

Comparison of WER between regular and flash attention models

