CHAPTER 08 / MASK

Who can read whom?

Attention mixes positions. The mask sets the boundary.

Attention mixes positions. The mask sets the boundary.



Allowed / not allowed · illustrative grid

Illustrative masks. Attention weights alone do not explain an answer. Encoder-decoder cross-attention is additional.

Attention

Mix information across the positions the mask permits.

Feed-forward

Transform each position separately, with add-back paths around the sublayers.

Masks

A five-token causal mask permits 15 cells; full visibility permits 25; a three-token prefix permits 18.

↗ **Mix information across the positions the mask permits.**

PUT IT TO WORK

Read the five-token sentence one position at a time. A causal row never reads a later token.

CLOSE THE PAGE. EXPLAIN IT.

If words are appended, can an earlier causal position read them?