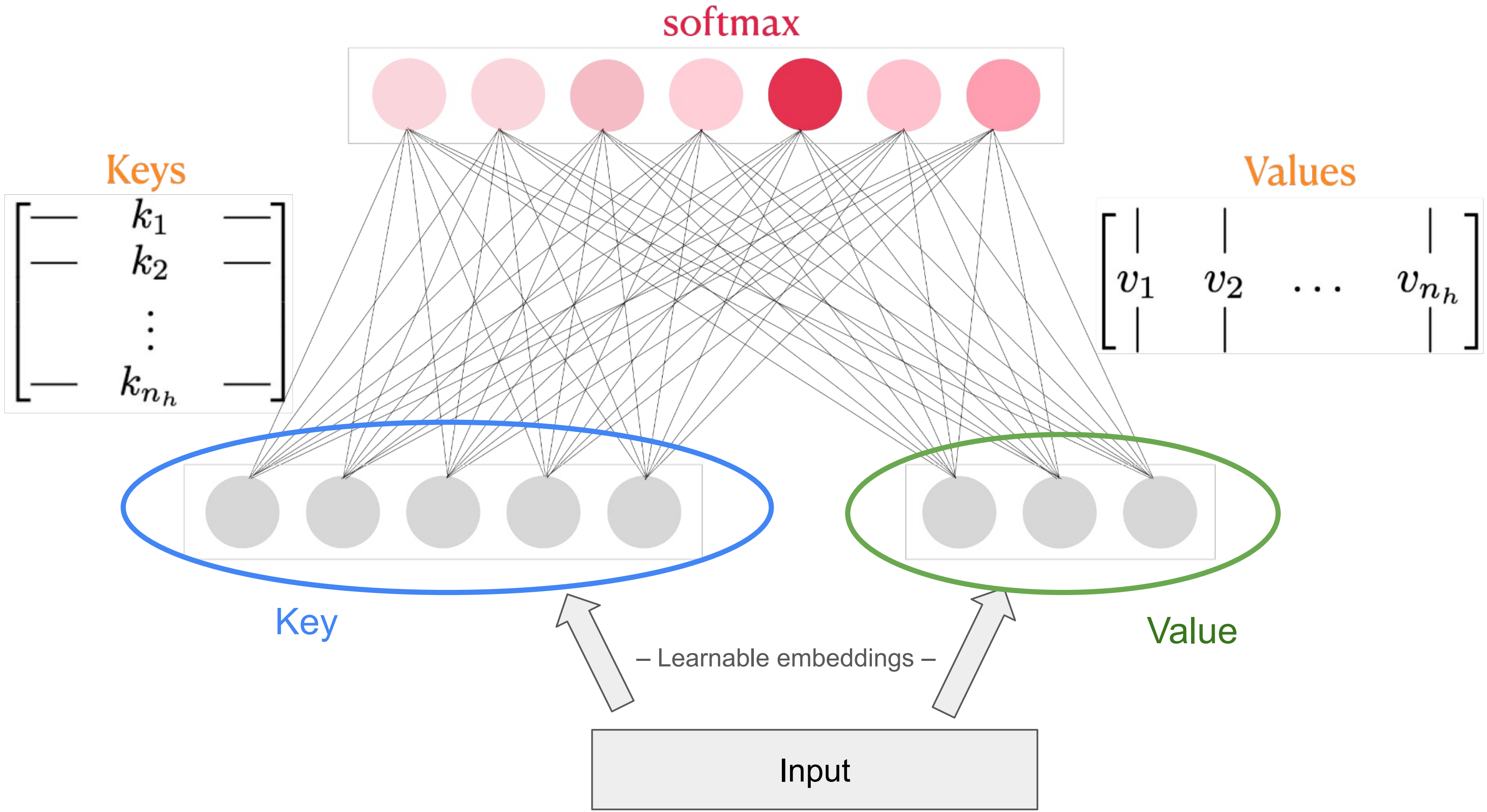
Symbolic Induction Head

Retrieval Head

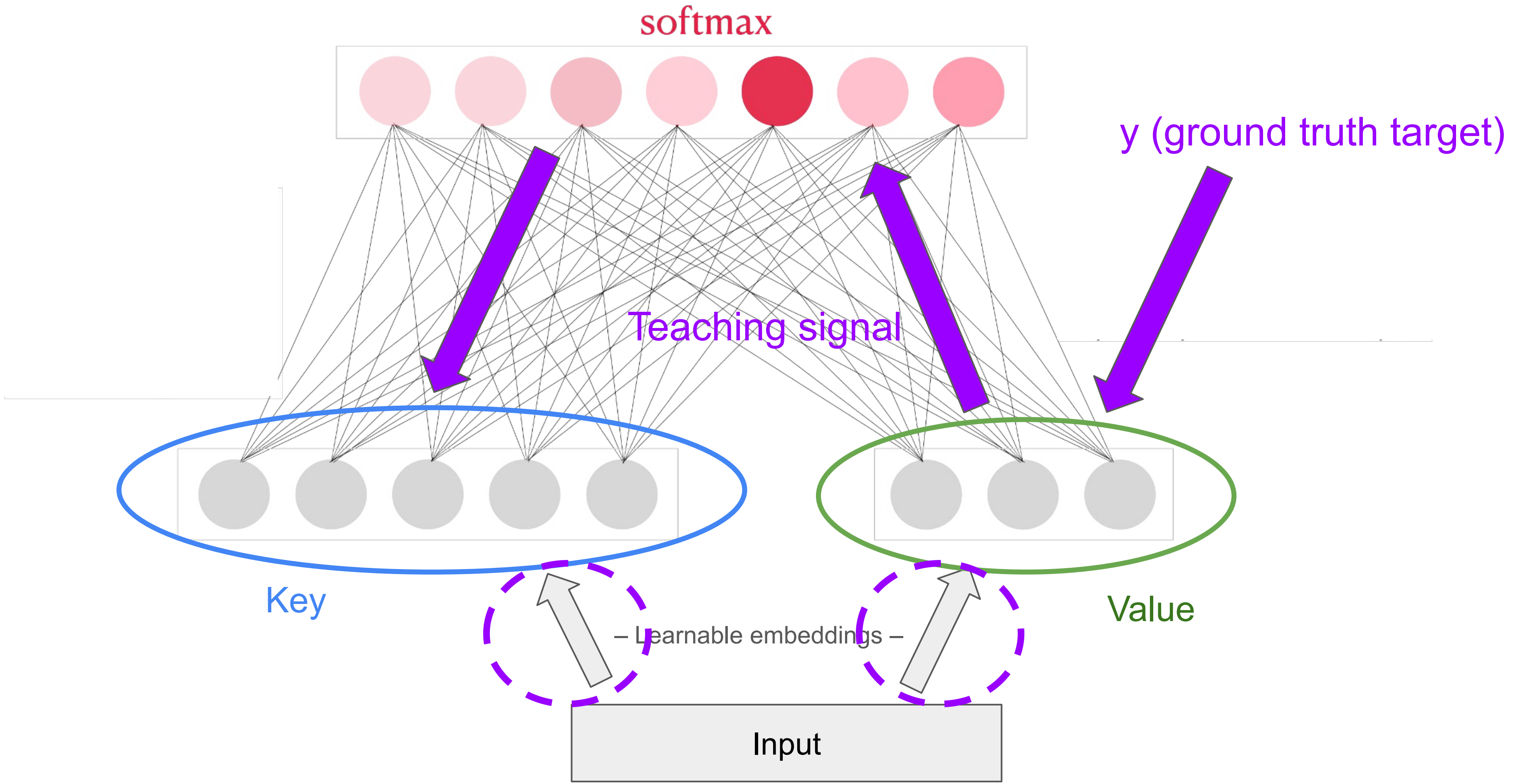
The Relational Bottleneck in the Mind and Brain

- ★ The relational bottleneck can model **inductive transitions** in early development
  - ○ Ex. children learning to gradually count to ~4 and then rapidly learning to count much higher
  - ○ A transition from memorization → abstraction
- ★ At a neural level, the relational bottleneck may arise as a consequence of **episodic memory** (e.g. hippocampus)
  - ○ Relates to deeper underlying computational principles of conjunctive versus compositional coding that could explain the tradeoff between processing and representational efficiency

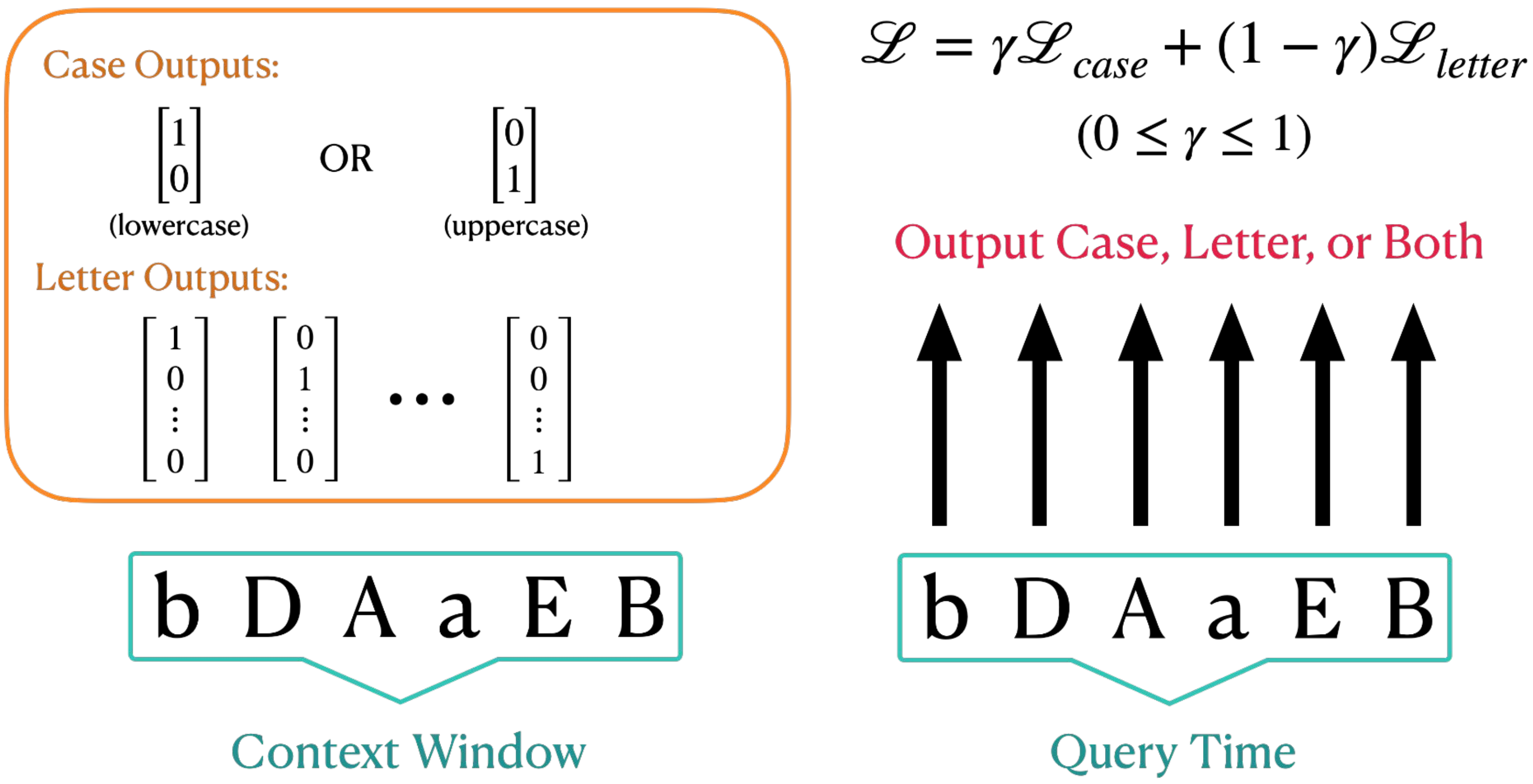
EM as a Hetero-associative Modern Hopfield Network

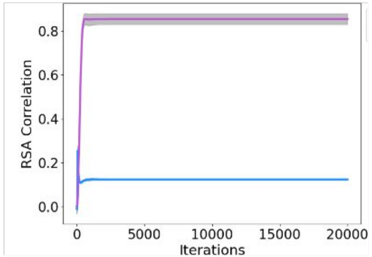
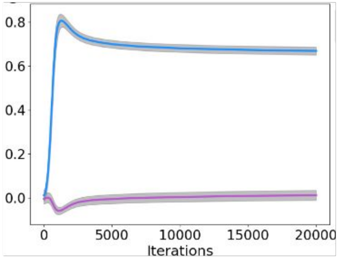
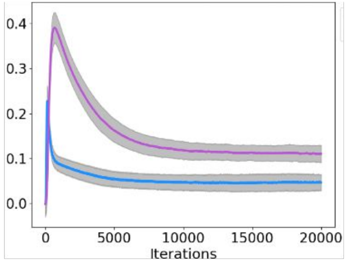
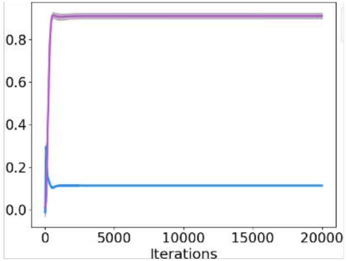


EM as a Hetero-associative Modern Hopfield Network



The Case-Letter Task



Nonlinearity Used	Linear	Softmax	ReLU	Tanh
EM Model				

Thank you! Any questions?