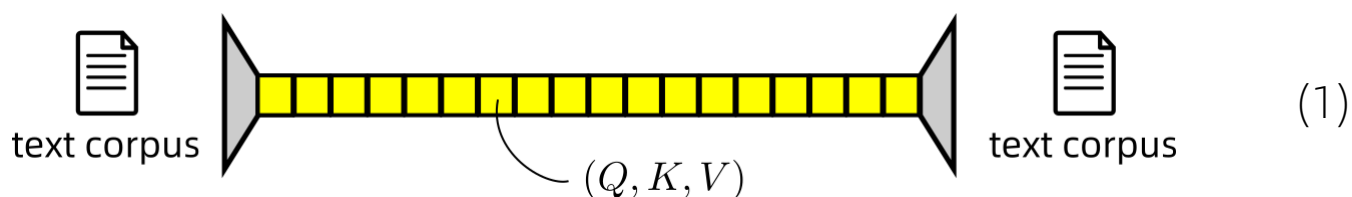


①

# Why I think reinforcement learning can solve the hallucination problem

The Transformer is an extremely **compressive** algorithm, due to its sharing the  $Q, K, V$  matrices across every layer (though there may be 100's of layers):



As a learning machine has only a finite number of weights (parameters), when it is “forced” to predict and re-construct the data, it must learn the abstract rules or patterns of the data.

**Occam's razor** is the idea that knowledge is acquired by finding the simplest explanations of the world (data).

Reinforcement learning is special in that it puts logic inference into a **closed loop**, ie, repeatedly iterating a **transition function**:



Under this setting, the transition function will be forced to explain the world with a compact set of rules, thus acquiring the knowledge / intelligence needed to discern truth from falsehood, thus solving the hallucination problem.

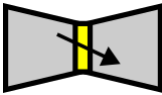
Indeed, the GPT / Transformer is already doing something like this, although it does not have an explicit closed-loop as in RL, it is still trained with an implicit loop over all data (the corpus).

My hope is for RL's closed loop to increase the **re-use** of logic rules, thus increasing the model's intelligence – according to Occam's razor.

2

## Two architectures

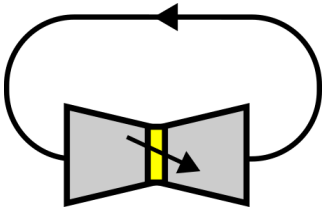
Recall that the **Auto-Encoder** (AE) has the dual functions of **compression** and **predicting the future**, so we use this symbol for it:



(3)

We can think of two architectures, based on how we interpret the functions of an Auto-Encoder.

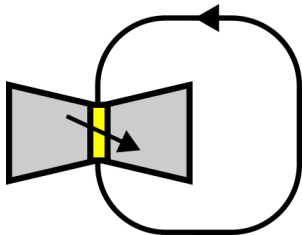
**Architecture #1**: AE emulates human thinking.



(4)

Thoughts = natural-language sentences.  
RL puts inference into a closed loop.  
Thoughts describe the world.

**Architecture #2**: AE = compressed world model.



(5)

Thoughts = hidden / latent state.  
RL helps AE to explain the world,  
successful explanations reward RL.

## Which architecture is correct?

It may appear strange that AE has two different interpretations. Because

human thinking  $\subset$  world, (6)

we can see that human thinking, as appears in text corpuses, can sometimes be wrong due to eg. lying, false beliefs, etc. Therefore Arch #1 can be faulty and may be the origin of hallucinations.

③

# Why iteration is important

Consider the following simplified scenario:

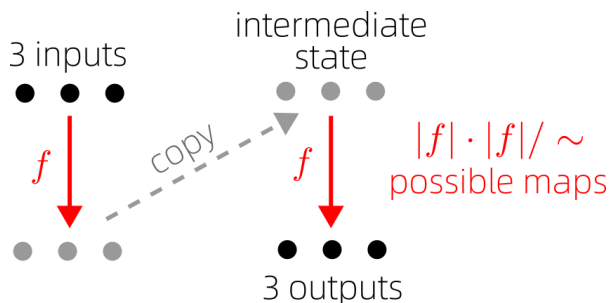
3 binary inputs,  $2^3 = 8$  states =  $\{0...7\}$

$f: 8 \rightarrow 8 = 8^8 = 16777216$  possible maps

(7)

3 binary outputs,  $2^3 = 8$  states =  $\{0...7\}$

If  $f \in \mathcal{F} = 8^8$  then all possible mappings are learnable. In general the family of  $f$  may be a subset of all possible maps, as long as some solution(s) exist in  $\mathcal{F}$ .



(8)

④

## About our group

We operate as a **DAO** (decentralized autonomous organization) based on transparent operations and reward system based on weighted voting, to enable global collaboration without racial (or other forms of) discrimination.

Our values:

- democracy
- freedom of speech
- racial equality
- transparency
- tolerance of mistakes
- a learning environment

It is OK for anyone to challenge other member's theories, ideas, proposals, etc.