WMT17 Pearson Correlation Coeeficient with Human Judgements

Translation Task	bleu	macro_1gF1	micro_1gF1	macro_1gF_2gP	micro_1gF_2gP	macro_1gF_3gP	micro_1gF_3gP	macro_1gF_4gP	micro_1gF_4gP	macro_1gF_5gP	micro_1gF_5gP
Mean	0.906	0.874	0.921	0.884	0.916	0.886	0.913	0.886	0.911	0.885	0.908
SD	0.045	0.114	0.049	0.103	0.052	0.1	0.053	0.099	0.054	0.1	0.055
wmt17 cs-en	0.966	0.571	0.881	0.617	0.868	0.632	0.869	0.634	0.87	0.629	0.871
wmt17 de-en	0.929	0.963	0.947	0.961	0.936	0.958	0.929	0.955	0.922	0.953	0.915
wmt17 en-cs	0.954	0.956	0.961	0.951	0.954	0.949	0.949	0.948	0.945	0.948	0.942
wmt17 en-de	0.82	0.722	0.782	0.728	0.775	0.728	0.768	0.727	0.763	0.725	0.759
wmt17 en-fi	0.918	0.95	0.967	0.944	0.961	0.942	0.957	0.941	0.954	0.941	0.952
wmt17 en-lv	0.839	0.912	0.939	0.9	0.924	0.894	0.919	0.891	0.916	0.889	0.914
wmt17 en-ru	0.902	0.952	0.923	0.947	0.915	0.945	0.915	0.943	0.916	0.942	0.917
wmt17 en-tr	0.887	0.886	0.918	0.902	0.937	0.905	0.937	0.906	0.937	0.906	0.937
wmt17 fi-en	0.895	0.871	0.917	0.897	0.909	0.907	0.915	0.911	0.924	0.913	0.931
wmt17 lv-en	0.96	0.951	0.963	0.958	0.962	0.958	0.957	0.957	0.95	0.955	0.943
wmt17 ru-en	0.903	0.937	0.927	0.943	0.923	0.943	0.92	0.944	0.917	0.943	0.915
wmt17 tr-en	0.941	0.878	0.947	0.919	0.967	0.927	0.966	0.929	0.965	0.929	0.963
wmt17 zh-en	0.867	0.811	0.899	0.83	0.883	0.832	0.871	0.83	0.86	0.827	0.848

Metrics