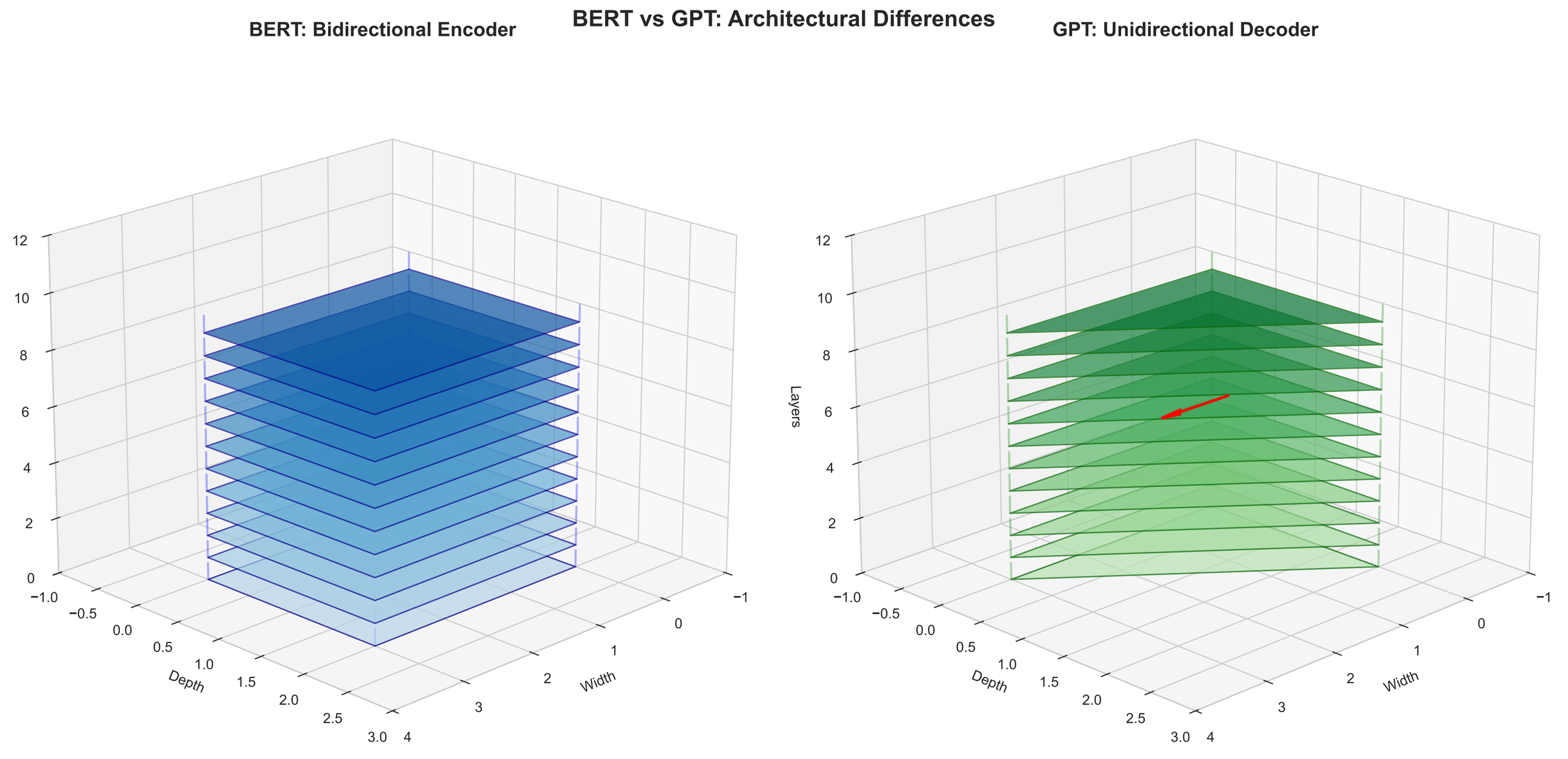
BERT vs GPT: Architectural Differences
BERT: Bidirectional Encoder
GPT: Unidirectional Decoder
Layers
Depth
Width
0
2
4
6
8
10
12
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
−1
0
1
2
3
4