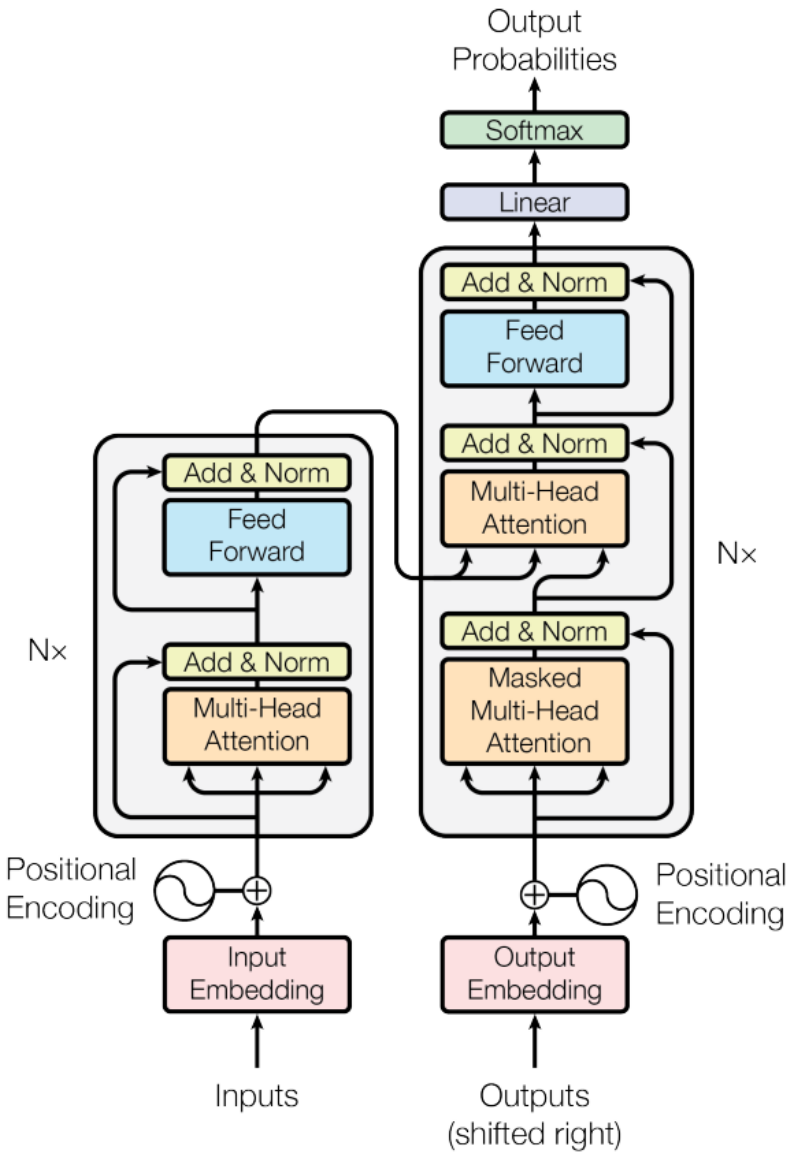
인공지능 심화교육
-자연어처리 실습 2

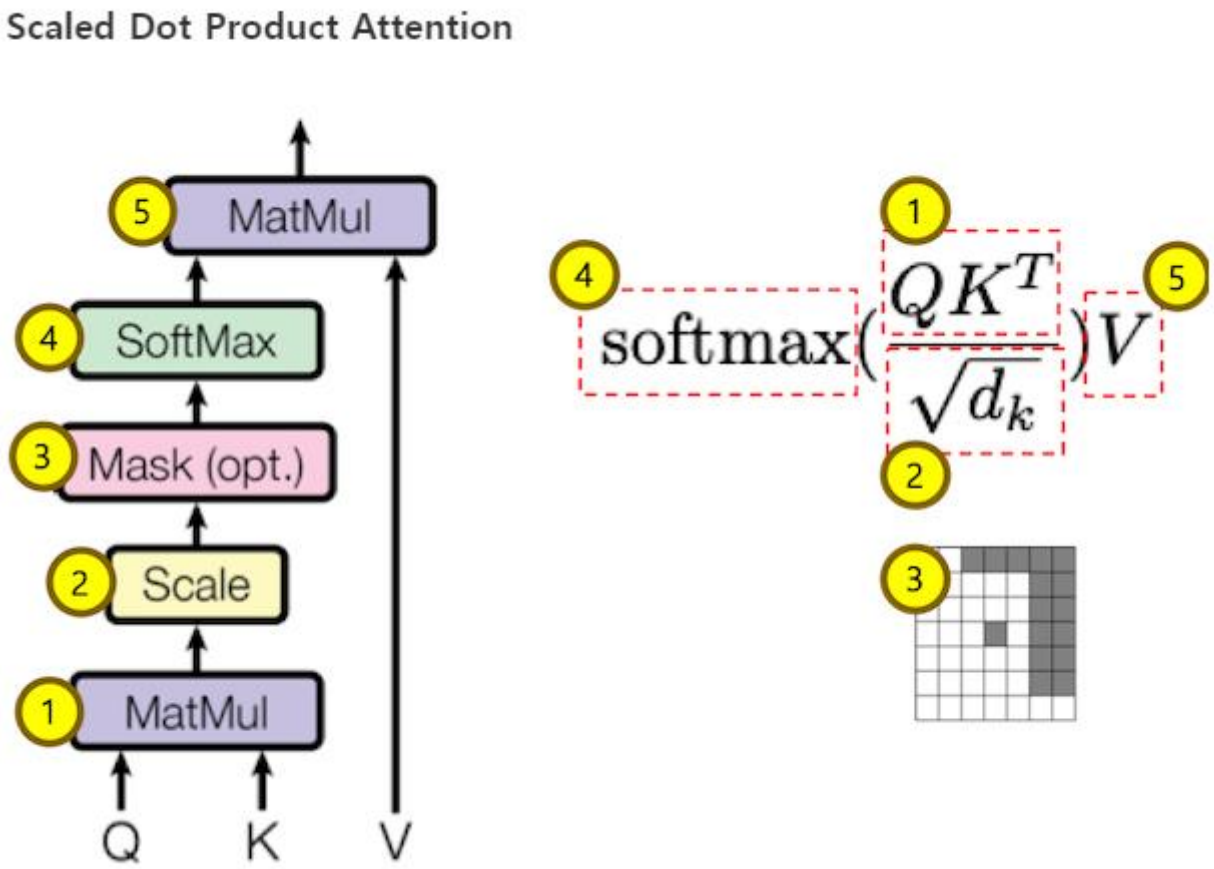
전북대학교 이성민

Transformer



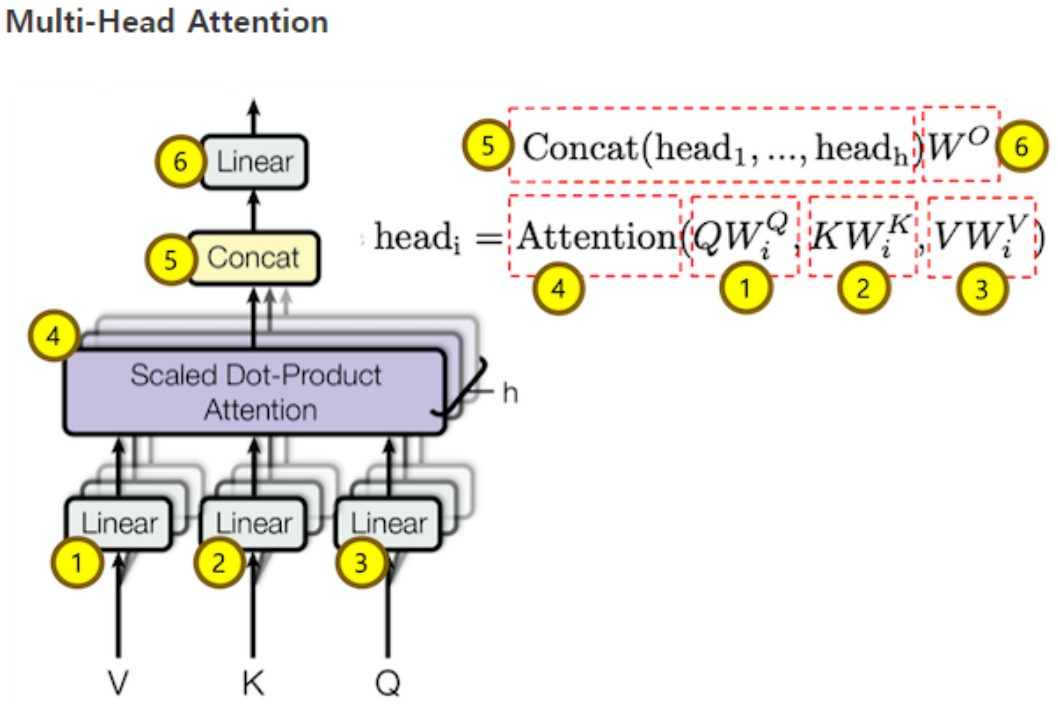
https://arxiv.org/pdf/1706.03762.pdf

Transformer



https://paul-hyun.github.io/transformer-02/

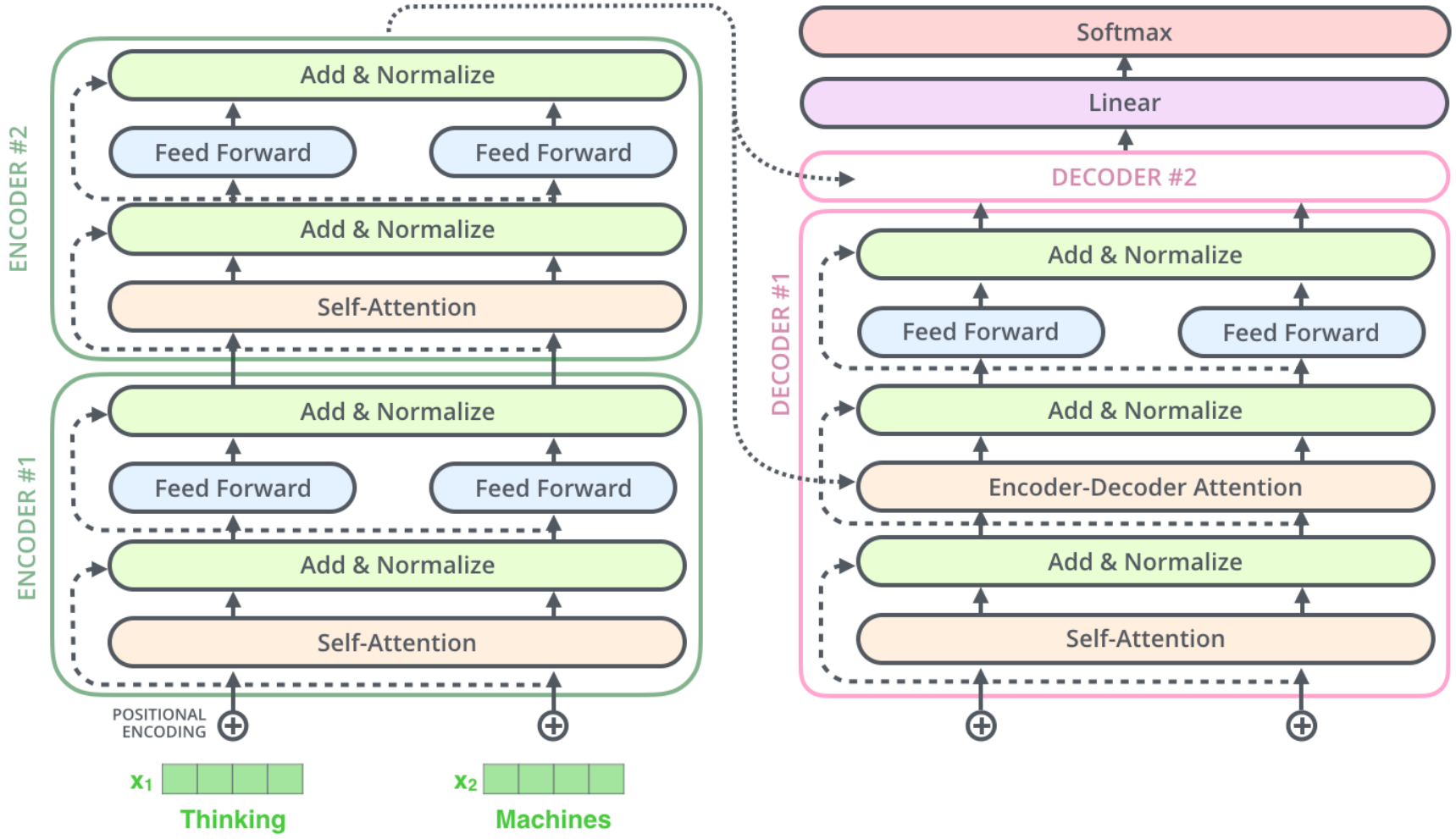
Transformer



$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, ..., \text{head}_\text{h})W^O$$
$$\text{where head}_\text{i} = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

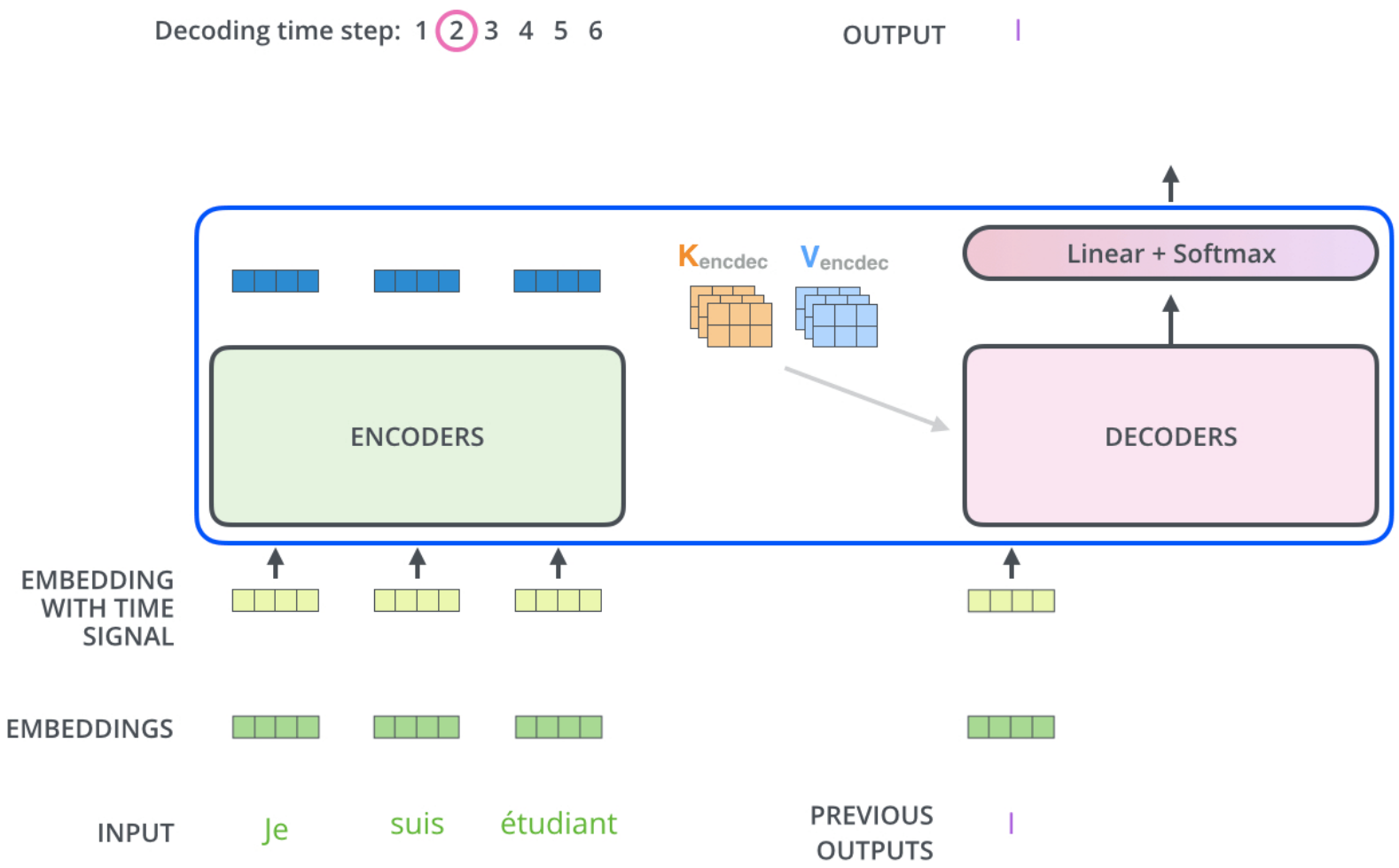
Where the projections are parameter matrices $W_i^Q \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_v}$ and $W^O \in \mathbb{R}^{hd_v \times d_{\text{model}}}$.

Transformer



https://jalammar.github.io/illustrated-transformer/

Transformer



https://jalammar.github.io/illustrated-transformer/

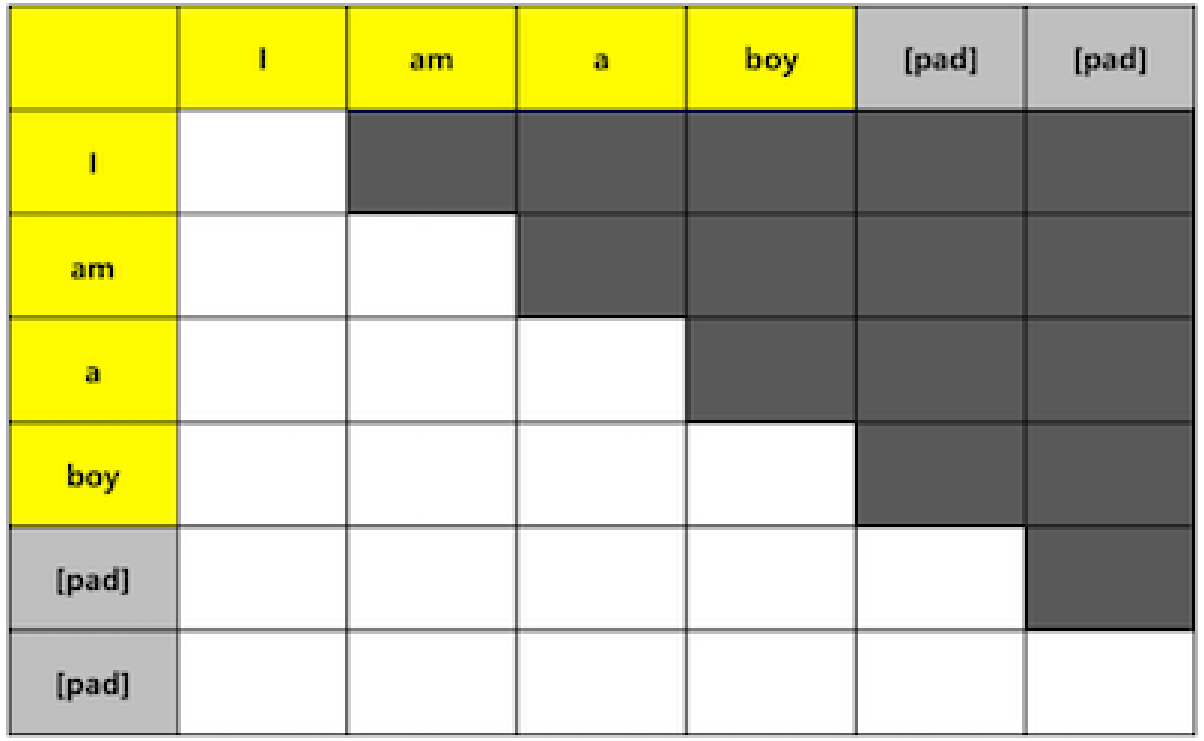
Transformer

Attention Pad Mask

https://paul-hyun.github.io/transformer-02/

Transformer

Attention Decoder Mask



https://paul-hyun.github.io/transformer-02/

Text Classification + Contrastive Loss

$$\mathcal{L}_{\text{s\_cl}} = -\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{1}_{y_i=y_j}\log\frac{\exp(\text{sim}(\mathbf{h}_i,\mathbf{h}_j)/\tau)}{\sum_{n=1}^{N}\exp(\text{sim}(\mathbf{h}_i,\mathbf{h}_n)/\tau)}$$

$$\mathcal{L}_{\text{intent}} = -\frac{1}{N}\sum_{j=1}^{C}\sum_{i=1}^{N}\log P(C_j|u_i)$$

$$\mathcal{L}_{\text{stage2}} = \mathcal{L}_{\text{s\_cl}} + \lambda'\mathcal{L}_{\text{intent}}$$

https://arxiv.org/pdf/2109.06349.pdf