# HOW BERT CAPTURES TEXT

Identify order of input (with embeddings)



| INPUT | [CLS] | mannheim | is | a | beautiful | city | [SEP] | ... |
|---|---|---|---|---|---|---|---|---|
| TOKEN EMBEDDING | $E_{[CLS]}$ | $E_{mannheim}$ | $E_{is}$ | $E_a$ | $E_{beautiful}$ | $E_{city}$ | $E_{[SEP]}$ | ... |
| SEGMENT EMBEDDING | $E_A$ | $E_A$ | $E_A$ | $E_A$ | $E_A$ | $E_A$ | $E_A$ | ... |
| POSITION EMBEDDING | $E_0$ | $E_1$ | $E_2$ | $E_3$ | $E_4$ | $E_5$ | $E_6$ | ... |

cosima_meyer