

Effect size (Hedges'  $g$ )

2  
1  
0  
-1

GPT-3.5

GPT-4

