HELMHOLTZ AI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

DKRZ DEUTSCHES KLIMARECHENZENTRUM

Attention

AI Consultants Earth & Environment @DKRZ

Mosaku Adeniyi

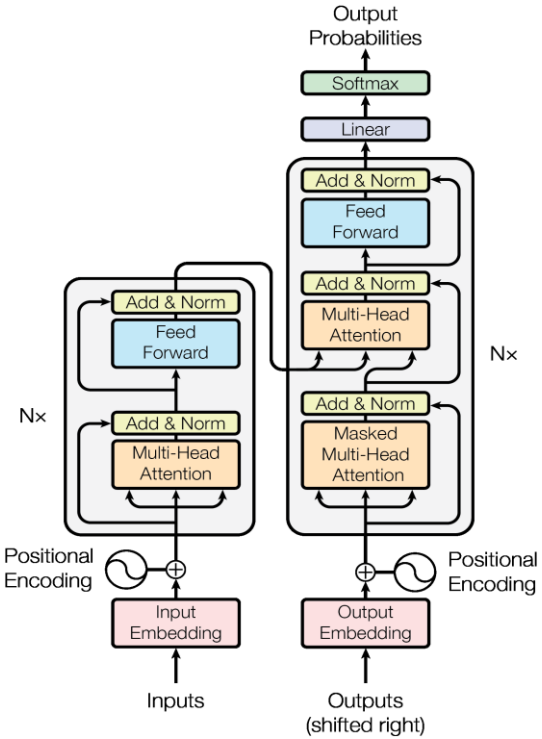
Helmholtz AI

Attention progression requirement

„Attention Is All You Need“

Self-Attention | Multihead Attention | Positional Encoding

- QKV are generated directly from the input data
- Multihead allows for more channels which are concatenated into linear layer
- Positional encoding brings about inclusion of spatial relationship / order

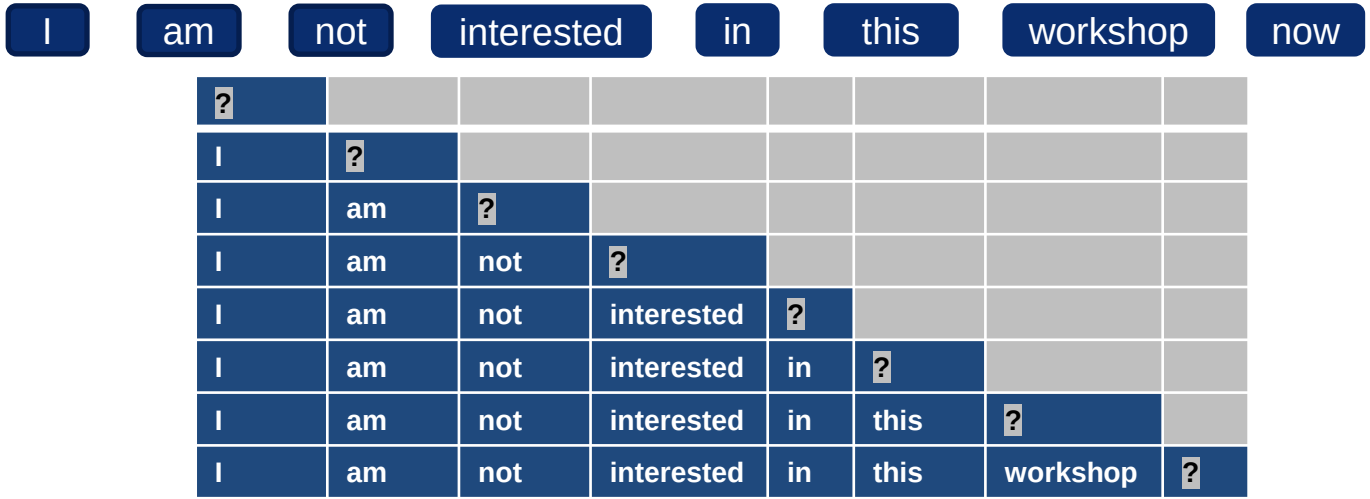


Vaswani: Attention Is All You Need

Helmholtz AI

DKRZ
DEUTSCHES
KLIMARECHENZENTRUM

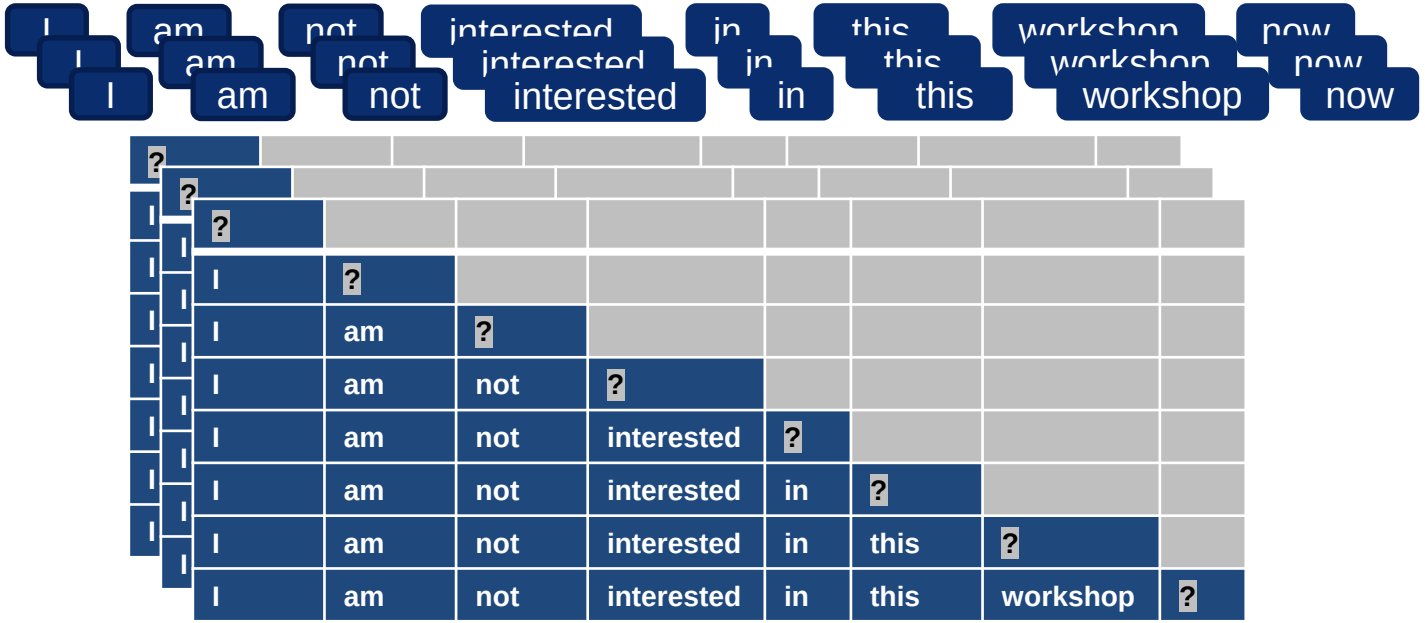
Average for communication using Matmul and Tril (Mask) Functions



- Communication through average of previous token
- Each token is thus summarized in the context of its history
- Using average assumes each token are equal, thus not efficient

Helmholtz AI

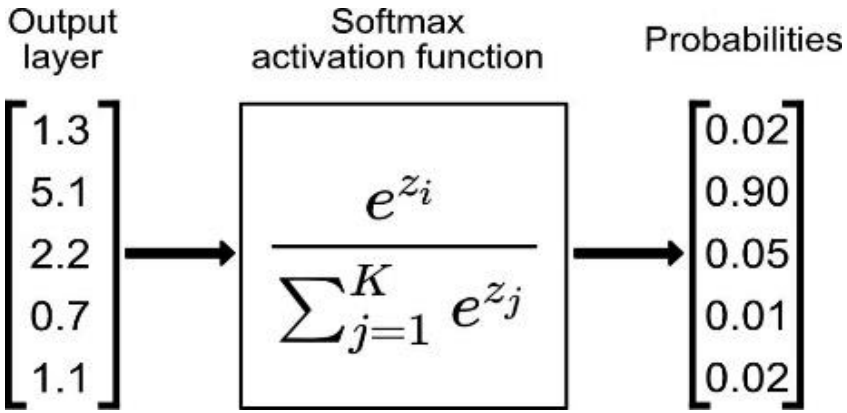
Input dimension as Batch Sequence Channel (B T C)



- Channel here just means depth, it could be RGB for images
- Batch means different group of inputs say three sentences or images

Helmholtz AI

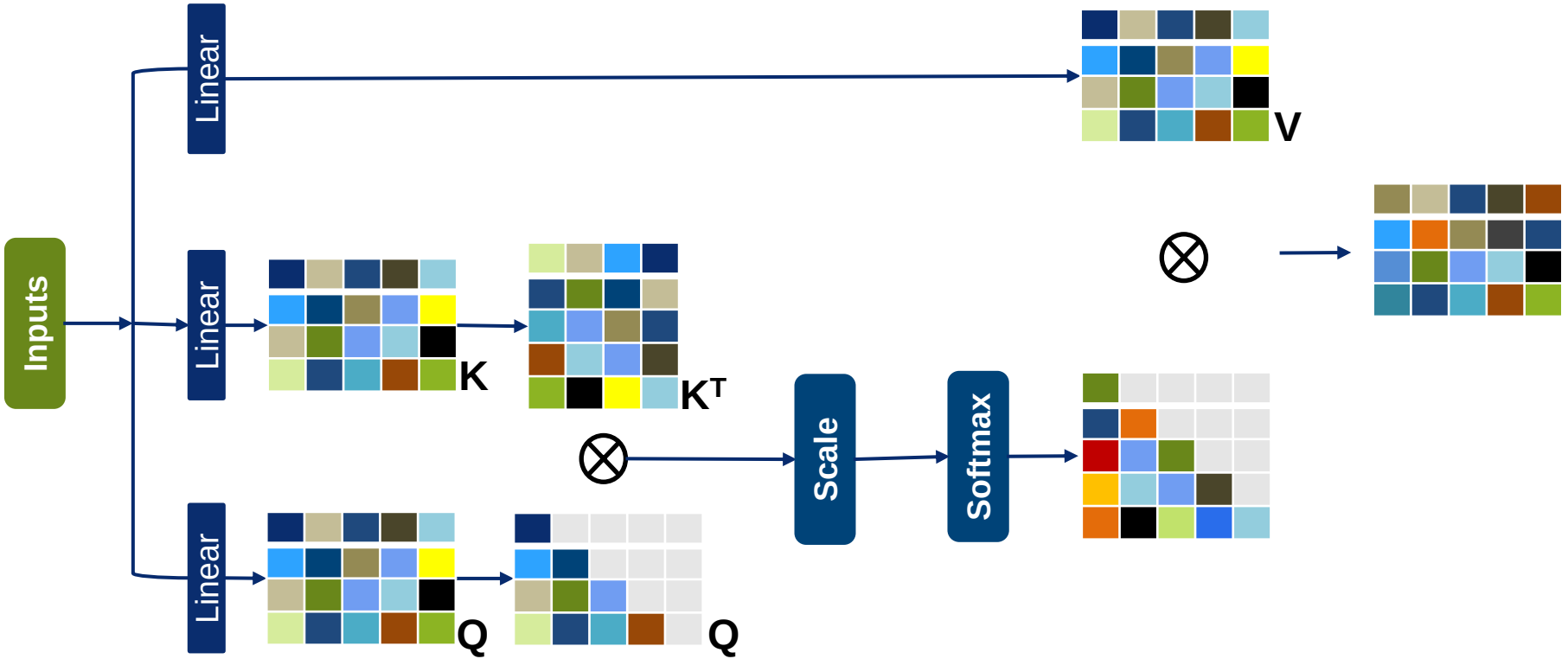
Data driven weights using probabilities (Softmax)



- Averaging as weight is lossy, thus not useful
- Softmax converts logits to probabilities

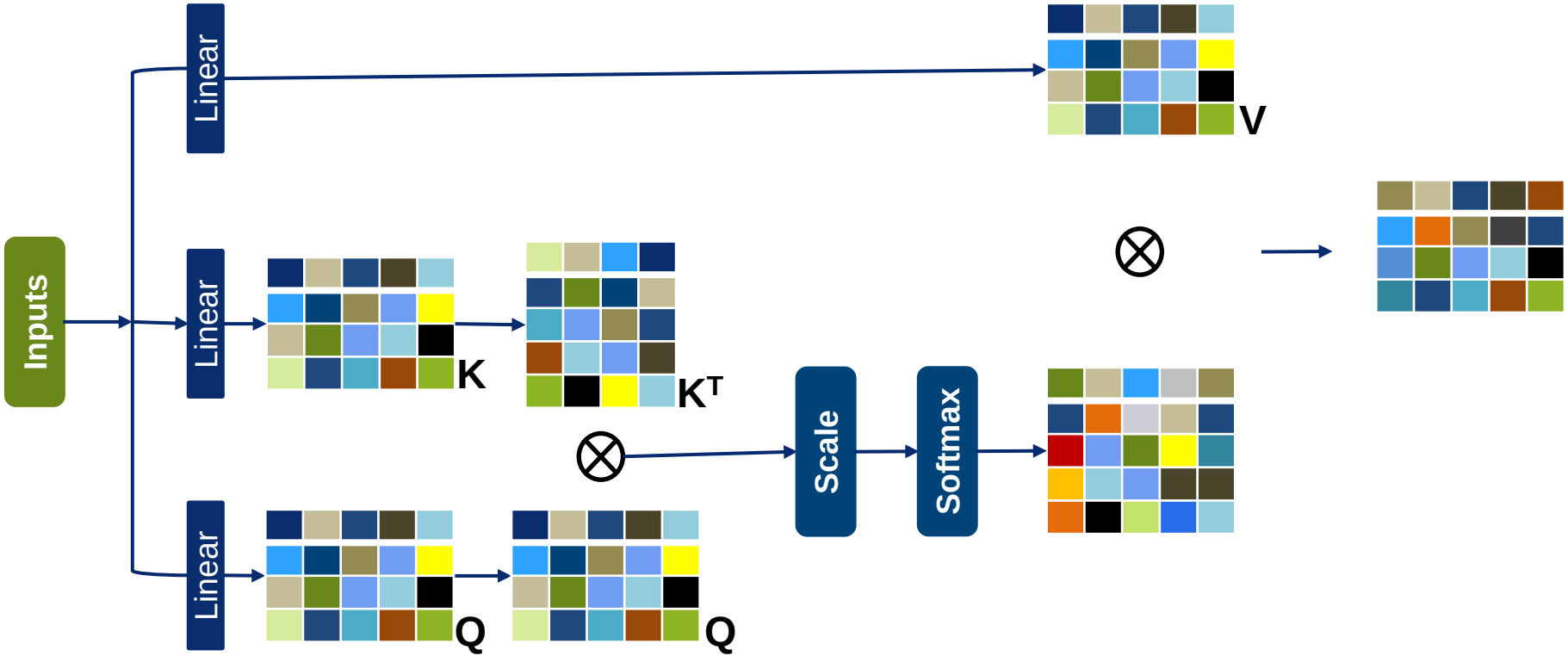
Helmholtz AI

Masked Self-attention



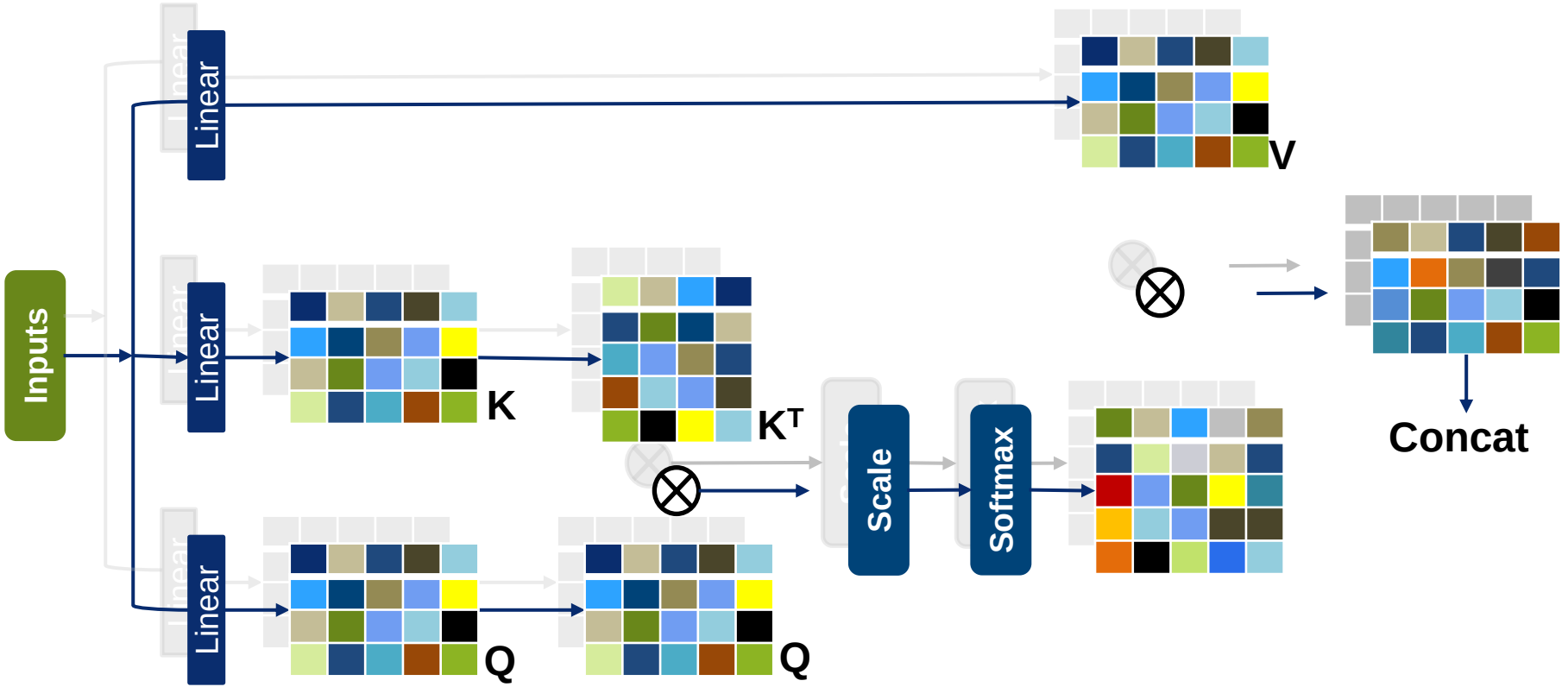
Helmholtz AI

Bi-directional Self-Attention



Helmholtz AI

Self-Attention Head size (BERT or Masked)



Helmholtz AI

DKRZ
DEUTSCHES
KLIMARECHENZENTRUM

QUESTIONS