Mean Context Saliency Score of the last 8 layers

Predicted Result:
- correct
- wrong

score value ($\times 10^{-3}$)

Context Length: 2k

Context Length	Evidence	Full	Attack	Remaining
correct	1.2	0.2	0.6	0.2
wrong	1.8	0.3	0.8	0.3