ATTENTION
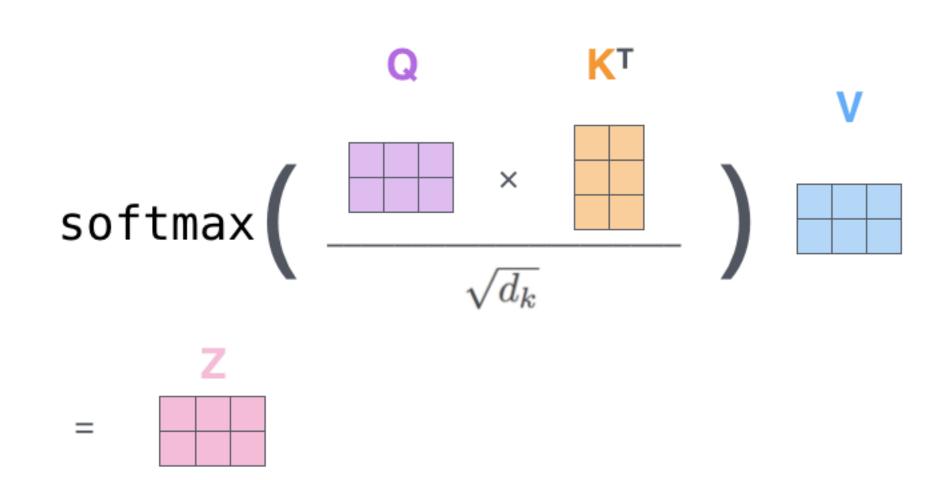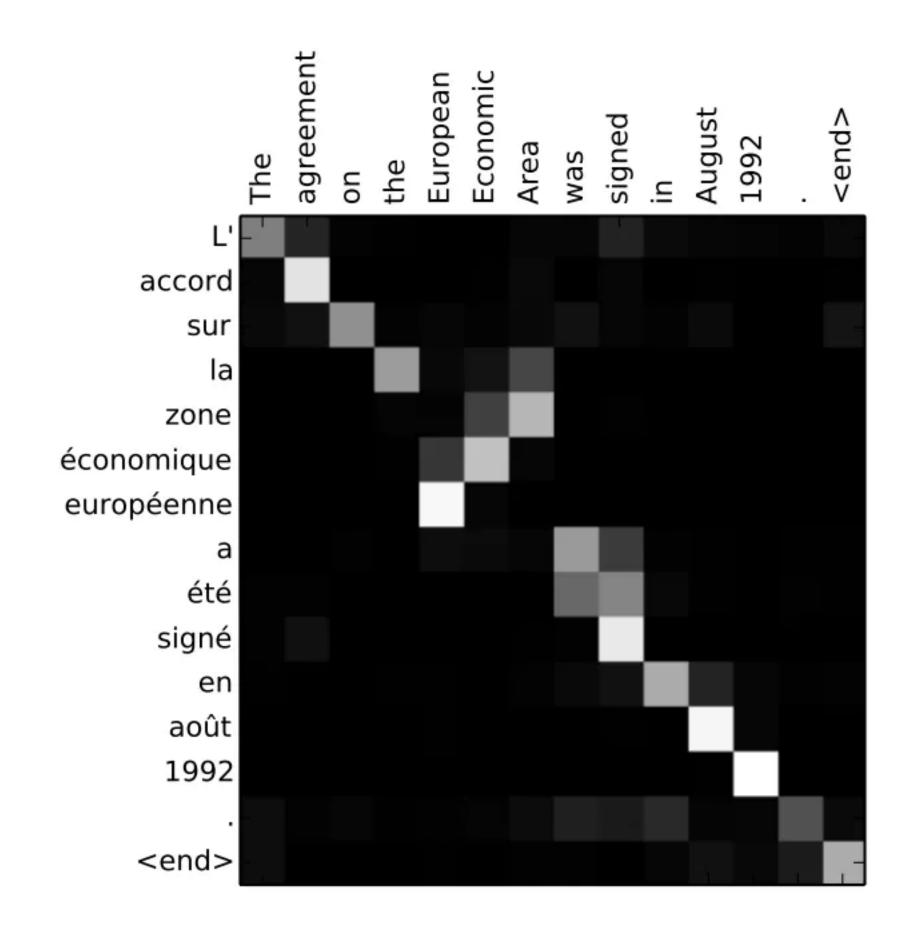
- **What was wrong with previous approaches?**

$$\text{softmax}\left(\frac{Q \times K^T}{\sqrt{d_k}}\right) V$$

$$Z =$$

THE MASKED SHOGGOTH

Unsupervised Learning

Supervised Fine-tuning

RLHF
(cherry on top 🙂)

Scott Alexander