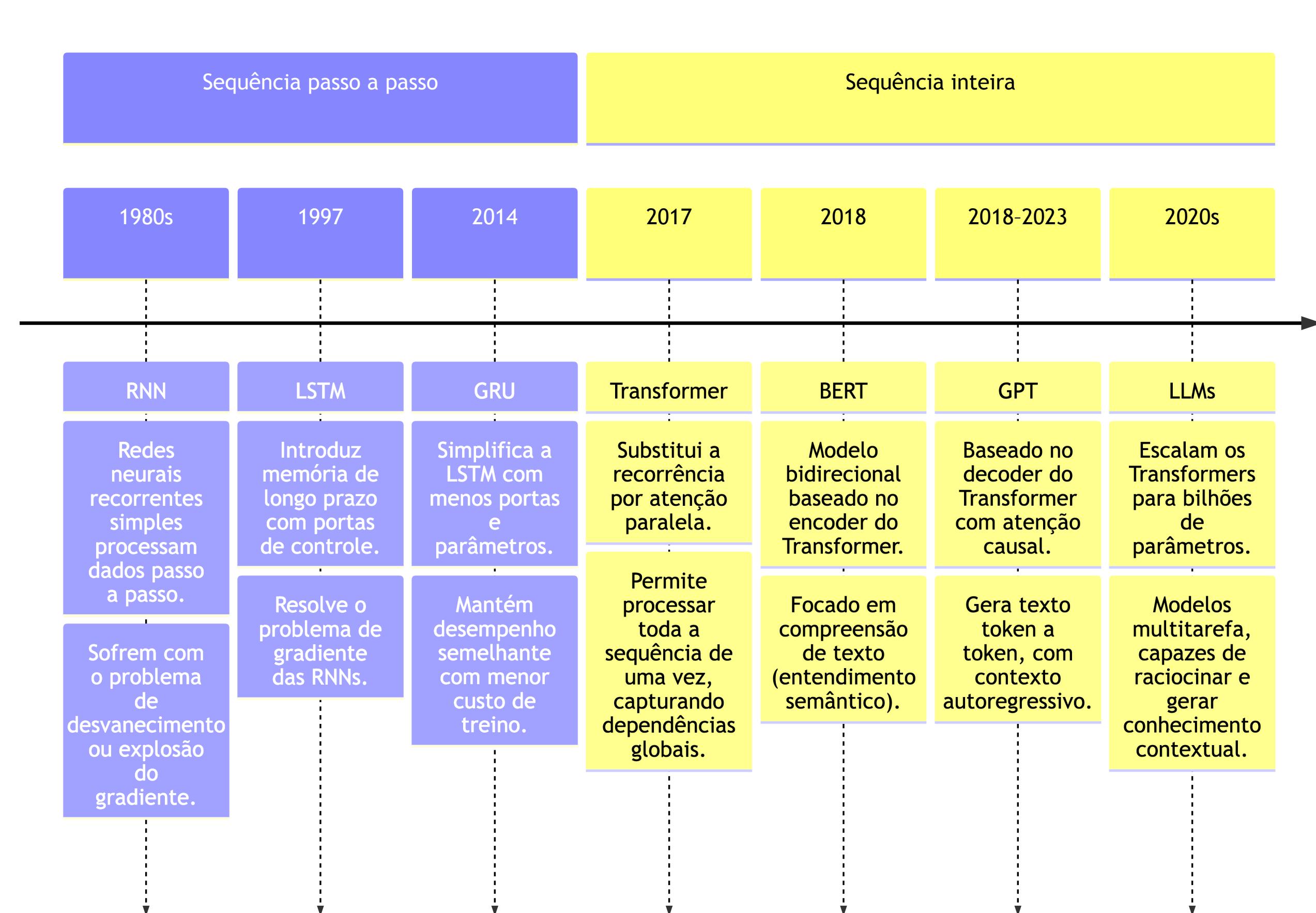
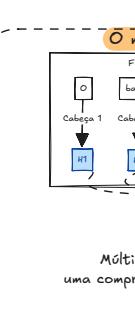


Evolução das Arquiteturas para Processamento de Sequências Variáveis



- Query
- Key (K)
- Value

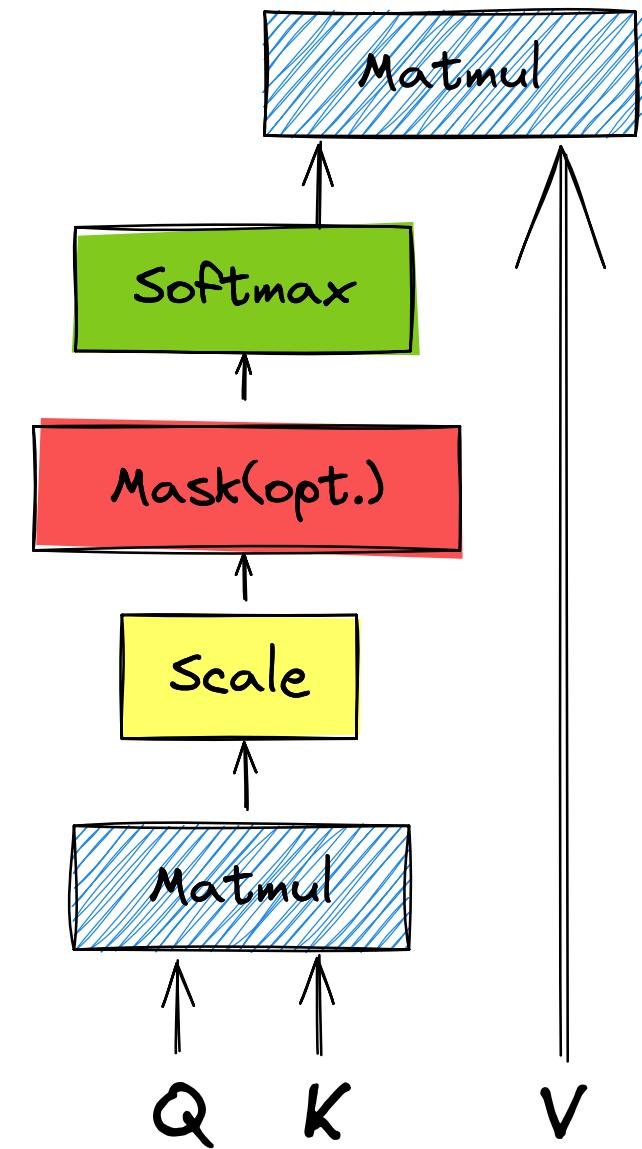
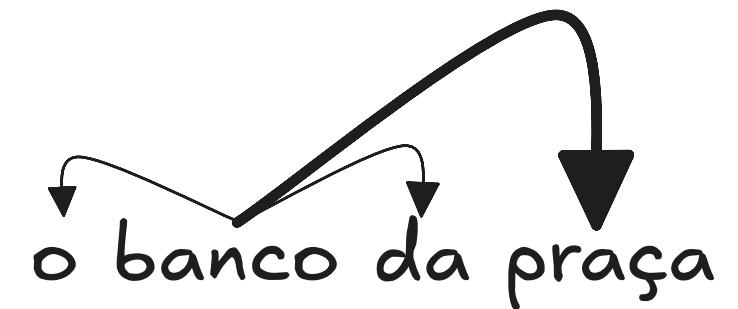


O mecanismo de atenção

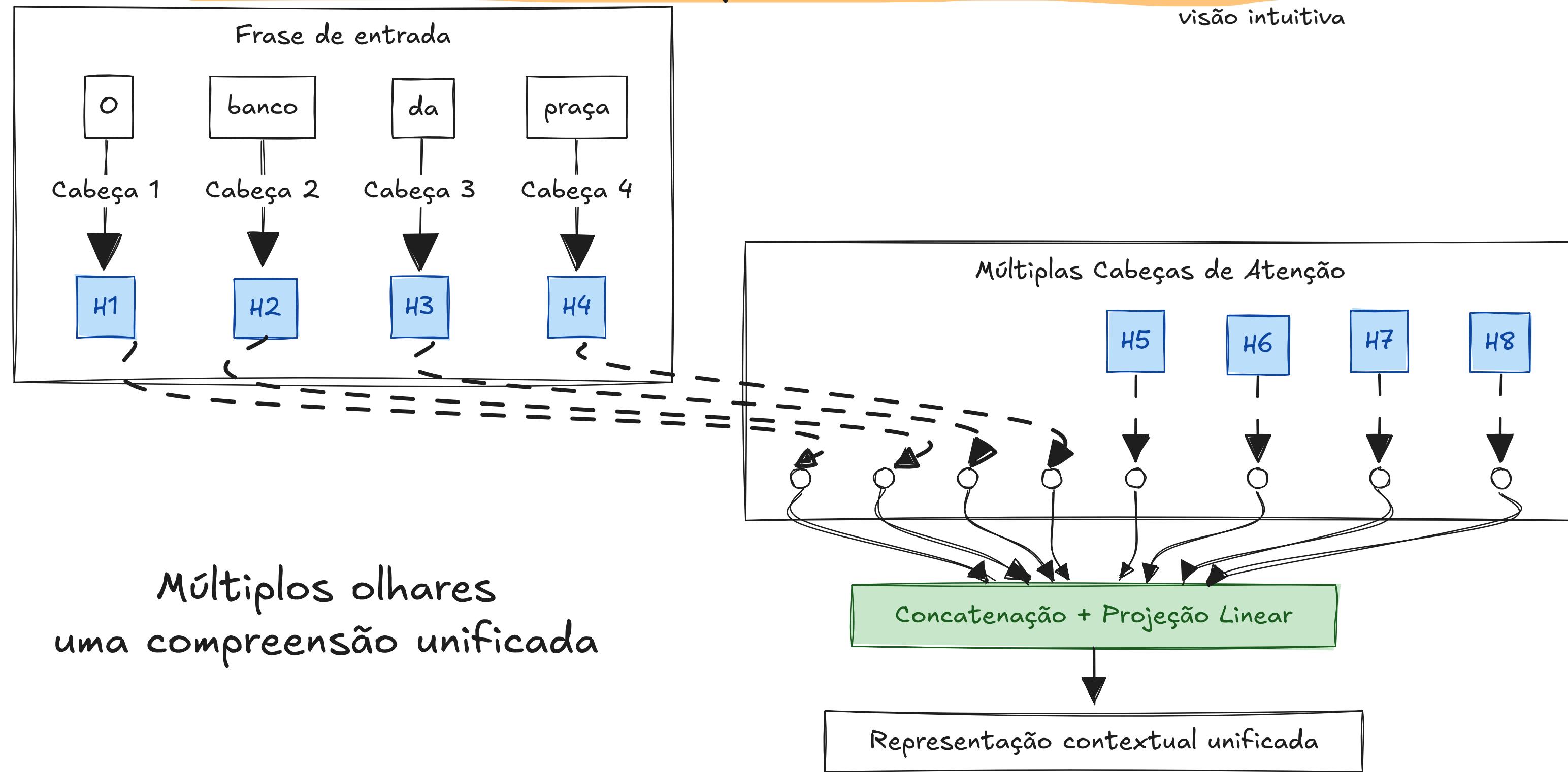
visão intuitiva

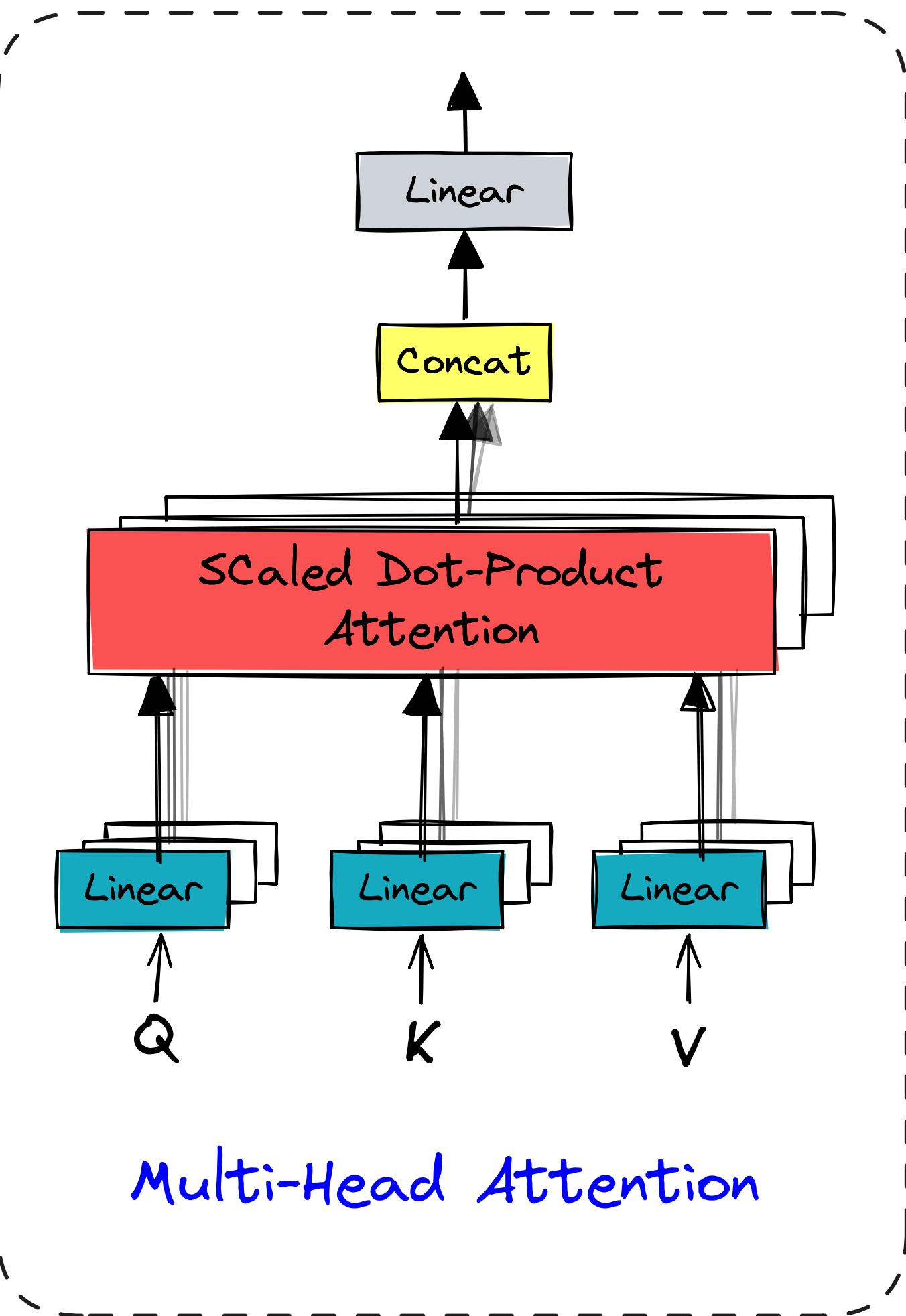
- Query (Q) — o que estou buscando;
- Key (K) — quem sou eu;
- Value (V) — o que tenho a oferecer.

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

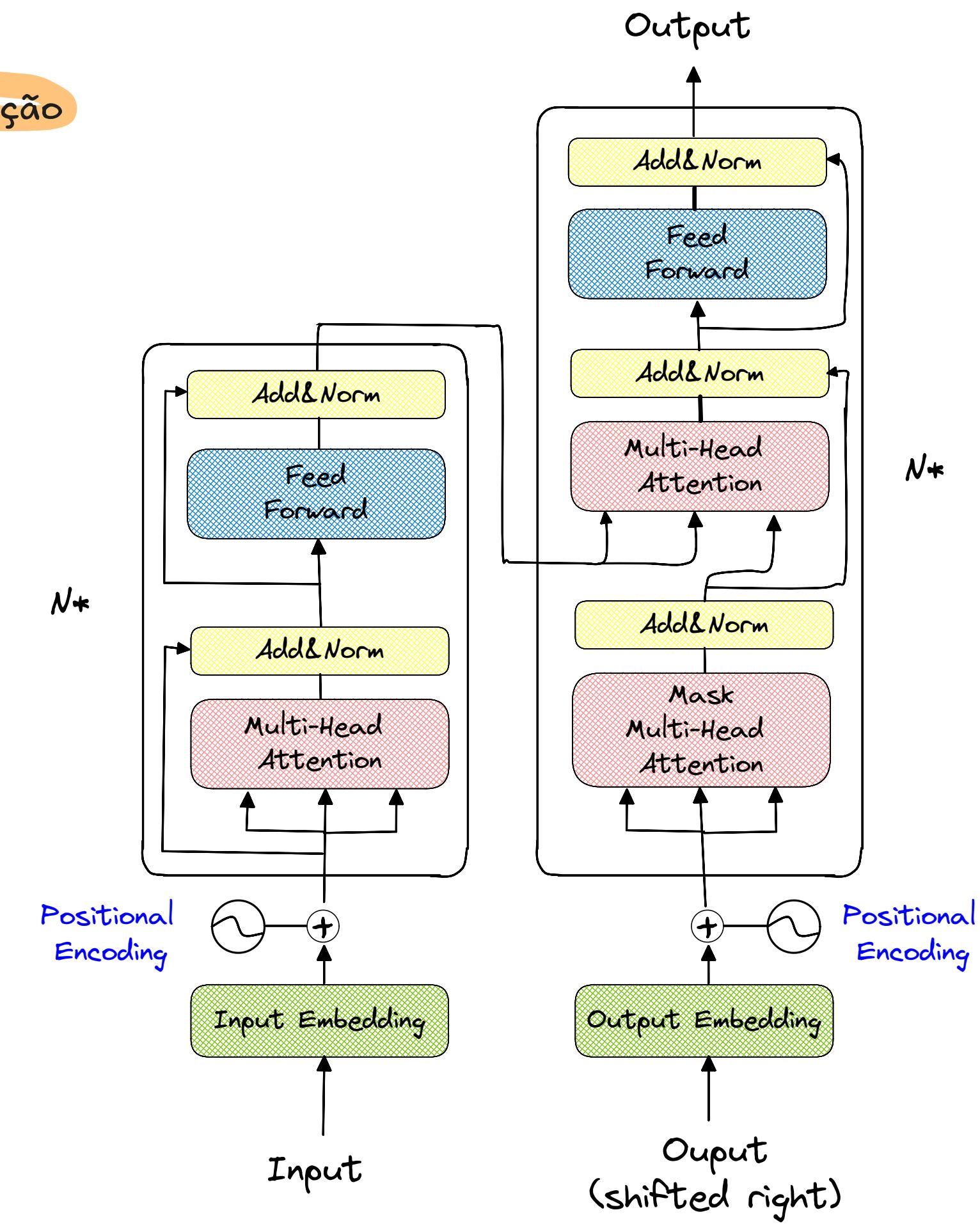


O mecanismo de múltiplas cabeças de atenção



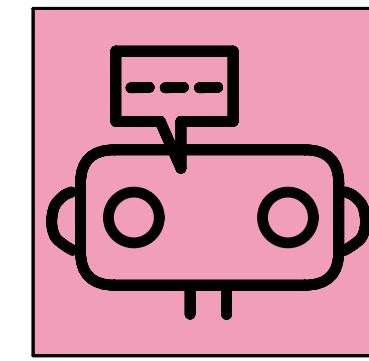
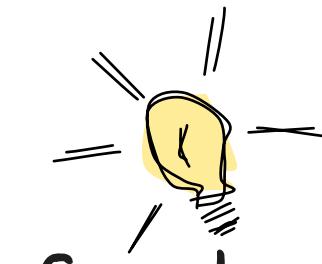
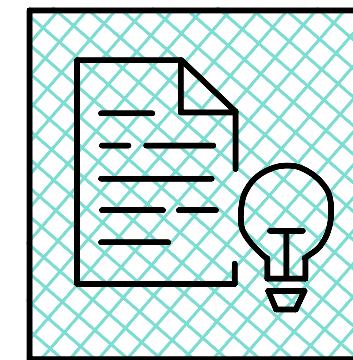
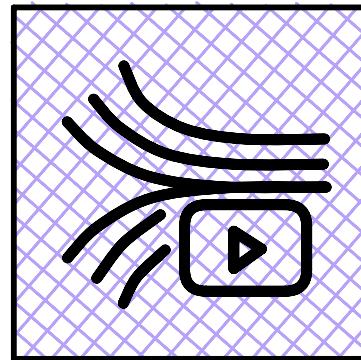
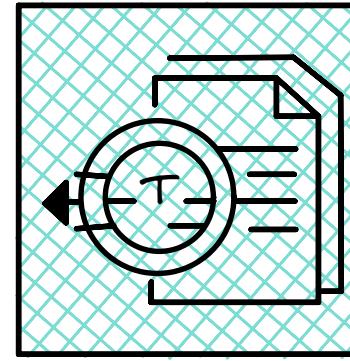


**Atenção e normalização
conduzem todo o fluxo de informação**



O legado dos Transformers para IA Moderna

- BERT — compreensão de linguagem.
- GPT — geração de texto.
- T5 — aprendizado multitarefa.
- ViT — atenção em visão computacional.



Atenção redefiniu a inteligência artificial