Is the attention similar to what humans do?

- *For humans:* **saves time**

Attention saves time when reading (i.e. we look only to the relevant parts of the sentence).

- *For machines:* **wastes time**

To compute the attention weights, the model carefully examines ALL the positions, thus wastes even more time.