



BERT<sub>base</sub>:  $H = 786$   $L = 12$  110M

BERT<sub>Large</sub>:  $H = 1024$   $L = 24$  340M