Model	SQuAD 1.1		SQuAD 2.0	
	EM	F1	EM	F1
Single models on dev, w/o data augmentation				
BERT$_{LARGE}$	84.1	90.9	79.0	81.8
XLNet$_{LARGE}$	**89.0**	94.5	86.1	88.8
RoBERTa	88.9	**94.6**	**86.5**	**89.4**
Single models on test (as of July 25, 2019)				
XLNet$_{LARGE}$			86.3[†]	89.1[†]
RoBERTa			86.8	89.8
XLNet + SG-Net Verifier			**87.0**[†]	**89.9**[†]