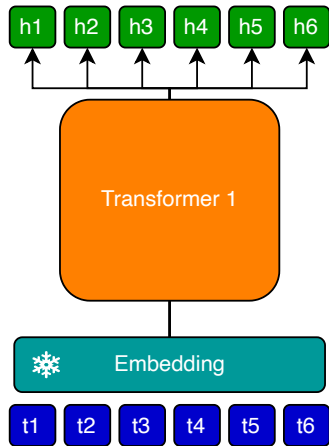


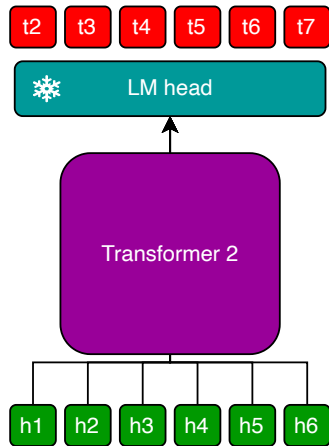
STEP 1:

Forward pass on model 1.
Produce dataset of hidden states.



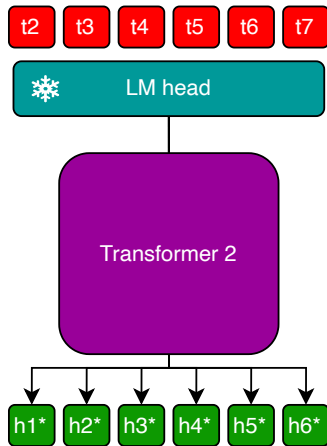
STEP 2:

Forward pass on model 2.
Calculate loss with respect to target
tokens.



STEP 3:

Backward pass on model 2.
Treat input-hidden-states as parameters
and update them.



STEP 4:

Forward pass on model 1; on the same tokens that produced the original
hidden states. Train against updated hidden states from previous step.
Optional: also train against the ground-truth tokens.

