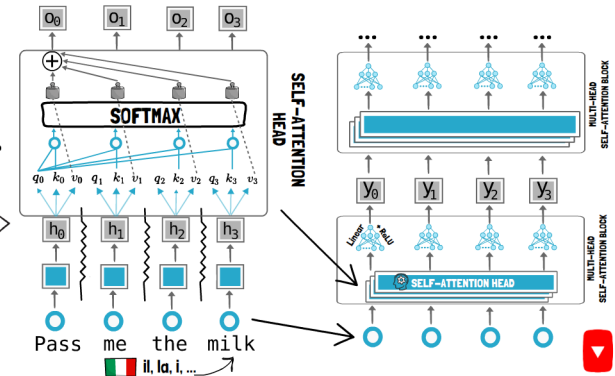
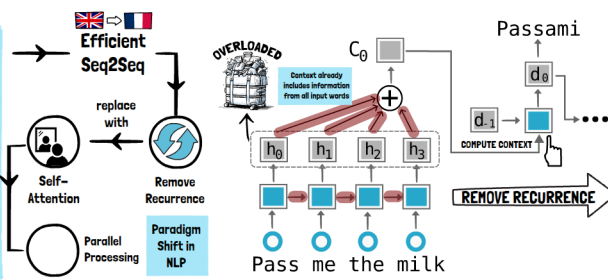
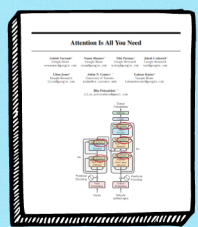
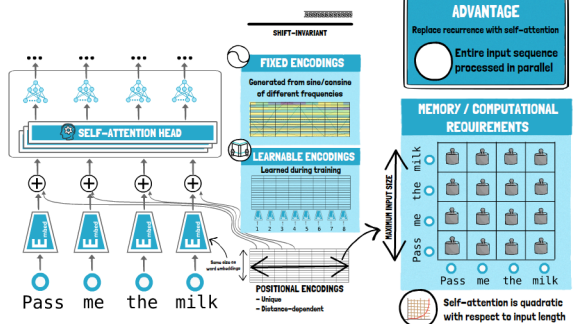
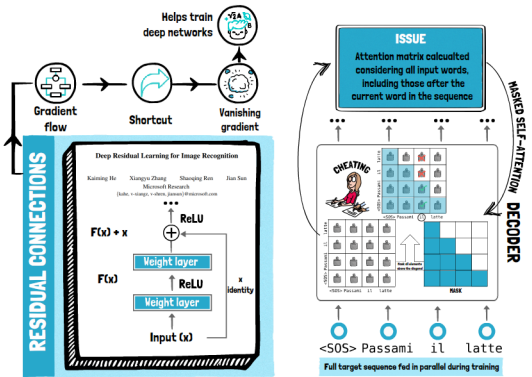
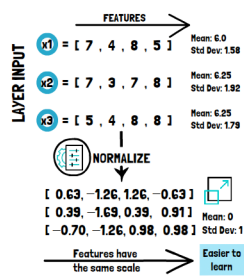


TRANSFORMERS



LAYER NORMALIZATION



ADVANTAGE

Replace recurrence with self-attention

Entire input sequence processed in parallel