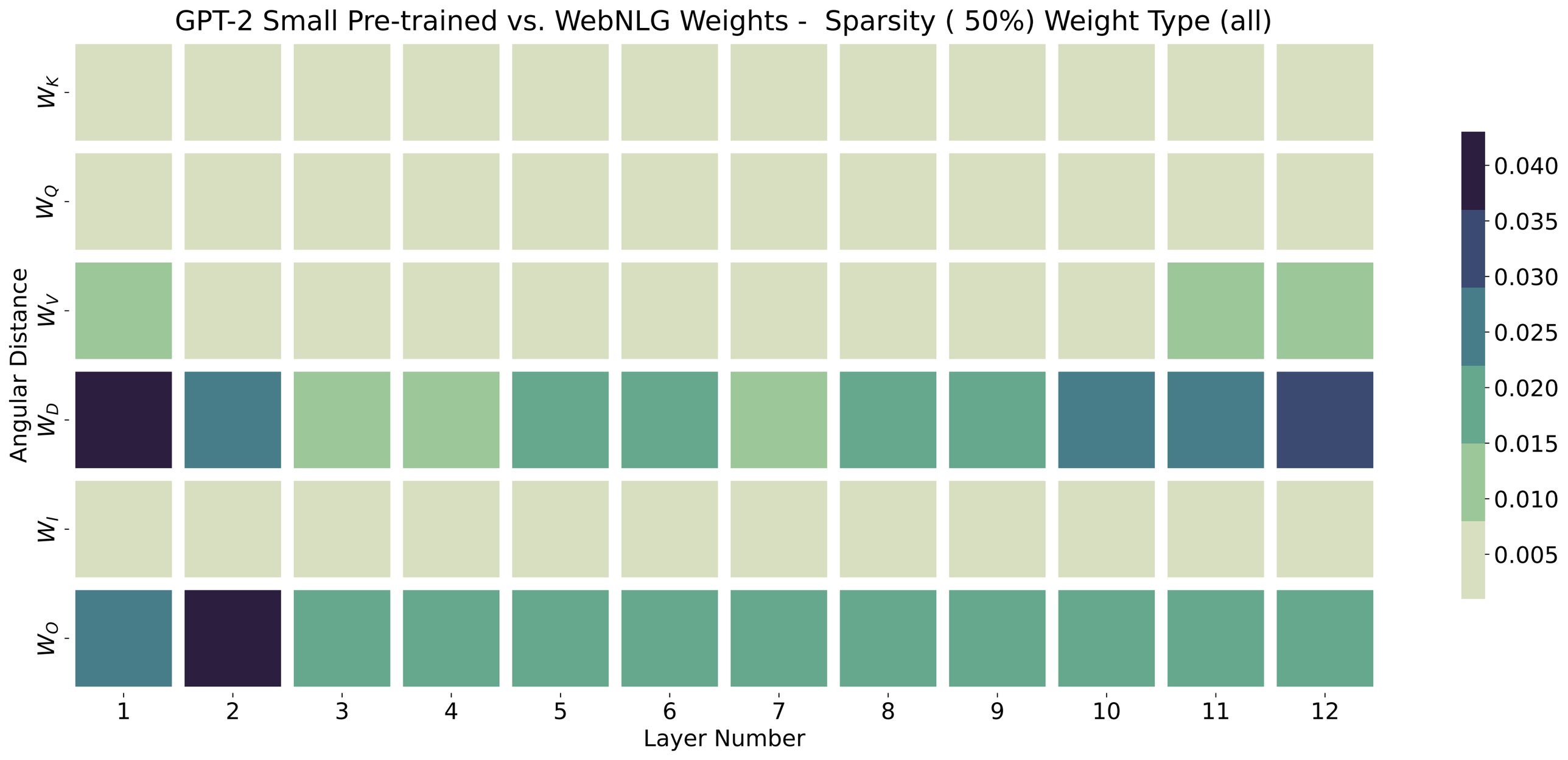
GPT-2 Small Pre-trained vs. WebNLG Weights - Sparsity (50%) Weight Type (all)
Angular Distance
W_K
W_Q
W_V
W_D
W_I
W_O
1
2
3
4
5
6
7
8
9
10
11
12
Layer Number
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005