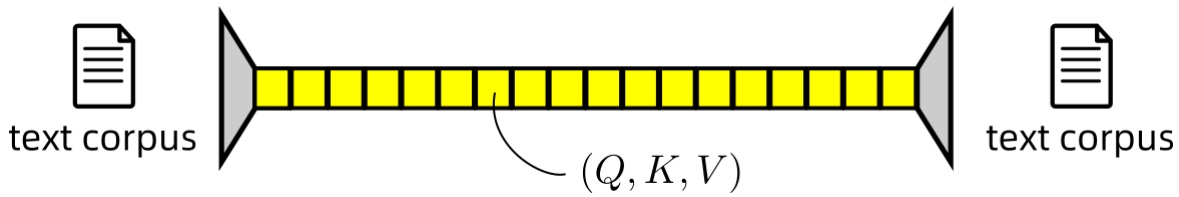
Why I think reinforcement learning can solve the hallucination problem

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):

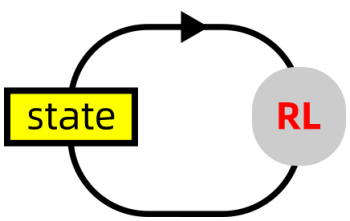


(1)

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**:



(2)

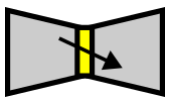
Under this setting, the transition function will be forced to explain the world with a compact set of rules, thus acquiring the knowledge / intelligence needed to discern truth from falsehood, thus solving the hallucination problem.

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.

Two architectures

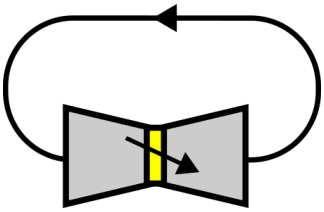
Recall that the **Auto-Encoder** (AE) has the dual functions of **compression** and **predicting the future**, so we use this symbol for it:



(3)

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

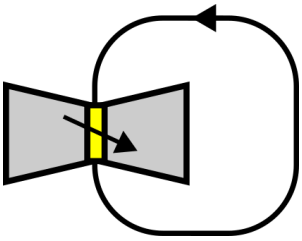
Architecture #1: AE emulates human thinking.



(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?

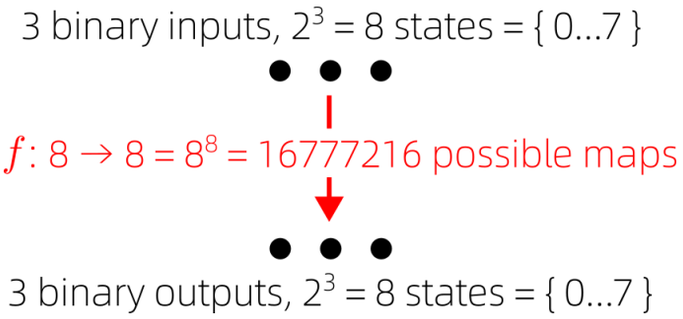
It may appear strange that AE has two different interpretations. Because

$$\text{human thinking} \subset \text{world}, \tag{6}$$

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.

Iteration helps compression

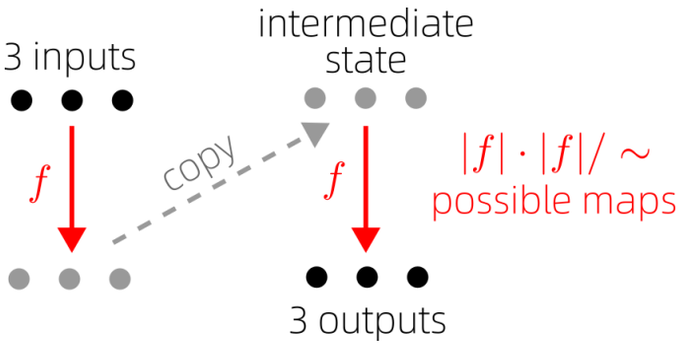
Consider the following simplified scenario:



(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 $\mathcal{F}$.

What if f is iterated?



(8)

Then the number of maps $f \circ f$ is $|f| \cdot |f|$ quotiented by redundancy.

In neural networks, $f \circ f$ is two neural networks **stacked** together. It has **twice** the number of layers (and weights). In general, the number of mappings of a neural network grows as the **exponent** of the number of layers[1], which is consistent with $|f| \cdot |f|$.

In other words, the number of maps representable by a big neural network is the same as if the network is broken into smaller pieces, as long as the layers add up to the same number. However, if the weights of each piece are **shared**, then it obviously results in more compact **information compression**, while training time $\propto$ number of weights which remains the same (shared weights counted with multiplicity).

Conclusions

- 有 loop 和没有 loop 的 **训练时间** 是差不多的，但其 **智能** 会提升。
- 中间状态的 语义是 **浮动** 的，只有通过训练才能确定。因此，GPT 训练出来的权重 不能在这里 **再用**，需要重新训练。

[1]For an intuitive explanation, think of f as a polynomial of degree d, then $f_1 \circ f_2$ has degree $d_1 \cdot d_2$. The number of "zero-crossings" of a polynomial is same as its degree. Thus in general a polynomial can "cut" the ambient space into d pieces.

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.