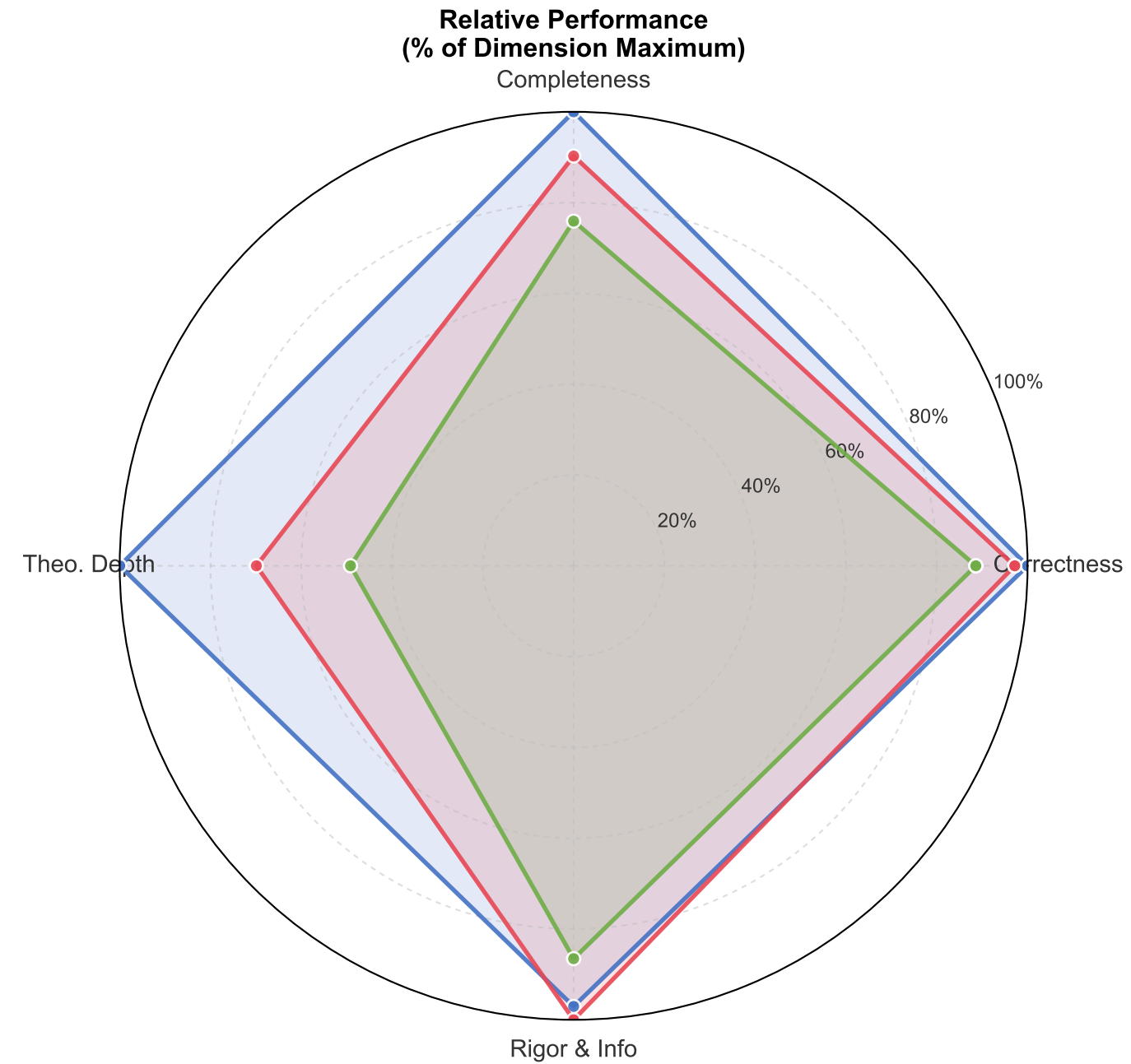
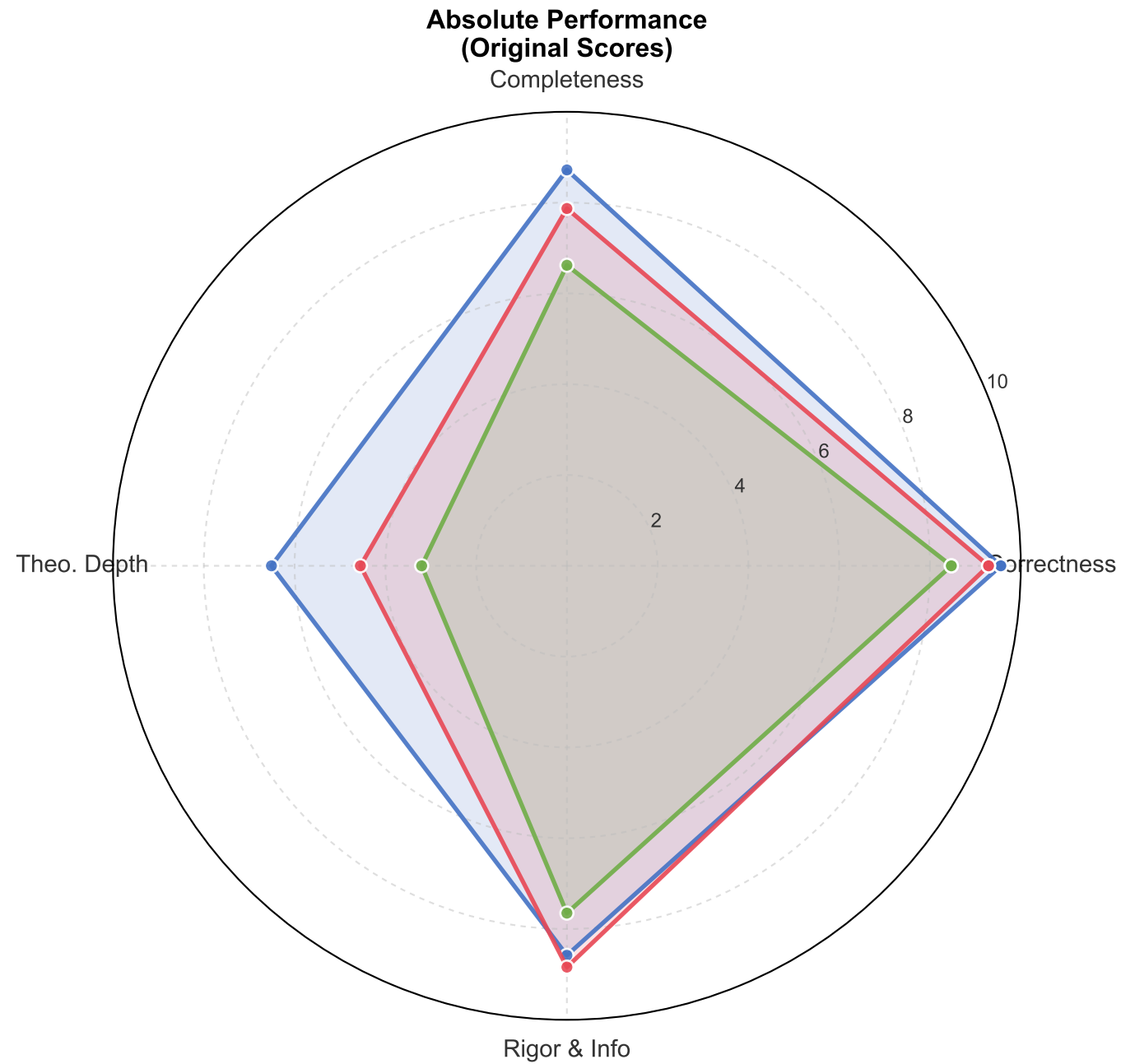


# Top Models Performance Analysis: Absolute vs. Relative Comparison



—●— MOSES —●— O3 —●— GPT-4.1