Top 2 Retained 0.00 0.69 0.59 0.84 0.89 0.59 0.66 0.69 0.61 0.59 0.54 0.54 0.56

BERT - DuoRC Integrated Gradients JSD

0.69 0.00 0.50 0.50 0.45 0.75 0.54 0.61 0.65 0.64 0.60 0.60 0.59

0.84 0.50 0.49 0.00 0.17 0.81 0.45 0.66 0.83 0.83 0.79 0.79 0.79

0.89 0.45 0.56 0.17 0.00 0.87 0.51 0.73 0.90 0.91 0.87 0.87 0.87

0.59 0.75 0.63 0.81 0.87 0.00 0.60 0.55 0.47 0.48 0.54 0.55 0.57

0.66 0.54 0.31 0.45 0.51 0.60 0.00 0.40 0.48 0.47 0.43 0.41 0.41

0.69 0.61 0.40 0.66 0.73 0.55 0.40 0.00 0.26 0.29 0.38 0.37 0.36

0.61 0.65 0.41 0.83 0.90 0.47 0.48 0.26 0.00 0.10 0.22 0.21 0.21

0.59 0.64 0.40 0.83 0.91 0.48 0.47 0.29 0.10 0.00 0.19 0.18 0.19

0.54 0.60 0.31 <mark>0.79 0.87</mark> 0.54 0.43 0.38 0.22 0.19 0.00 0.03 0.05

0.54 0.60 0.29 0.79 0.87 0.55 0.41 0.37 0.21 0.18 0.03 0.00 0.03

0.56 0.59 0.28 <mark>0.79 0.87</mark> 0.57 0.41 0.36 0.21 0.19 0.05 0.03 0.00

10

0.59 0.50 0.00 0.49 0.56 0.63 0.31 0.40 0.41 0.40 0.31 0.29 0.28