‹¡{—lèa R›wé–5ÿ Attention Matrixÿ f/Transformerj!W‹N-v„h8_ÃdÍO\ÿ Qvvîv„f/• •Ç‹¡{—•"Qe^•R N-kÏN*OI

**1. •"QeQÆY **
PG‹¾•"Qe^•R v„]LQeˆhy:ÿ embeddingÿ N: \(X \in \mathbb{R}^{N \times C} \)ÿ QvN-ÿ
- \(N \) f/^•R ••^¦ÿ tokenep˙Ïÿ 0
- \(C \) f/kÏN*tokenv„~ô^¦ÿ Y,512b 768ÿ 0

---

2. ~¿`'SØcbu b Q0 K0 V
• •ÇN N*Sï[fN`v„gC'Íwé–5 \(W_Q, W_K, W_V \in \mathbb{R}^{C \times d} \) \ •"Qe \(X \) b•_qR0gå‹âÿ
\[
Q = X W_Q, \quad K = X W_K, \quad V = X W_V
\]
QvN- \(d \) f/lèa R›Y4v„~ô^¦ÿ • ^8 \(d = C / h \)ÿ \(h \) N:Y Y4lèa R›v„Y4epÿ 0

---

3. ‹¡{—lèa R›R ep
• •Çwé–5NXlÕ‹¡{—b@g token[ùv„vøO<^¦_—R ÿ g*_RN S v„lèa R›R epÿ ÿ
\[
P = Q K^T \in \mathbb{R}^{N \times N}
\]
kÏN*QC} \(P_{ij} \) ˆhy:{, \(i \) N*token[ù{, \(j \) N*tokenv„Qs€T_:^¦0

**4. •)e>N _RN S **
[ù \(P \) •Û^L•)e>ÿ –2kbh¯^¦mˆY1ÿ ^v^"u(Softmax_RN S ÿ
\[
A = \text{Softmax}\left(\frac{P}{\sqrt{d}} \right)
\]
QvN-ÿ
- •)e>Và[P \(\sqrt{d} \) u(NŽc§R6R epƒ Vô0
- SoftmaxxnOÝkÏˆLv„gC'ÍTŒN:1ÿ Ss \(\sum_j A_{ij} = 1 \)0

**5. lèa R›c©x ÿ Sï• ÿ **
,å— \O…=eàeHOM•nÿ Y,XkQEOM•nb g*geOá`oÿ ÿ W(SoftmaxRM\ [ù^"OM•nv„ \(P_{ij} \) QÏS»N N*g•
\[
P_{\text{masked}} = P + M, \quad M_{ij} =
\begin{cases}
0 & \text{QA‹¸Qslè}, \\

```
-\infty & \text{\O…=}.
\end{cases}
\]
```

**6. R gC€ZT **
u(_RN S v„lèa R›gC'Í \(A \) [ùP<wé–5 \(V \) R gClBTŒÿ _—R0•"Qúÿ
```
\[
\text{Output} = A V \in \mathbb{R}^{N \times d}
\]
```

**7. Y Y4lèa R›ÿ bi\Uÿ **
,åO•u(Y Y4lèa R›ÿ 'ÍY N •ðkeš¤ \(h \) k!ÿ büc¥b@g Y4v„•"Qú^v~¿`'b•_qÿ
```
\[
\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) W_O
\]
```
QvN- \(W_O \in \mathbb{R}^{h d \times C} \) f/•"Qúb•_qwé–50

Qs•.Ql_ `;~Ó
```
\[
\text{Attention}(Q, K, V) = \text{Softmax}\left( \frac{Q K^T}{\sqrt{d}} \right) V
\]
```

R¨ ViTv„e9•Û
W(DynamicViTN-ÿ • •Ç**lèa R›c©x ÿ Attention Maskingÿ **R¨ Rjg•Q—OYtokenÿ Y,ᵒe‡N-v„Ql_ 10-11ÿ
1. h9cn~„mKj!WWu b v„NŒEP<c©x \(\hat{D} \in \{0,1\}^N \) Œ etlèa R›‹¡{—0
2. OÝuY•ê•Þc¥ÿ \(G_{ii}=1 \)ÿ ÿ \O…=ˆ«Rjg•token[ùQvNÖtokenv„_qTÍÿ \(G_{ij} = \hat{D}_j \) _S \(i \n
3. g ~Èlèa R›wé–5‹¡{—N:ÿ
```
\[
\tilde{A}_{ij} = \frac{\exp(P_{ij}) G_{ij}}{\sum_k \exp(P_{ik}) G_{ik}}
\]
```

• •ÇN •ðkeš¤ÿ lèa R›wé–5€ýY R¨ • ^"•"QeQ…['ÿ QÏ\ ‹¡{—Q—OY0