Human Eye-gaze Human Annotation and Input Plot Attention Matrix (G) Recording Eye-gaze Data Value of leading corporate brands in the Asia Pacific region in 2019 (in billion euros) Value of leading corporate brands in the Asia Pacific region in 2019 (in billion euros $\longrightarrow \mathcal{L}_{\text{W-MSE}} = \frac{1}{N} \sum_{i=1}^{N} w_i \cdot (G_i - A_i)^2 -$ Value of leading corporate brands in the Asia Pacific region in 2019 (in billion euros PetroChin Pingan Insurance Brand value in billion euros **Question and Answers Attention Between Text and Generated Based on the Chart Vision Tokens Averaged over** $\mathcal{L}_{ ext{total}} = \mathcal{L}_{ ext{LM}} + \mathcal{L}_{ ext{W-MSE}}$ **Summaries From VisText All Heads and Text Tokens** Brand value in billion euros QA₁ **LARGE VISION** LANGUAGE MODEL **Average Over the First** Was Alibaba Group the most **M** Layers valuable brand at over €50 billion? **Answer: True**

Model Output Tokens:

True </s>

 $\mathcal{L}_{LM} = -\sum \log P(y_t \mid x_{\leq t})$