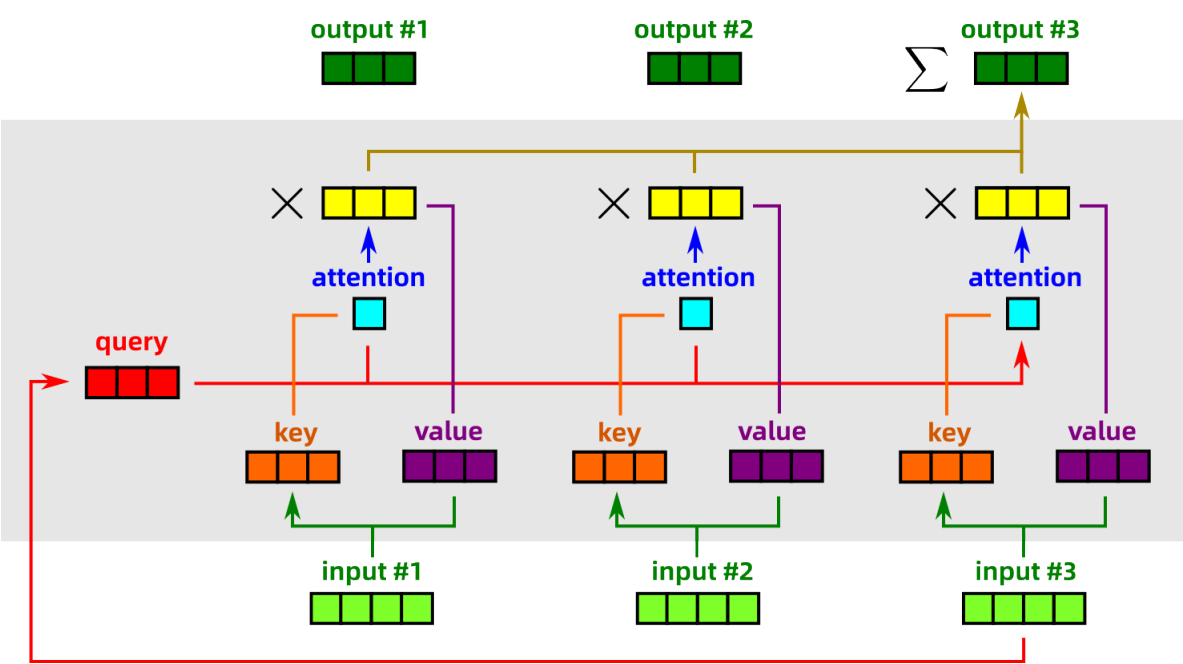


①

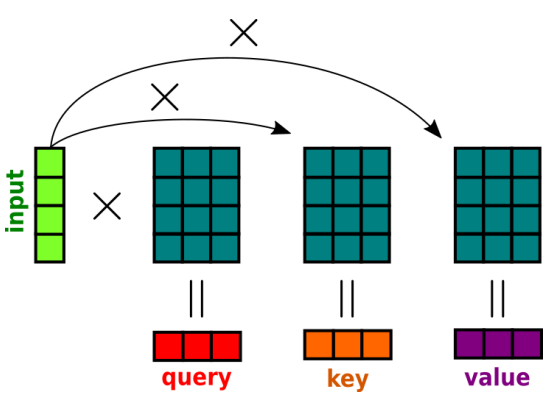
Transformer has logic structure

In this infographic I'd explain a major finding of several years of my research: that the Transformer is a symbolic-logic machine.

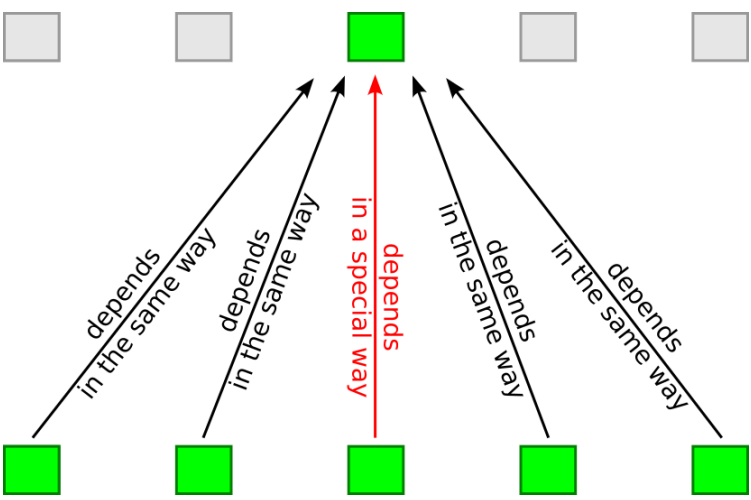
For your convenience let's refresh on the Transformer's **Self-Attention** mechanism:



"Input" tokens are translated to Q, K, V values via matrix multiplication, which can be regarded as a kind of table look-up, or **memory store**:



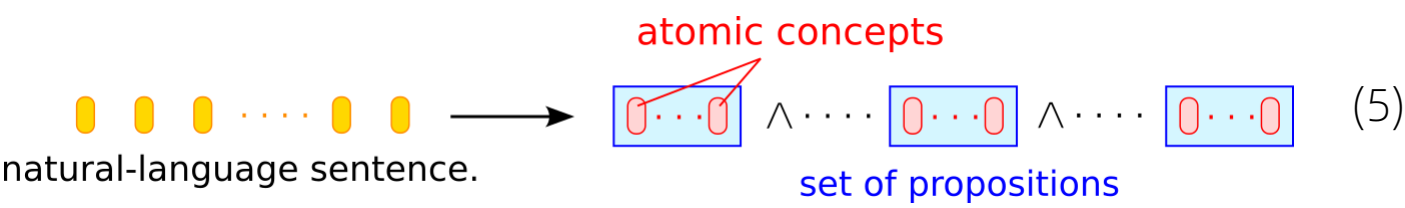
From an abstract point of view, the Transformer has this structure, which gives it the **equivariant** property:



The equivariance property corresponds to the **exchangeability** of logic propositions:

$$A \wedge B \Leftrightarrow B \wedge A \tag{4}$$

Given a natural-language sentence, we'd like to convert or **decompose** it into a bunch of logic propositions:



The structure on the right is a **mental state** of a logical AI system. It is composed of (exchangeable) propositions, which are in turn made up of atomic concepts. This 2-level structure is characteristic of all **logical** systems.